ADVANCES IN INTELLIGENT SYSTEMS AND COMPUTING 206

Álvaro Rocha
Ana Maria Correia
Tom Wilson
Karl A. Stroetmann (Eds.)

Advances in Information Systems and Technologies

Springer

Advances in Intelligent Systems and Computing 206

For further volumes:
http://www.springer.com/series/11156

Álvaro Rocha, Ana Maria Correia, Tom Wilson,
and Karl A. Stroetmann (Eds.)

Advances in Information Systems and Technologies

Springer

Editors
Dr. Álvaro Rocha
LIACC
University of Porto
Porto
Portugal

Dr. Ana Maria Correia
Instituto Superior de Estatística e
Gestão de Informação
Campus de Campolide
Universidade Nova de Lisboa
Lisboa
Portugal

Dr. Tom Wilson
Sheffield
United Kingdom

Dr. Karl A. Stroetmann
Empirica GmbH
Bonn
Germany

ISSN 2194-5357 ISSN 2194-5365 (electronic)
ISBN 978-3-642-36980-3 ISBN 978-3-642-36981-0 (eBook)
DOI 10.1007/978-3-642-36981-0
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013932785

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

These proceedings contain all papers accepted for presentation and discussion at The 2013 World Conference on Information Systems and Technologies (WorldCIST'13). This Conference is organised by AISTI (Associação Ibérica de Sistemas e Tecnologias de Informação /Asociación Ibérica de Sistemas y Tecnologías de Información / Iberian Association for Information Systems and Technologies) and took place on 27th-30th of March in Olhão, Algarve, Portugal.

The World Conference on Information Systems and Technologies (WorldCIST) is a global forum for researchers and practitioners to discuss their most recent research, innovations, trends, results, experiences and concerns, also in view of recent and foreseeable technical trends as well as national and global policies. The meeting covers all relevant domains of Information Systems and Technologies, including Knowledge Management, Organisational Models, Decision Support Systems, Software Systems – Architectures - Applications – Tools, Computer Networks - Mobility - Pervasive Systems, Radar Technologies, Human-Computer Interaction. One of its main aims is to strengthen the drive towards symbiosis between academy, society and industry in all these fields and their application domains.

The Program Committee of WorldCIST'13 is composed of a multidisciplinary group of researchers and experts who are intimately concerned with Information Systems and Technologies research and application. They have had the responsibility for evaluating, in a 'blind review' process, the papers received for each of the main themes proposed for the Conference: A) Information and Knowledge Management (IKM); B) Organizational Models and Information Systems (OMIS); C) Intelligent and Decision Support Systems (IDSS); D) Software Systems, Architectures, Applications and Tools (SSAAT); E) Computer Networks, Mobility and Pervasive Systems (CNMPS); F) Radar Technologies (RAT); G) Human-Computer Interaction (HCI).

WorldCIST'13 received contributions from thirty-three countries. The papers accepted for its presentation and discussion at the Conference are published by Springer and will be indexed by ISI, EI, SCOPUS, DBLP and EBSCO, among others. Selected papers will be published in relevant journals and in a SCI (Studies in Computational Intelligence) series book.

We acknowledge all those contributing to the staging of WorldCIST13 (authors, committees and sponsors); their involvement is very much appreciated. It is our aim that WorldCIST becomes the global forum for discussing both latest developments in information systems and technologies RTD as well as in its varied application fields.

January 2013

Álvaro Rocha
Ana Maria Correia
Tom Wilson
Karl A. Stroetmann

Organization

Co-chairs

Ana Maria Ramalho Correia, Full Visiting Professor, University Nova de Lisboa, Rtd, PT; Associate Professor, Department of Information Sciences, University of Sheffield, UK

Tom Wilson, Emeritus Professor, University of Sheffield, UK; Visiting Professor, University of Leeds, UK; Visiting Professor, University of Boras, SE

Karl Stroetmann, Senior Research Fellow, Empirica Communication & Technology Research, DE

Álvaro Rocha, President of AISTI; Senior Research Fellow, LIACC, University of Porto, PT; Invited Professor, University of Santiago de Compostela, ES

Advisory Committee

Bezalel Gavish	Southern Methodist University, Dallas, US
Carlos Vaz de Carvalho	IEEE; Polytechnic of Porto, PT
Ezendu Ariwa	London Metropolitan University, UK
Gustavo Rossi	La Plata National University, Lifia, AR
Ivan Lukovic	University of Novi Sad, RS
Janusz Kacprzyk	Systems Research Institute, Polish Academy of Sciences, PL
Lei Li	Hosei University, JP
Luis Paulo Reis	University of Minho, PT
Marcelo Sampaio Alencar	Federal University of Campina Grande, BR

Program Committee

A. Aladwani	Kuwait University, KW
Abdallah AL Sabbagh	University of Technology, Sydney (UTS), AU
Adrian Florea	'Lucian Blaga' University of Sibiu, RO
Alexandre Balbinot	Federal University of Rio Grande do Sul, BR
Ali Elrashidi	University of Bridgeport, US
Alisson Brito	Federal University of Paraíba, BR
Amnart Pohthong	Prince of Songkla University, TH
Andrey Koptyug	Mid Sweden University, SE
Ankit Chaudhary	Bits Pilani, IN
Arthur Taylor	Rider University, US
Arun Gupta	M.S.College, IN
Avireni Srinivasulu	Vignan University, IN
Berk Canberk	Istanbul Technical University, TR
Bernard Cousin	University of Rennes 1 - IRISA, FR
Bharati Rajan	Loyola College, Chennai, IN
Bin Zhou	University of Maryland, Baltimore County, US
Bridget Kane	Norwegian University of Science and Technology, NO
Carla Pinto	Polytechnic of Porto, PT
Carlos Costa	ISCTE - IUL, PT
Cathryn Peoples	University of Ulster, UK
Cesar Alexandre de Sousa	University of São Paulo, BR
Chien-Ta Bruce Ho	National Chung Hsing University, TW
Ching-Hsien Hsu	Chung Hua University, TW
Christopher Martinez	University of New Haven, US
Christos Bouras	University of Patras and CTI&P-Diophantus, GR
Christos Douligeris	University of Piraeus, GR
Concha Soler	Radiotelevisión Valenciana, ES
Constantinos Hilas	Technological Educational Institute of Serres, GR
Cristina Alcaraz	University of Malaga, ES
Cristina Olaverri Monreal	Technische Universität München, DE
Damon Shing-Min	National Chung Cheng University, TW
David Mason	Victoria University of Wellington, NZ
Dinko Vukadinovic	University in Split, HR
Dirk Thorleuchter	Fraunhofer INT, DE
Dragan Milivojevic	Mining and Metallurgy Institute - Informatics Department, RS
Duggirala Venkata	TJPS PG College, IN
Eda Marchetti	CNR-ISTI, IT
Eli Upfal	Brown University, US
Elisa Francomano	University of Palermo, IT
Enrique Herrera-viedma	University of Granada, ES
Ernesto Exposito	LAAS-CNRS, FR
Erol Egrioglu	Ondokuz Mayıs University, TR

Fabrizio Montesi	IT University of Copenhagen, DK
Farhan Siddiqui	Walden University, CA
Fernando Almeida	University of Porto, PT
Fernando Bobillo	University of Zaragoza, ES
Fernando J Aguilar	Universiity of Almería, ES
Fernando Morgado Dias	University of Madeira, PT
Francesco Bianconi	Università degli Studi di Perugia, IT
Filipe Portela	Universidade do Minho, PT
Filippo Neri	University of Malta, MT
Fu-Chien Kao	Da-Yeh University, TW
Gabriele Oliva	University Campus Biomedico, IT
Garyfallos Arabatzis	Democritus University of Thrace, GR
Gitesh Raikundalia	Victoria University, AU
Gonçalo Paiva Dias	University of Aveiro, PT
Guenther Ruhe	University of Calgary, CA
Habiba Drias	University of Sciences and Technology Houari Boumediene, DZ
Hamdi Tolga Kahraman	Karadeniz Technical University, TR
Hamid Alinejad-Rokny	University of Newcastle, AU
Hari Om	Indian School of Mines Dhanbad, IN
Hartwig Hochmair	University of Florida, US
Hernani Costa	Universidade de Coimbra, PT
Hing Kai Chan	Norwich Business School, University of East Anglia, UK
Hironori Washizaki	Waseda University, JP
I-Hsien Ting	National University of Kaohsiung, TW
Isabel Pedrosa	Polytechnic Institute of Coimbra, PT
Ilaim Costa Junior	Fluminense Federal University, BR
J. Octavio Gutierrez-Garcia	ITAM, MX
Jaflah University of Bahrain	University of Bahrain, BH
Jan Nolin	Swedish School of Library and Information Science, SE
Jannica Heinstrom	Abo Akademi University, FI
Jari Makinen	Tampere University of Technology, FI
Jérôme Darmont	University of Lyon, FR
Jezreel Mejia	Centro de Investigacion en Matematicas Unidad Zacatecas, MX
Jianhua Chen	Louisiana State University, US
Jian Li	Tulane University, US
Jingfeng Xia	Indiana University, US
José Braga de Vasconcelos	University Fernando Pessoa, PT
Jose C. Valverde	University of Castilla-La Mancha, ES
Jose Luis Herrero Agustin	University of Extremadura, ES
José Luis Reis	ISMAI, PT
José Machado	University of Minho, PT

José Vicente Rodríguez Muñoz	University of Murcia, ES
Juan Carlos Torres	University of Granada, ES
Judith Broady-Preston	Aberystwyth University, UK
Jyh-Horng Chou	National Kaohsiung First University of Science and Technology, TW
Kaoru Hirota	Tokyo Institute of Technology, JP
Kewei Sha	Oklahoma City University, US
Kevin Ho	University of Guam, GU
Khaled Sayed	Modern University for Technology and Information, EG
Konstantinos Baltzis	Aristotle University of Thessaloniki, GR
Konstantinos Lakakis	Aristotle University of Thessalonikim GR
Kuan Yew Wong	Universiti Teknologi Malaysia, MY
Kun Chang Lee	Sungkyunkwan University, KR
Lori McCay-Peet	Dalhousie University, CA
Luis Velez Lapão	University Nova de Lisboa, PT
Lynne Conventry	Northumbria University, UK
Manfred Thaller	University at Cologne, DE
M. Rosario Fernández Falero	University of Extremadura, ES
Maciej Dabrowski	Digital Enterprise Research Institute, NUI Galway, IE
Manolis Vavalis	University of Thessaly, GR
Mansaf Alam	Jamia Millia Islamia - A Central University, IN
Manuel Cota	University of Vigo, ES
Manuel Mazzara	United Nations University, CN
Manuel Silva	Polytechnic of Porto, PT
Marco Scaioni	Tongji University, Shanghai, CN
Maria Bermudez-Edo	University of Granada, ES
Maria Lee	Shih Chien University, TW
Mário Pinto	Polytechnic Institute of Porto, PT
Martin Zelm	INTEROP VLab, BE
Masoud Abbaszadeh	Maplesoft, CA
Matthias Galster	University of Groningen, NL
Mazdak Zamani	Universiti Teknologi Malaysia, MY
Michal Strzelecki	Technical University of Lodz, PL
Mirjana Kljajic Borstnar	University of Maribor, SI
Mirna Ariadna Muñoz Mata	Centro de Investigación en Matemáticas- Unidad Zacatecas, MX
Miroslav Koncar	Oracle, HR
Mohammad Reza Daliri	Iran University of Science and Technology, IR
Mu-Song Chen	Da-Yeh University, TW
Natheer Gharaibeh	Al-Balqa Applied University, JO
Newlin Rajkumar Manokaran	Ann University of Technology, Rajkumar, IN
Niels Lohmann	Universität Rostock, DE

Noemi Emanuela Cazzaniga	Politecnico di Milano, IT
Nuno Fernandes	Polytechnic Institute of Castelo Branco, PT
Patrick Wang	Northeastern University, US
Panos Balatsoukas	University of Manchester, UK
Paolo Bientinesi	RWTH Aachen, DE
Pascal Lorenz	University of Haute Alsace, FR
Paweł Karczmarek	The John Paul II Catholic University of Lublin, PL
Pedro Peris	Computer Security Lab (COSEC), Carlos III University of Madrid, ES
Pedro Sousa	Universidade do Minho, PT
Peng-Sheng Chen	National Chung Cheng University, TW
Persephone Doupi	National Institute for Health and Welfare - THL, FI
Pierpaolo D'Urso	Sapienza - University of Rome, IT
Pierre L'Ecuyer	Université de Montréal, CA
Pierre Robillard	Polytechnique Montréal, CA
Preben Hansen	Swedish Institute of Computer Science, SE
Radu Prodan	University of Innsbruck, AT
Rainer Schmidt	HTW - Aalen University, DE
Ramiro Gonçalves	University of Trás-os-Montes e Alto Douro, PT
Rashid Ali	Taif University, SA
Rébecca Deneckère	University of Paris 1 Panthéon-Sorbonne, FR
Ren-Song Ko	National Chung Cheng University, TW
Rimvydas Skyrius	Vilnius University, LT
Robert Walker	University of Calgary, CA
Roberto Montemanni	Dalle Molle Institute for Artificial Intelligence (IDSIA), CH
Rubén González Crespo	Pontifical University of Salamanca, ES
Saeed Tavakoli	University of Sistan and Baluchestan, IR
Saleem Abuleil	Chicago State University, US
Salim Bitam	University of Biskra, DZ
Sandra Costanzo	University of Calabria, IT
Sangkyun Kim	Kangwon National University, KR
Sarang Thombre	Tampere University of Technology, FI
Santosh Kumar Nanda	Eastern Academy of Science and Technology, IN
Sergio Escalera	University of Barcelona, ES
Shaikh Abdul Hannan	Vivekanand College (LPP), Maharashtra, IN
Shamim Khan	Columbus State University, US
Sherali Zeadally	University of the District of Columbia, US
Sirje Virkus	Tallinn University, EE
Stasa Milojevic	Indiana University Bloomington, US
Stefan Jablonski	University of Bayreuth, DE
Stefan Wagner	University of Stuttgart, DE
Suksant Sae Lor	HP Labs, UK
Thomas Blaschke	University of Salzburg, AT
Tossapon Boongoen	Royal Thai Air Force Academy, TH

Tzung-Pei Hong	National University of Kaohsiung, TW
Valentina Emilia Balas	Aurel Vlaicu University of Arad, RO
Vitalyi Igorevich Talanin	Zaporozhye Institute of Economics and Information Technologies, UA
Waqas Bangyal	Iqra University Islamabad, PK
Wojciech Cellary	Poznan University of Economics, PL
Wolf Zimmermann	Martin-Luther-University Halle-Wittenberg, DE
WY Szeto	The University of Hong Kong, HK
Xiangmin Zhang	Wayne State University, US
Xiaoli Li	Institute for Infocomm Research, SG
Yafia Radouane	Ibn Zohr University, MA
Yaohang Li	Old Dominion University, US
Yair Wiseman	Bar-Ilan University, IL
Yi Gu	University of Tennessee at Martin, US
Yo-Ping Huang	National Taipei University of Technology, TW
Yong-Hyuk Kim	Kwangwoon University, KR
Yuhua Li	University of Ulster, UK
Yunqing Xia	Tsinghua University, CN
Zehua Chen	Taiyuan University of Technology, CN
Zhefu Shi	University of Missouri, Kansas City, US
Zhisheng Huang	Vrije University Amsterdam, NL
Zonghua Gu	Zhejiang University, CN

Contents

Information and Knowledge Management

Organizational Models and Information Systems

Intelligent and Decision Support Systems

Software Systems, Architectures, Applications and Tools

Computer Networks, Mobility and Pervasive Systems

Radar Technologies

Human-Computer Interaction

Knowledge Acquisition Activity in Software Development

Olivier Gendreau and Pierre N. Robillard

Department of Computer and Software Engineering
Polytechnique Montréal,
C.P. 6079, Succ Centre-Ville
Montréal, Qc, Canada, H3C 3A7
{olivier.gendreau,pierre.robillard}@polymtl.ca

Abstract. Data from four field studies are analyzed to find the patterns of knowledge acquisition activity in software development projects with respect to other cognitive activities such as documentation, coding and V&V. The data are obtained from self-recorded activity time slips approach. Data are codified based on an information source model, which is related to Nonaka and Takeuchi's knowledge creation model. It shows that knowledge acquisition activities account for almost 15% of the total effort. We also find out that this effort, in most cases, cannot be restricted to the first phase of the project during requirement and architectural design, which is expected from waterfall or disciplined processes. About half of the learning is done during the code implementation even within a disciplined process. This finding is in line with one value of the Agile philosophy that promotes team interactions and users involvement for the whole project duration.

Keywords: Knowledge acquisition, software development, cognitive activities, knowledge flow, cognitive factors, empirical studies, field studies.

1 Introduction

Software engineering is a knowledge-intensive activity [1,2,3,4]. Software development requires programmers to gather and absorb large amounts of knowledge distributed over several domains, such as application and programming, and to encode that knowledge in the software [3].

In order to better understand the complexity of software development, Ko et al. [5] suggest analyzing software activities from a knowledge perspective. However, the nature of knowledge poses a methodological challenge. Since knowledge is the product of various cognitive activities and mostly resides in a software developer's mind, it might be better described by the developers themselves. Therefore, this study is based on a data acquisition approach in which software developers record their activities from a knowledge viewpoint. This approach is used to gain an understanding of how knowledge acquisition needs evolve throughout the development of a software project.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1–10.
DOI: 10.1007/978-3-642-36981-0_1

In the cognitive sciences, four knowledge models are referred to widely: Kolb's model of experiential learning [6], Argyris and Schön's double-loop learning theory [7], Wenger's theory of communities of practice [8], and Nonaka and Takeuchi's theory of knowledge creation [9], which is the model used for this study. We describe six cognitive factors involved in software development projects. Data were collected from four selected industrial capstone projects, which were based on requirements supplied by a single avionics industrial partner.

Section 2 presents the knowledge model used to describe the cognitive activities. Section 3 describes the four field studies that provide the data. Section 4 presents the self recording Activity Time Slip (ATS) approach. Section 5 presents the knowledge acquisition patterns observed in the field studies.

2 Information Source Model

We used an information source model (see Fig. 1) which is related to Nonaka and Takeuchi's knowledge creation model [10] to analyze software development from a knowledge acquisition perspective.

The three round-cornered boxes in Fig. 1 represent information sources from which developers can built their knowledge. External information may come from various sources, such as the Web, a paper, or a book or technical documentation related to the product to be developed. Development artifact information comes from any of the project's documentation. Source code strictly includes executable statements and comments. Tacit knowledge is individual knowledge built from interaction with information sources. The arrows in Fig. 1 represent the various cognitive factors that constitute the knowledge built up from the various information sources. The acquisition cognitive factor is involved when a developer needs to increase his tacit knowledge by taking in external information. The crystallization cognitive factor is the translation of a developer's mental representation of a concept (tacit knowledge) into an artifact (explicit knowledge), such as a use case diagram or an architectural plan. The realization cognitive factor involves the translation of tacit knowledge into explicit information or documentation artifacts, but requires, in addition, technical know-how, which is related to source code production. The validation cognitive factor involves bidirectional information flow between tacit knowledge and development artifacts (explicit knowledge), in order to validate the consistency between the knowledge and the information source. The verification cognitive factor is like validation, except that source code is the information source, and so it involves technical know-how. The work planning cognitive factor mostly involves developers' synchronization of the project's planning and progress information.

This information source model is limited to software development activities. The management activities related to the software project are not taken into consideration in it, because they are not specific to software development, and also because they frequently involve several projects.

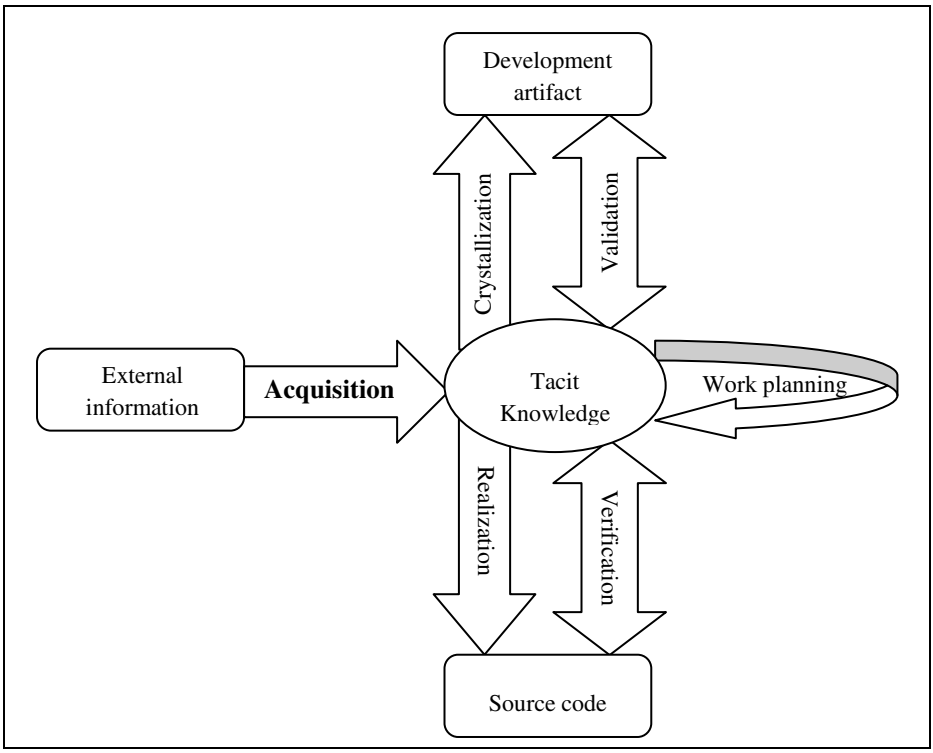


Fig. 1. Information source model

3 Field Studies

Data were collected from four selected industrial capstone projects conducted at our engineering school. The projects labeled P06 to P09 were successful in terms of teamwork and functional deliverables. The four projects were based on requirements supplied by a single avionics industrial partner.

The teams of five students were formed based on four criteria: number of cumulative total credits, past internship experience in industry, current grade point average, as well as software design and process course grades. They were chosen with the objective of balancing the know-how and experience on the teams.

The capstone projects were conducted over one semester (14 weeks) on a fixed schedule of three half-day team working sessions per week, and a flexible schedule of up to three extra half-days per week. The teams had access to an equipped dedicated room on campus for the duration of the project. All the projects, which are briefly described, are related to avionic applications and required adding some functionality to existing software systems.

The P06 project required to add a graphical interface to a design and configuration system used to build avionic model and enable editing of the various system model components and messages.

The P07 project required to build a graphical interface to create, delete or modify the hierarchy of the groups in Doxygen, which is a documentation system (under GNU General Public License) that generates documentation from source code [11].

The P08 project required to build a tool that will automatically extract data from a video of a plane cockpit dashboard. This tool is based on OCR (Optical character recognition) software. Once the video is loaded, the user can define the areas from which the data must be extracted. The extracted data are saved on a CVS file for further processing.

The P09 project required to translate a proprietary file format into a Microsoft Windows Presentation Foundation Format (WPF) [12]. The new files must have the same functionality as the old ones statically and dynamically. This project involved writing from scratch a home-made parser.

The external validity of empirical studies with students is a commonly raised concern. According to Carver et al. [13], more and more students are employed for either a summer internship or a full internship in an industrial environment. Höst et al. [14] conclude that only minor differences exist between students and professionals, and their research does not challenge the assumption that final year software engineering students are qualified to participate in empirical software engineering research. Similar results were obtained in a study on detection methodologies for software requirement inspection conducted by Porter et al. [15] among students, and then replicated among professionals [16]. Consequently, the external validity of our study is increased because it was conducted among senior students who have some internship experience in industry.

4 Data Acquisition Approach

The ATS (Activity Time Slip) approach, which is used in these field studies, is an augmented work diary approach focusing on the activity instead of the task being performed [17, 18]. The meaning of *work* (as in "work diary") is different from that of *activity* (as in the ATS). On the one hand, work is related to a task, and is often part of a schedule and is related to project resources. On the other hand, an activity is a personal endeavor undertaken while a developer is executing a task. Examples of activities reported on an ATS are: browsing the Web, reading about an API, talking to teammate about a concern, etc.

The ATS approach requires that the developer log, in an ATS token, the details of every activity performed. Table 1 gives an example of ATS token fields. Each ATS entry takes into account activities that may last more than an hour or only few minutes and records the teammates who are involved.

Each developer uses a preformatted spreadsheet to detail activities on an ongoing basis, at the rate of roughly one entry per hour. The ATS approach was applied throughout the entire duration of the projects.

Table 1. Activity Time Slip (ATS) token content

Field	Description	Example
ID	Unique token identifier	75
Date	Activity date	2012-05-05
Start time	Activity start time	9:20
End time	Activity end time	10:15
Effort	Activity duration (computed from the start/end time fields)	55
P_1 .. P_n	P_1 to P_n participants involved in executing the activity	GL, PN
Input artifact	Main input artifact of the activity	SRS
Output artifact	Main output artifact of the activity	CPA
Activity description	Detailed description of the activity	Use-Case Realization A,B,C,S
Process	Process discipline related to the activity	REQ
Task	Prescribed task	Modeling interface

Table 2 presents the total software development effort in hours, the number of tokens, and the tokens-per-hour ratio for each of the four projects.

Table 2. Software development effort and tokens

Project	Effort (Hrs)	Number of tokens	Tokens/hour
P06	997	1426	1.4
P07	750	1408	1.9
P08	810	887	1.1
P09	628	621	1.0

In order to extract knowledge behavior from self-reported developer activities, a coding scheme, based on the information source model, has been designed. An ATS token is codified according to the cognitive factor concerned. However, some tokens involve more than one cognitive factor. In this case, the coder needs to determine the dominant cognitive factor, mainly based on the description of the token and its context (input artifact, process, etc.). For instance, fixing a code defect involves both the verification and realization cognitive factors. First, it requires locating the defect in the code, which is related to the verification cognitive factor. Then, the actual fixing of the code involves the realization cognitive factor. In this situation, the dominant cognitive factor remains verification.

All the tokens of the four projects were codified by two independent coders, who had to decide which cognitive factor was dominant. However, tokens related to academic and technical activities were not accounted for in the codification, since they were not specific to project development. Academic activities are related to the lectures given by the instructors or presentations by the students, such as teamwork

training and project presentation. Technical activities are related to tasks which can be performed by technicians, such as configuring the network or setting up and maintaining the development environment.

5 Knowledge Acquisition in Software Projects

Every project has three phases, ending with a milestone. Many development artifacts are produced during the first phase, such as SRS, use case documentation, architecture and design documentation. The first milestone requires the development team to present their system architecture to the industrial partner. This occurs between 20% and 35% of project completion, depending on the project. Most of the system is coded during the second phase. The second milestone requires the team to package its application for acceptance testing. This occurs between 85% and 90% of project completion. Integration and acceptance testing are performed during this third phase. The third milestone occurs at the end of the semester, when the product is delivered to the client.

We recall that the four projects had the same industrial client, used the same disciplined process, and developed similar but different avionics applications. All participants had similar background and experience, but the teams were different for each project.

Fig. 2 to 5 show the total effort expended on each cognitive factor in relation to the four project completion. Each of the 6 curves of the graphs represents the relative total effort expended (Y-axis) for a given cognitive factor with respect to the percentage of project completion (X-axis). For example, in P06 (Fig. 2), at 30% of project completion (X-axis), 14% of the total effort (Y-axis) had been expended on crystallization.

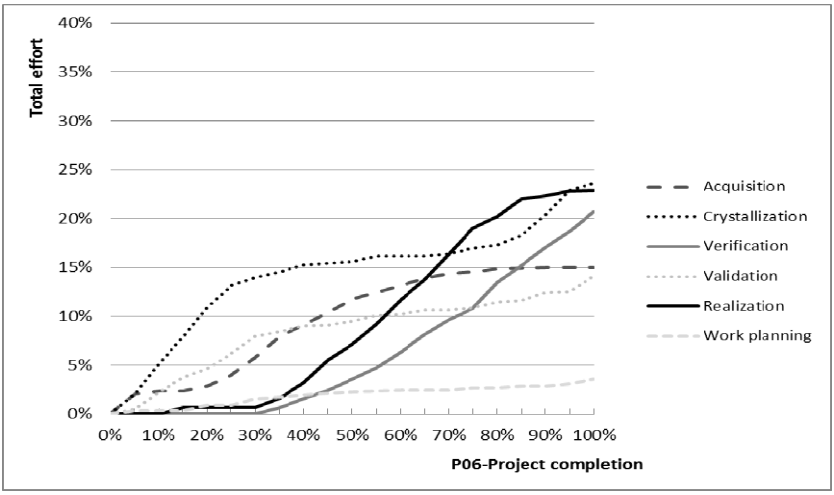


Fig. 2. Project P06 total effort distribution (Add graphical interface)

Validation represented 9% of the total effort, acquisition 6%, and realization and work planning both accounted for 1%. No verification effort had been expended to that point. Analysis of the slopes of the 6 curves in Fig. 2 provides a better understanding of the relationships between cognitive factors throughout the project. At project completion (X-axis 100%) we see that acquisition accounts for 15% of the total effort. In project P06 (Fig. 2), the learning activity (acquisition) is following the realization activity and the new knowledge is documented (crystallization) at the end of the project. Half of the learning effort (acquisition) was expended during the coding phase. Team members learned as they develop the product. There is as much total effort in realization activity as in crystallization activity.

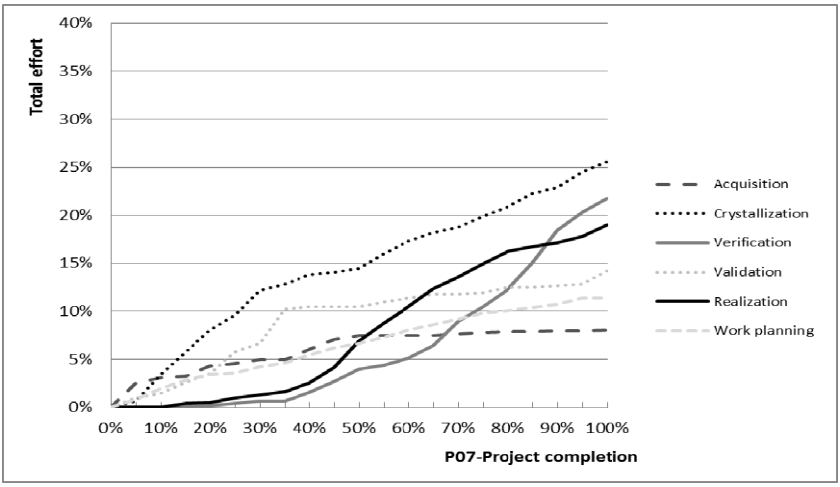


Fig. 3. Project P07 total effort distribution (Doxygen grouping)

In project P07 (Fig. 3) most of the acquisition occurred at the beginning of the project, during the first phase, and there is almost no acquisition during the coding activities. The learning behavior of this team was different in many respects. It is the project that required the least total learning activity (8%) and there is no learning during the coding phase. The cognitive factor requiring the most total effort was crystallization, followed by verification and realization. This cognitive behavior is related to the low level of technical difficulty of the project.

In project P08 (Fig. 4), the team expended as much total effort in documenting (crystallization) as in testing (verification). As in P06, half of the learning effort (acquisition) was expended during the coding phase. This new learning was documented as it is shown by the crystallization curve that followed the acquisition curve.

In the project P09 (Fig. 5), which is clearly coding oriented (realization), a minimum of acquisition activity occurred at the beginning of the project, and this produced documentation that would be revisited only during the third phase. The realization phase began early in the project (at 20% of project completion). However,

the acquisition activity was maintained throughout the coding activities. At 60% of project completion, the team needed to learn more about the project. However, part of this new knowledge was updated (crystallization) only at the very end of the project.

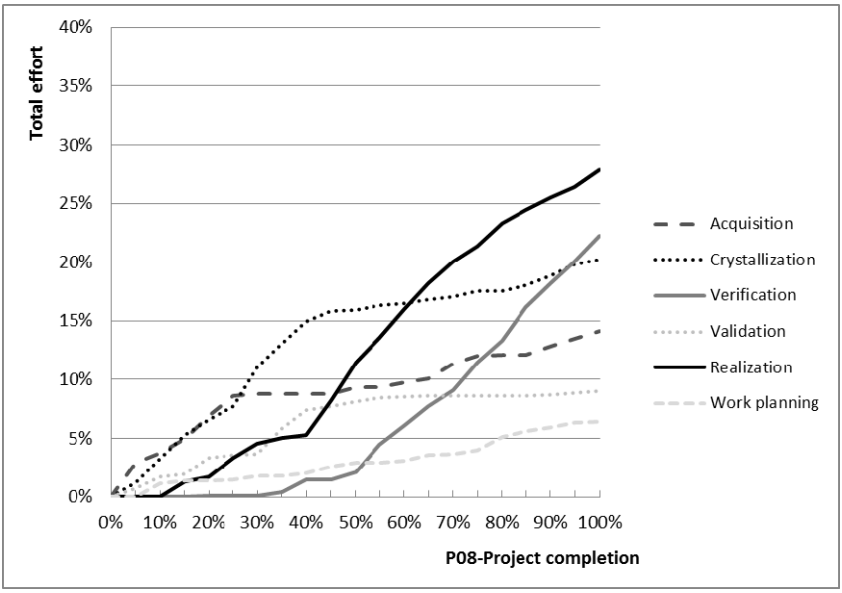


Fig. 4. Project P08 total effort distribution (OCR from video)

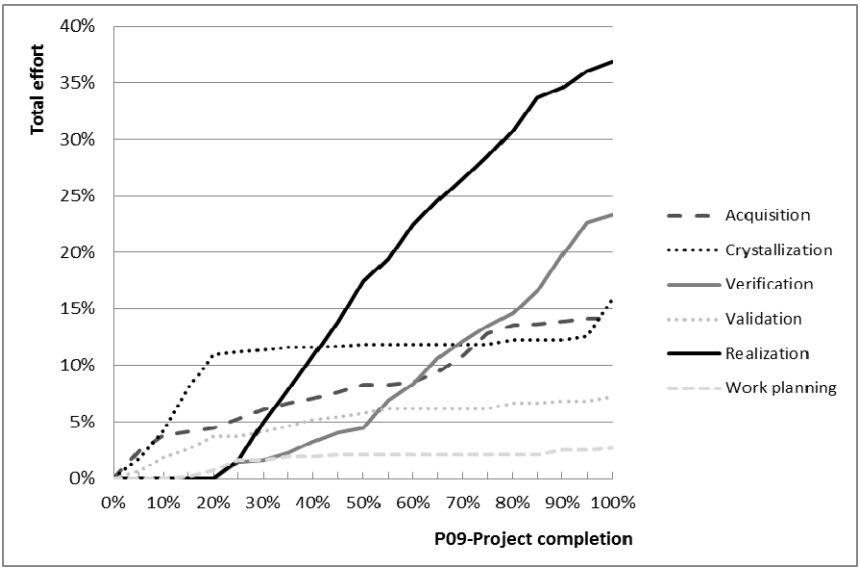


Fig. 5. Project P09 total effort distribution (Home-made parser)

Preliminary results from these field studies seem to indicate that the acquisition activity has an impact on the coding effort. For example, we observed that all the projects but one (P07) required around 15% of the acquisition effort, and that the later in the project the acquisition activity occurred, the greater the coding effort (see P08 and P09).

It seems that the need for more information or more learning emerged from activities performed during the coding phase. This could have an impact on the efficiency of Agile philosophy development, for example. Both projects P06 and P08 involved reusing software components. In P06, the team had to enhance a system by adding a graphical interface, while in P08, OCR components were used to build the required system. From a knowledge acquisition standpoint, it is noteworthy that in both cases, half of the acquisition was expended during the coding phase, even though a disciplined software process was used. This tends to confirm that each cognitive factor (acquisition, crystallization, validation, realization, and verification) is important throughout to whole project duration, as promoted in the Agile philosophy.

6 Concluding Remarks

There is a growing need to consider the knowledge perspective in software development, since developers' activities are mostly cognitive. Such knowledge cannot be measured directly, since it is mostly tacit, that is, it resides in the developer's mind. However we can measure the activities that lead to learning.

The ATS approach presented in this paper is a compromise between the acquisition of very accurate (think-aloud) data on participant cognitive activities in a short time and that of self-reported data on these activities over the duration of a project. The level of accuracy obtained with the ATS approach is sufficient to explore various knowledge acquisition perspectives in software development.

We find that the acquisition of information, which is learning, leads to improved knowledge and requires almost 15% of the total team effort. We also find that this effort, in most cases, cannot be restricted to the first phase of the project during requirement and architectural design, which is expected from waterfall or disciplined processes. About half of the learning is done during the code implementation even within a disciplined process. This finding is in line with one value of the Agile philosophy that promotes team interactions and users involvement for the whole project duration.

Future works will involve repeating these field studies with similar projects but with Agile teams. It could be interesting to see how the information acquisition patterns are modified with these new software development approaches.

Project managers should be aware that learning is an important component of software development and they should provide the social and physical environments to facilitate these learning activities.

Acknowledgments. This research was supported in part by NSERC grant A-0141.

References

1. Henninger, S.: Tools Supporting the Creation and Evolution of Software Development Knowledge. In: International Conference on Automated Software Engineering (ASE 1997), pp. 46–53 (1997)
2. Robillard, P.N.: The Role of Knowledge in Software Development. Commun. ACM 42(1), 87–92 (1999)
3. Xu, S., Rajlich, V., Marcus, A.: An Empirical Study of Programmer Learning During Incremental Software Development. In: 4th IEEE International Conference on Cognitive Informatics, pp. 340–349 (2005)
4. Bjornson, F.O., Dingsoyr, T.: Knowledge Management in Software Engineering: A Systematic Review of Studied Concepts, Findings and Research Methods Used. Inf. Softw. Technol. 50(11), 1055–1068 (2008)
5. Ko, A.J., DeLine, R., Venolia, G.: Information Needs in Collocated Software Development Teams. In: 29th International Conference on Software Engineering, pp. 344–353 (2007)
6. Kolb, D.: Experiential Learning: Experience as the Source of Learning and Development. Prentice Hall, Englewood Cliffs (1984)
7. Argyris, C., Schon, D.A.: Organizational Learning: A Theory of Action Perspective. Addison-Wesley, Reading (1978)
8. Wenger, E.: Communities of Practice: Learning, Meaning and Identity. Cambridge University Press, Cambridge (1998)
9. Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company – How Japanese Companies Create the Dynamics of Innovation. Oxford University Press, Oxford (1995)
10. Gendreau, O., Robillard, P.N.: Exploring Knowledge Flow in Software Project Development. In: International Conference on Information, Process, and Knowledge Management (eKNOW 2009), pp. 99–104 (2009)
11. Doxygen Manual, http://www.doxygen.org
12. WPF Tutorial.net, http://www.wpftutorial.net
13. Carver, J., Jaccheri, L., Morasca, S., Shull, F.: Issues in Using Students in Empirical Studies in Software Engineering Education. In: Ninth International Software Metrics Symposium (METRICS 2003), pp. 239–249 (2003)
14. Höst, M., Regnell, B., Wohlin, C.: Using Students as Subjects – A Comparative Study of Students and Professionals in Lead-Time Impact Assessment. Empir. Softw. Eng. 5(3), 201–214 (2000)
15. Porter, A., Votta, L., Basili, V.R.: Comparing Detection Methods for Software Requirements Inspection: A replicated experiment. IEEE Trans. Softw. Eng. 21(6), 563–575 (1995)
16. Porter, A., Votta, L.: Comparing detection methods for software requirements inspection: A replication using professional subjects. Empir. Softw. Eng. 3(4), 355–380 (1998)
17. Germain, E., Robillard, P.N.: Engineering-Based Processes and Agile Methodologies for Software Development: A comparative case study. J. Syst. Softw. 75(1-2), 17–27 (2005)
18. Gendreau, O., Robillard, P.N.: Knowledge Conversion in Software Development. In: Nineteenth International Conference on Software Engineering and Knowledge Engineering (SEKE 2007), pp. 392–395 (2007)

An Electronic Learning System for Integrating Knowledge Management and Alumni Systems

Amnart Pohthong[1] and Panumporn Trakooldit[2]

[1] Information and Communication Technology Programme
[2] Software Engineering and Applications(SEA) Group, Department of Computer Science, Faculty of Science, Prince of Songkla University, Hat Yai, Thailand
amnart.p@psu.ac, pnon6ster@gmail.com

Abstract. Nowadays, most people worldwide accept the fact that knowledge is a very valuable asset for their success. Educational institutions such as universities, whose main objective is to educate students to gain more knowledge and skills for their careers as well as how to adapt and live in their community. Over the past decades, teaching and learning paradigms have rapidly changed from traditional styles to computer-based styles. Electronic learning (e-learning) has been recognized as an effective computer-based technology for knowledge management, life long learning, and distant learning. Although most universities have created electronic alumni networks and systems, those systems lack the concern of knowledge management. Therefore, in this paper, we propose an electronic learning system for integrating knowledge management and alumni systems.

Keywords: E-learning, Knowledge management, Alumni system.

1 Introduction

Nowadays, universities and colleges have become knowledge-intensive organizations. Teaching and learning processes in educational institutes are expected to educate their students to meet all requirements stated in their curricula. When these students complete their studies and become graduates, they bring their knowledge to apply in their careers or further studies, as well as in their lives in their community. M. Shih, J. Feng, and C-C Tsai also addressed learning and teaching perspectives from *South Africa's Draft white paper on education: transforming learning and teaching through ICT (2003)* and from the British government-commissioned report on *Teaching and Learning 2020* in [1] about the importance of developing learners' critical thinking, decision-making, problem-solving skills through collaborative learning environments and the focus on the differences of each individual learner.

Self development and individual learning would increase graduates' knowledge. However, the results of their knowledge improvement and competency depend on each individual potential and opportunity. Some alumni still need to be supported by their previous institutions. Therefore, knowledge management (KM) becomes a mechanism for knowledge sharing, especially in subject areas where knowledge

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 11–21.
DOI: 10.1007/978-3-642-36981-0_2

changes rapidly such as Computer Science (CS) and Information Technology (IT). Knowledge management has also emerged from IT supports [2, 3].

Although most universities have developed their on-line alumni systems in order to link and communicate with their alumni, it still lacks the KM integration to those alumni systems. Therefore, the research reported in this paper proposes the framework for integrating knowledge management into the alumni system. The prototyped system can be used as an electronic learning system for knowledge sharing among alumni, current students, and institutional staff.

2 Electronic Learning and Knowledge Management

2.1 Electronic Learning

E-learning technologies have been widely used to deliver information, including learning materials, electronically from instructors to learners, especially for education and training. The results of the study reported in [4] confirm that instructors are willing to use e-learning environments to aid their teaching activities while learners also respond favorably to e-learning environments for complementing to their learning activities. Some critical factors influencing learner satisfaction for e-learning are reported in [5]. These factors can be grouped into six dimensions: learner, instructor, course, technology, design, and environmental dimensions. Among thirteen factors, seven factors were identified to be critical factors: learner computer anxiety, instructor attitude toward e-learning, e-learning course flexibility, e-learning course quality, perceived usefulness, perceived ease of use, and diversity in assessments. J. Andrade, J. Ares, R. Garcia, and S. Rodriguez suggested the three main blocks for organizing the elaboration process of e-learning actions: didactical material, follow-up and tutoring, and alternative learning [6]. J. Ismail also suggested the critical components for designing an e-learning system in [7]: learning management system, learning content design system, learning content management system, and learning support system.

2.2 Knowledge Management

Over the past decades, most universities worldwide have focused on data, information and knowledge as the important factors for their businesses. R.D. Corbin, C.B. Dunbar, and Q. Zhu highlighted in [8] that information results from the collection and assembly of “facts (data)” while knowledge involves the human intelligence traits. The terms of data and information management seem to gain better understanding than knowledge management (KM). KM is a systematic approach for capturing and creating, storing and accessing, validating, disseminating and applying knowledge to accomplish organizational goals and objectives [3]. Hence, most organizations expect that KM can lead to competitive advantages. KM in organizations relies on many systems and processes. In [9], Debowski suggests three types of organizational infrastructure for KM: managerial, technological, and social infrastructures.

KM activities can be addressed in various perspectives such as design, IT, management, artificial intelligence, and ontology perspectives [10]. Among several activities suggested in these perspectives, knowledge sharing has gained public attention in order to value and make use of both tacit and explicit knowledge as well as individual and organizational knowledge. Tacit knowledge can be transformed to be explicit knowledge and vice versa. However, knowledge sharing becomes more difficult than information sharing. Community of Practice (CoP) is one effective mechanism used to share knowledge among practitioners within the same organization or across different organizations.

Nowadays, IT tools have emerged to support knowledge sharing more efficiently than in the past, especially within and between CoP[11,12]. Many innovative technologies have been introduced and adopted to leverage KM such as knowledge portals [13,14] and e-learning [15,16].

3 System Framework and Implementation

3.1 Case Study

The Department of Computer Science at Prince of Songkla University in Thailand was selected as a case study for employing the proposed system. The department firstly recruited graduate students in 1986 and undergrad students in 1991. At present, there are around 1400 alumni. The current alumni system has only news and photo galleries.

3.2 Alumni System Components

The requirements for an alumni system are common to many universities, for example, the following alumni systems appear on their universities' web sites.

- Harvard medical school alumni system consists of news, events, alumni council, community, alumni benefits, and giving [17].
- Harvard business school alumni system consists of career development, reunions, travel, events, boards & volunteers, giving to HBS, FAQs, Clubs, Bulletin, and tools [18].
- MIT alumni system consists of alumni association, networks, benefits & services, volunteering, learn, travel, news & views, MIT students, parents association, and giving to MIT [19].
- University of Cambridge, UK., the Cambridge alumni relations office (CARO)'s main page consists of news, Olympics 2012, alumni events, alumni groups, alumni benefits, travel programme, and contact us [20].

3.3 Knowledge Management Components

The components of web-based knowledge management system suggested by Debowski in [9] are as follows:

(1) Business process management
(2) Content management
(3) Web content management
(4) Knowledge applications management

3.4 System Analysis and Proposed Framework

From the basic requirements for an alumni system and the components of web-based knowledge management, more requirements were gathered during the system analysis phase from the following groups of related users at the case study site.

(1) Alumni
Alumni can access provided courses, share learning media and knowledge. However, these media and knowledge will be evaluated by the departmental executive or committee before being deposited into a repository. Alumni can also manage some attributes of their personal data such as a degree earned and work experience.
(2) Students
Students can access courses provided on the departmental virtual class room (VCR) or those provided on the university learning management system called LMS@PSU. They can also access the knowledge repository via the knowledge sharing space.
(3) Support staff
Support staff can manage alumni profiles, news and activities, the departmental web-board and social networks.
(4) Teachers/Lecturers
In addition to manage their courses on the VCR or LMS@PSU, teachers or lecturers can manage courses for alumni, evaluate shared media and knowledge.
(5) Executive/Committee
The departmental executives and student affair committee can do the same activities similar to teachers and lecturers. They can also access some reports or some graphs of students and alumni statistics.

The main activities for the proposed system are specified by using the use-case diagram as shown in Fig.1. From information analysis, the proposed framework for an integrating KM and alumni system via an e-learning system was designed into six main parts as shown in Fig.2.

(1) Virtual class room (VCR)/Learning Management System (LMS)
(2) Course Materials for Alumni
(3) Alumni Competency and Learning Material Evaluation
(4) Knowledge Sharing Space
(5) Knowledge Repository
(6) Alumni Profile / Personal Profile

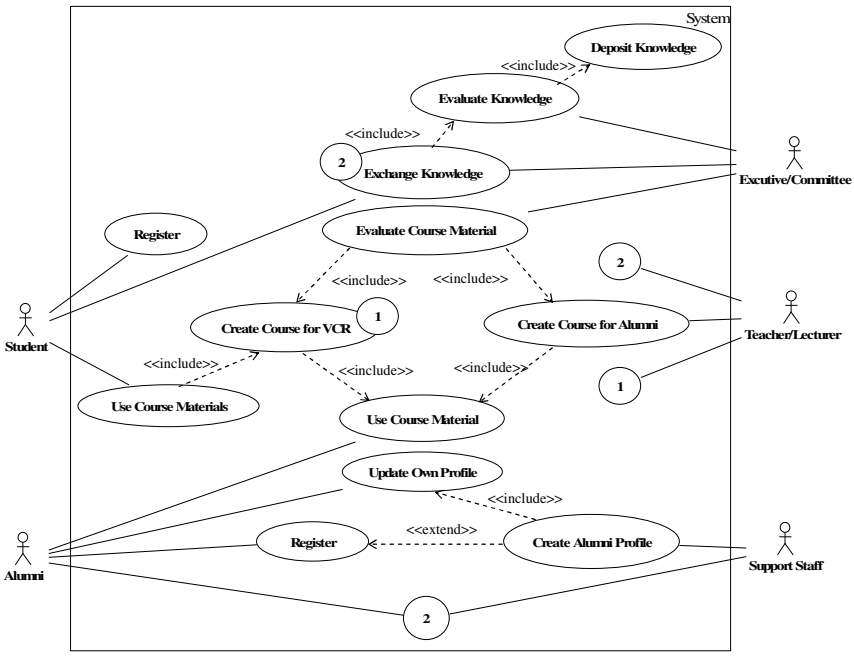


Fig. 1. The use-case diagram for the proposed system

3.5 System Design and Implementation

From the proposed framework, the system architecture was designed as shown in Fig.3. It consists of thirteen components: (1) *Membership* for managing the system users, (2) *AlumniData* for managing alumni data and profiles, (3) *Associate* for managing alumni groups, (4) *CourseforAlumni* for managing courses, (5) *Materials* for managing shared materials, (6) *Knowledge* for managing shared knowledge, (7) *KnowledgeEvaluation* for managing knowledge evaluation, (8) *Repository* for managing knowledge repository, (9) *YearStat* for managing statistical reports, (10) *News* for managing news, (11) *Event* for managing events, (12) *Webboard* for managing alumni webboard, and (13) *SocialNetwork* for managing social networks. The system users were managed into two main groups: members and non-members. The system members consist of alumni, current students, support staff, teachers/lecturers, executive/committee, and system administrators while non-members are end-users. The system database and its relational schemas were designed using Entity-Relationship (E-R) technique as shown in Fig.4.

The proposed system was prototyped as a web-based system. Some examples of user interfaces in the prototyped system such as the main menu, and the menu of knowledge submission by alumni are shown as Fig.5-6.

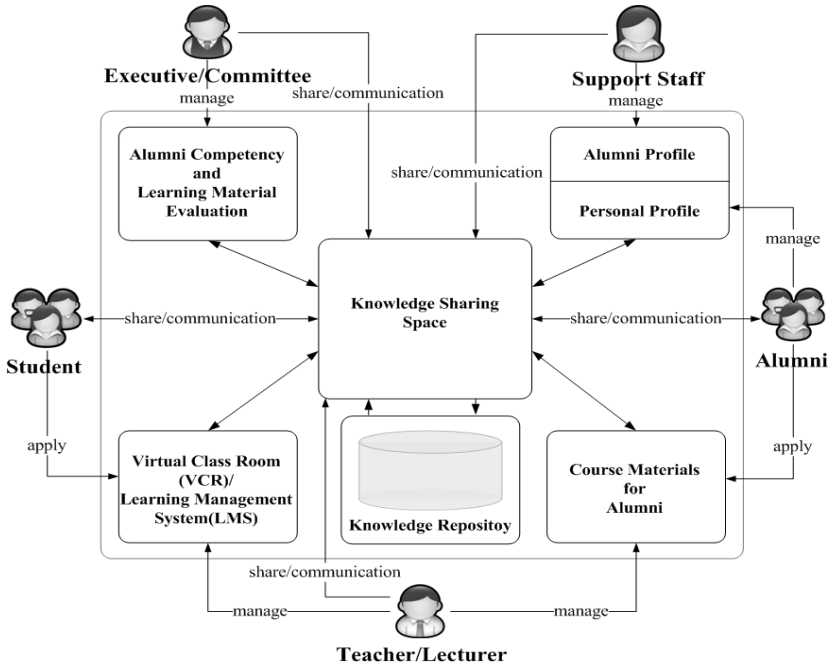


Fig. 2. The proposed framework for the prototyped system

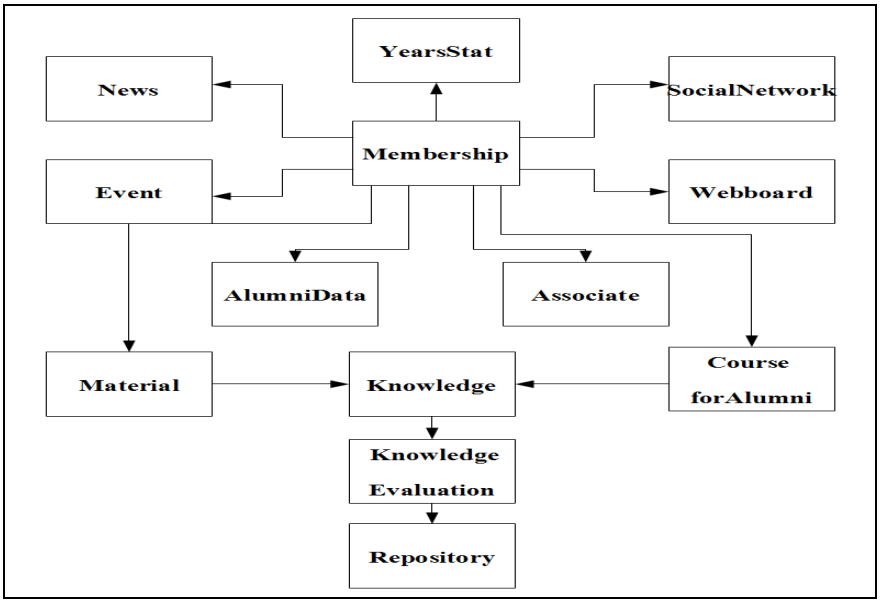


Fig. 3. The system architecture

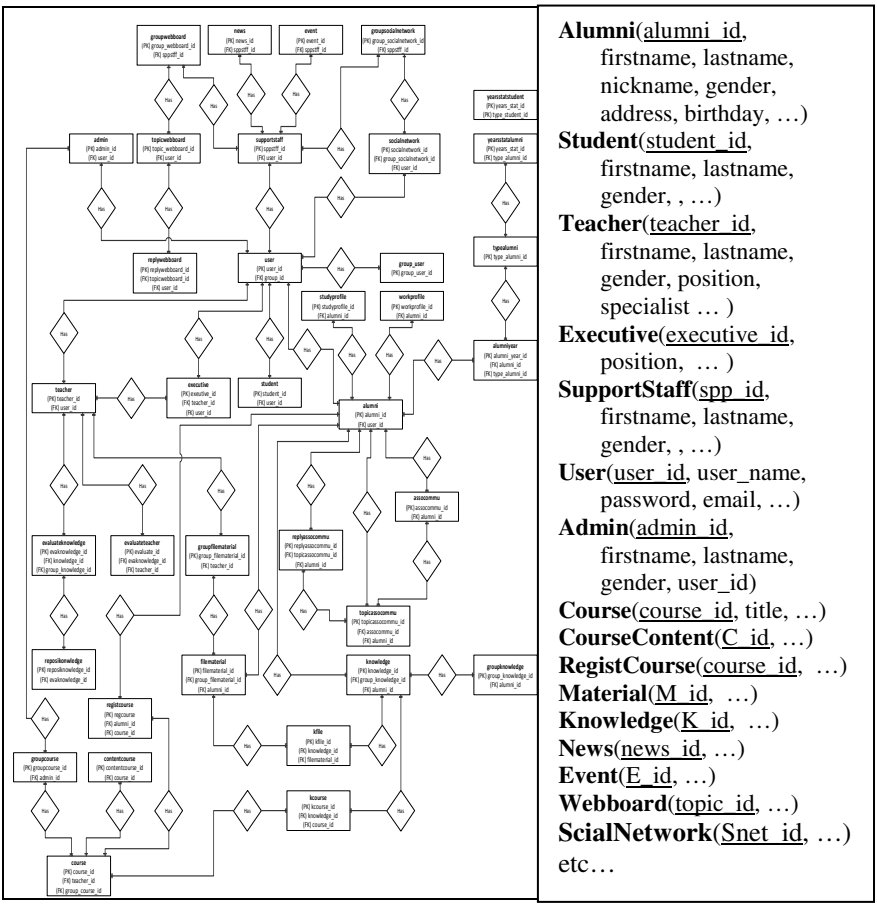


Fig. 4. The database structure for the system

4 System Evaluation

The system was evaluated in our laboratory for users' satisfaction. Thirty-five subjects were voluntarily selected. They consisted of one departmental executive who was in charge with the deputy head of department for students' affair, three lecturers, thirty alumni, and one support staff who was in charge of the departmental alumni system.

The evaluation tools were (1) the prototyped system (2) an introductory tutorial for the system (3) the given problems and (4) the questionnaire. The given problems corresponded to the type of users and consisted of several questions, allowed the subjects to explore and use the most related features of the system. The questionnaire consisted of three parts: personal data, questions for rating satisfaction level using Likert scale (from 1=very low to 5=very good), and open suggestions. Content validity in the questionnaire was evaluated by three experts and then calculated the

index of item-objective congruence (IOC). Questions that have the IOC value from 0.6 to 1.00 were included in the questionnaire. The revised questionnaire was used with five dry-run subjects in order to evaluate its reliability, using the Alpha-Coefficient method. The Alpha-Coefficient value was 0.91, showing that the questionnaire has good reliability.

The evaluation procedures started with giving a brief introduction about the system and its evaluation method as in the introductory tutorial for the system, taking around 10 minutes. Then, the subjects performed the given problems. Finally, all subjects completed the questionnaires.

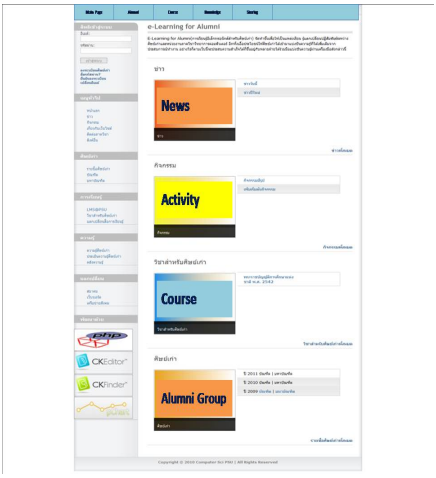


Fig. 5. The system main menu

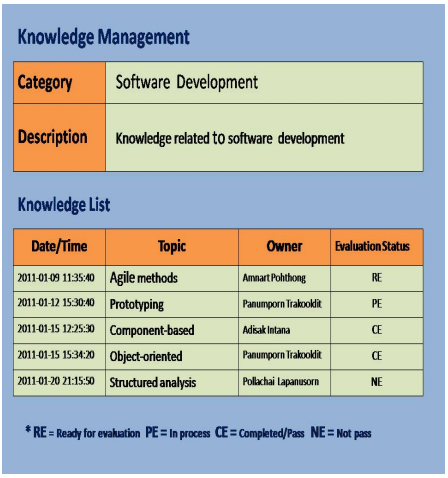


Fig. 6. The menu of knowledge submission by alumni

The results of system evaluation performed by two main groups of subjects: (1) members of staff (consisting of teachers/lectures, executive/committee, and support staff), and (2) alumni, are described as shown in Table 1 and Table 2.

Table 1. The result of each evaluation item performed by the departmental members of staff

Evaluation Items	Average	SD	Satisfaction Level
1. System Usage			
1.1 Convenience of data entry	3.2	0.55	Moderate
1.2 Convenience of data modification	3.2	0.45	Moderate
1.3 Automatic prevention of input errors	3.0	0	Moderate
1.4 Search support	3.4	0.55	Moderate
1.5 Appropriation of presentation sequence	3.4	0.55	Moderate
1.6 Completeness of requirements	3.0	0.70	Moderate
1.7 Appropriation of screen design	3.8	0.84	Good
1.8 Clarity of communication	3.4	0.89	Moderate

Table 1. (*continued*)

2. Efficiency			
2.1 Accuracy of data processing	3.2	0.84	Moderate
2.2 Accuracy of data retrieval	3.4	0.89	Moderate
2.3 System sustainability	3.0	0.71	Moderate
2.4 Efficiency of data storage	3.6	0.55	Good
3. Access ability			
3.1 Coverage of users	4.0	0.71	Good
3.2 Security	3.8	0.84	Good
4. System Value			
4.1 Learning support	3.8	0.84	Good
4.2 Organizational development	3.4	0.55	Moderate
4.3 Knowledge resource	3.6	0.89	Good
4.4 Knowledge sharing	3.6	0.89	Good
5. Overall Satisfaction	3.4	0.55	Moderate

Table 2. The result of each evaluation item performed by the departmental alumni

Evaluation Items	Average	SD	Satisfaction Level
1. System Usage			
1.1 Convenience of data entry	4.5	0.51	Very Good
1.2 Convenience of data modification	3.6	0.55	Good
1.3 Automatic prevention of input errors	3.4	0.72	Moderate
1.4 Search support	3.6	0.57	Good
1.5 Appropriation of presentation sequence	3.5	0.51	Good
1.6 Completeness of requirements	3.6	0.63	Good
1.7 Appropriation of screen design	3.6	0.57	Good
1.8 Clarity of communication	3.5	0.58	Good
2. Efficiency			
2.1 Accuracy of data processing	4.5	0.51	Very Good
2.2 Accuracy of data retrieval	3.5	0.57	Good
2.3 System sustainability	3.6	0.57	Good
2.4 Efficiency of data storage	3.6	0.57	Good
3. Access ability			
3.1 Coverage of users	3.5	0.57	Good
3.2 Security	3.6	0.57	Good
4. System Value			
4.1 Learning support	3.4	0.57	Moderate
4.2 Organizational development	4.5	0.51	Very Good
4.3 Knowledge resource	4.5	0.51	Very Good
4.4 Knowledge sharing	3.8	0.68	Good
5. Overall Satisfaction	3.5	0.51	Good

5 Conclusion

An electronic learning system for alumni was proposed. This system allows alumni to learn new knowledge and share their knowledge with other groups of users, especially with the departmental executives and lecturers. The system was prototyped for the case study at the Department of Computer Science, Prince of Songkla University, Thailand. The system was evaluated by thirty-five subjects for users' satisfaction in four categories: system usage, efficiency, access ability, and system value. The overall results of users' satisfaction were rated as moderate quality by the members of staff and as good quality by alumni. In future, the proposed system should share students' data with the university database, since current students will eventually become alumni in order to evaluate alumni competency after their graduation.

Acknowledgments. We wish to express our sincere thanks to all experts and all who acted as our subjects in both the dry run and the experiment.

References

1. Shih, M., Feng, J., Tsai, C.-C.: Research and trends in the field of e-learning from 2001 to 2005: a content analysis of cognitive studies in selected journals. Computers & Education 51, 955–967 (2008)
2. Liao, S.-H.: Knowledge management technologies and applications – literature review from 1995 to 2002. Expert Systems with Applications 25, 155–164 (2003)
3. Carnes, W.B.: Knowledge mangement: a stste of the practice summary. In: IEEE 7th Human Factors Meeting, Scottsdale Arizona (2002)
4. Liaw, S.-S., Huang, H.-M., Chen, G.-D.: Surveying instructor and learner attitudes toward e-learning. Computers & Education 49, 1066–1080 (2007)
5. Sun, P.-C., Tsai, R.J., Finger, G., Chen, Y.-Y.: What drives a successful e-learning? An empirical investigation of the critical factors influencing learn satisfaction. Computers & Education 50, 1183–1202 (2008)
6. Andrade, J., Ares, J., Garcia, R., Rodriguez, S.: Guidelines for the development of e-learning systems by means of proactive questions. Computers & Education 51, 1510–1522 (2008)
7. Ismail, J.: The design of e-learning system beyond the hype. The Internet and Higher Education 4, 329–336 (2002)
8. Corbin, R.D., Dunbar, C.B., Zhu, Q.: A three-tier knowledge management scheme for software engineering support and innovation. The Journal of Systems and Software 80, 1494–1505 (2007)
9. Debowski, S.: Knowledge Management, John Wiley & Sons Australia, Ltd. (2006)
10. Lai, L.F.: A knowledge engineering approach to knowledge management. Information Sciences 177, 4072–4094 (2007)
11. Pan, S.L., Leidner, D.E.: Bridging communities of practice with information technology in pursuit of global knowledge sharing". Strategic Information Systems 12, 71–88 (2003)
12. Chua, A.: Knowledge management system architecture: a bridge between KM consultants and tecnologists. International Journal of Information Management 24, 87–98 (2004)

13. Baalen, P.V., Bloemhof-Ruwaard, J., Heck, E.V.: Knowledge sharing in an emerging network of practice: the role of a knowledge portal. European Management Journal 23(3), 300–314 (2005)
14. Fernandes, K.J., Raja, V., Austin, S.: Portals as a knowledge repository and transfer tool – VIZCon case study. Technovation 25, 1281–1289 (2005)
15. Chen, R.-S., Hsiang, C.-H.: A study on the critical success factors for corporations embarking on knowledge community-base e-learning. Information Sciences 177, 570–586 (2007)
16. Weller, M., Pegler, C., Mason, R.: Use of innovative technologies on an e-learning course. The Internet and Higher Education 8, 61–71 (2005)
17. Harvard Medical Alumni Association, http://alumni.hms.harvard.edu/
18. Harvard Business School Alumni, http://www.alumni.hbs.edu/
19. MIT Alumni Association, http://alum.mit.edu/
20. University of Cambridge, UK., alumni relations office (CARO), http://my.alumni.cam.ac.uk/

Knowledge Management Systems and Intellectual Capital Measurement in Portuguese Organizations: A Case Study

Mário Pinto

ESEIG, Polytechnic Institute of Porto, Portugal,
KMILT Research Group
mariopinto@eu.ipp.pt

Abstract. This paper presents the results of an exploratory study on knowledge management in Portuguese organizations. The study was based on a survey sent to one hundred of the main Portuguese organizations, in order to know their current practices relating knowledge management systems (KMS) usage and intellectual capital (IC) measurement. With this study, we attempted to understand what are the main tools used to support KM processes and activities in the organizations, and what metrics are pointed by organizations to measure their knowledge assets.

Keywords: knowledge management systems, intellectual capital, intangible assets, metrics.

1 Introduction

Knowledge and Knowledge management (KM) are increasingly recognized as a key driver to innovation, competitive advantage and future sustainability [1], [2], [3]. In the new economy, knowledge based resources can be considered the main source of value creation [4], [5]. The competitiveness of organizations as well as their ability to develop distinctive capabilities of its competitors, is closely related with their capacity to create, store, share and apply their knowledge assets [4], [6].

In this context, KMS play a role of increasing importance. These systems contribute to support organizational processes and activities, which enable the knowledge sharing and knowledge application across organizations [7]. KMS also increase communication and collaboration, promoting a culture of knowledge sharing, and managing knowledge as a crucial asset for the organization [23]. Despite its importance in the modern economy, these intangible assets are not yet clearly measured and reported. Measuring these intangible assets shows their impact in value creation and its benefits for organization [5], [7], [8]. According some authors, evaluating the economic impact of knowledge in organizations, i.e., the Intellectual Capital (IC) measurement is a key issue in KM [5], [9], [11].

The aim of this paper is to know what Portuguese organizations are doing in terms of KM practices, namely KMS usage and IC measurement. A survey was made with two main purposes: i) identify the KMS used by Portuguese organizations; ii) identify the metrics specified to measure the main components of IC. A brief literature review

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 23–32.
DOI: 10.1007/978-3-642-36981-0_3

about IC measurement and KMS is made in the second section of this paper, as the background of this study. Section three describes the research methodology used in this study, while section four shows the results obtained and presents a brief discussion of them. The fifth section provides some conclusions and draws some directions for future research.

2 Background

2.1 Intellectual Capital Measurement Models

Numerous definitions of IC have been proposed, focusing IC as knowledge that can be converted into value [9], intellectual material [10] or combined intangible assets which enable the company to function [11]. However, almost all definitions have three common elements [12]: i) intangibility; ii) knowledge that can create value; iii) effect of collective practice. From these perspectives it is possible to describe IC as intangible assets that may be used as a source of sustainable competitive advantage, creating wealth in organizations.

There is a general agreement that intangible assets may be decomposed in a set of components. Almost all authors refer to IC as consisting on a set of human, relationship and structural capital [9], [12], [13]:

- Human capital is concerned with individual capabilities, knowledge, skills, experience and abilities to solve problems. It represents the employee's competence, attitude and intellectual agility [14], [15]. Competences include skills and education, while attitude covers the behaviour of the employees. Intellectual agility enables to think on innovative solutions and to change practices in order to solve problems [12].
- Structural capital is concerned with systems, organizational processes, technologies, concepts and models of how business operate, databases, documents, patents, copyrights and other codified knowledge. According to Roos [16], structural capital is what remains in the company when employees go home at night.
- Relationship capital is concerned with alliances and relationships with customers, partners, suppliers, investors and communities. It also includes brand recognition, organization image and market position. The relationship capital represents the knowledge embedded and the value added from the relationships with other external entities [17].

While IC represents the intangible assets that brings competitive advantage and value creation to the organization, its measurement reflects the influence and the impact of these assets in the organization [9], [12], [14]. Measuring the knowledge value and their impact in the organizations is a growing area of interest in the KM field, which reflects the value added by knowledge to the organizations and enables to monitor the performance of the knowledge resources and KM activities [18].

According to Luthy and Williams [2], [8], [19] there are two general approaches for measuring IC and its main components:

- Direct Intellectual Capital Methods, which estimate the monetary value of IC by identifying its various components. Once these components are identified, they can be evaluated, either individually or as an aggregated coefficient. It represents an attempt to fill the gap between market and book value.
- Scorecard Methods, which identify the knowledge resources that bring value added. Metrics for measuring these knowledge resources are reported in scorecards or graphs, giving a more detailed picture of the value of knowledge in organization. No estimates are made of monetary value of IC.

Tables 1 summarizes a review of the IC measurement models grouping them according Williams classification [20]:

Table 1. IC measurement models review

IC Approaches	IC Measurement Model	Authors
Scorecard Methods	Skandia Navigator	Edvinsson e Malone
	Balanced Scorecard	Kaplan & Norton
	Intangible Assets Monitor	Sveiby
	Intelect Model	Euroforum
	Intellectual Capital Index	Roos & Edvinsson
	Nova Model	Camisón, Palácios et al.
	Intangible Value Framework	Allee
	IC Rating	Edvinsson
	Intellectual Capital Rating	Joia
	Heng Model	Heng
	Meritum Guidelines	Meritum Guidelines
	Danish Guidelines	Mouritzen & Bukh
	Value Chain Scoreboard	Lev
	Chen, Zhu & Xie Model	Chen, Zhu & Xie
	VAIC	Pullic
	Intellectus	IADE & CIC
Direct Intellectual Capital Methods	Technology Broker	Brooking
	Citation-Weighted Patents	Bontis
	Inclusive Valuation Methodology	M´Pherson & Pike
	Total Value Creation	Anderson & McLean
	The Value Explorer	Andriessen & Tissen
	The 4-Leaf Model	Leliaert, Candries et al.
	Value Added Intellectual Coefficient	Pullic

2.2 Knowledge Management Systems

KMS are systems developed with the purpose of supporting KM processes, namely knowledge creation, storage and retrieval, knowledge transfer and application, as well as the flows between them [7], [21]. These systems enable an environment that facilitates the creation of knowledge, its sharing and application, and also the

communication and collaboration among the organization employees. More than technological tools, the KMS could be viewed as virtual spaces that promote knowledge conversion between explicit and tacit dimensions of knowledge [24]. According to Nonaka [24], Knowledge conversion from one form to another occurs frequently and leads to the creation of new knowledge.

Not all KMS are based on technologies. A non-virtual community of practice or a face meeting are ways to create and share knowledge. However, nowadays almost all KMS are based in information technologies. The amount of knowledge that needs to be captured, stored and shared, the geographic distribution of people and the dynamic knowledge evolution make the use of technology a necessity [21].

Many authors have written about the use of different types of KMS [21], [22], [23], [25], [26]. The variety of classifications referred by these authors takes us to develop a systematization of KMS categories [15], regarding their addressed issues, capabilities and functionalities [4], [27]. Table 2 summarizes this categorization, presenting the KMS categories considered and their main functionalities:

Table 2. Knowledge management systems categorization

Categories	Main functionalities
Document management systems (knowledge repositories)	Document management; edition collaboration; versions control; documents sharing; support for all content types (text, audio, video, graphs, xml, web, etc.); searching and retrieval advanced mechanisms.
Knowledge maps	Categorizing and indexing knowledge in taxonomies; creating knowledge maps; pointing to organizational knowledge; inserting tags and labels in documents; alerting to relevant information.
Collaboration systems (groupware)	Synchronous or asynchronous communication; process and people collaboration; virtual meetings; instant messenger, videoconference; real-time conversation; grouping calendar and scheduling, etc.
Workflow systems	Business processes automation; support automated flows of activities, tasks and information; support documental flows.
Business intelligence and Data mining tools	Statistical, OLAP analysis; reveal patterns and hidden relationships between data; generate new knowledge from existing one; query and reporting tools; data mining and data warehousing tools.
Expert systems	Expert identification; connect users with experts to solve certain problems; ask questions, provide recommendations and explain logical processes; capture and store new questions and rules in a knowledge base.
Competence management	Employees profiles; experts, customers, vendors or others profiles in some systems; competence maps; individual competence analysis; training programs recommendation based on employees skills; recruitment and selection support.

Table 2. (*continued*)

E-learning systems	Environment personalization; evaluation and progress tracking: exercises quiz and tests; collaboration tools; reusable learning and object libraries; support different types of contents: text, audio, video, etc.; classes' workgroups; authoring, scheduling and reporting tools; searching and matching tutorials.
Help-desk systems	Self-desk and help-desk functionalities; FAQs access and maintenance; on-line customer support; expert help; customer profiles; customers queries.
Corporative portals	Environment personalization; filtering relevant information; search and retrieval advanced mechanisms; news, activities, tasks and calendar management; unified access environment to other tools: documents management, workflow, knowledge maps, groupware, etc.; integration with other applications.
Web 2.0 tools	Interaction, collaboration, participation of people: blogs, wikis, social bookmarking, tagging, platforms for content sharing.

3 Research Methodology

A survey was made with the purpose of knowing the current practices of the Portuguese organizations, regarding KMS usage and IC measurement. With this study we seek to understand which type of tools are most used by Portuguese organizations in supporting KM processes, and what metrics they generally use to measure knowledge resources.

The survey was based on a questionnaire, sent to one hundred of the main Portuguese organizations. The organizations were selected from a publication that produces an annual ranking of companies, based on their value creation for the Portuguese economy. The questionnaire was sent to the director of the knowledge management department or information systems department (when the first did not exist in the company). With the questionnaire one letter was also sent, explaining the concepts of KMS, IC and their main components: human, structural and relationship capital; also explaining the aims of the study, assuring confidentiality and requesting collaboration from the most suitable person in the organization. The questionnaire was structured in three main sections:

- Organization identification: it includes the organization name and business area.
- Knowledge management systems identification: It comprises the identification of KMS categories used in the organization. The questionnaire presents the eleven KMS categories described in table 2 and the organizations could select the adequate categories or add new ones.
- Intellectual capital metrics identification: It comprises the identification of the metrics used in the organization to measure IC and their components: human, structural and relationship capital. The questionnaire contains a comprehensive list of qualitative and quantitative metrics, resulting from an extensive review of IC measurement models [15]. However, the respondents could also complete this list, adding the metrics used in their organizations.

An extensive review of survey studies shows that some forms of follow-ups can increase response rates [12], [29]. According some authors [28] the resistance response rate is continuously increasing during the survey: it is relatively high at first, drops for a short period after the follow-ups and then starts to increase. Thus, two follow-ups were carried out using letters, telephone calls and e-mails. The delay between these two follow-ups was six weeks. Twenty-one valid questionnaires answers were received, corresponding to a response rate of 21%.

4 Results and Discussion

From the questionnaires received, six were from service organizations (e.g. telecommunications, energy) and fifteen were from industry organizations (e.g. automobile industry, electronic, pneumatics). All of the respondent organizations have identified a set of KMS, used to support knowledge processes, but only twelve organizations have specified a set of metrics to measure IC assets. The remaining organizations said that they didn't make a regular management and measurement of intangible assets.

4.1 Knowledge Management Systems

One issue addressed in the survey was the use of KMS in Portuguese organizations. Table 3 summarizes the several categories mentioned by respondents in the questionnaire, presenting the respective occurrence rate.

Table 3. Knowledge management systems usage in Portuguese organizations

KMS Categories	Rate
Business Intelligence and Data mining tools	67%
Knowledge Maps	25%
Document Management Systems (repositories)	75%
Collaboration Systems (groupware)	43%
Workflow Systems	50%
Expert Networks	13%
Competence Management Systems	55%
E-learning Systems	25%
Help-desk tools	75%
Corporative Portals	67%
Web 2.0 tools	75%

Based on the results presented above, some important conclusions can already be drawn:

- The findings presented in table 3 point that KMS supporting mainly explicit knowledge were most mentioned by respondents. Document management systems, business intelligence, competence management and help-desk systems are examples of tools that lead mainly with explicit to explicit conversion of knowledge, according to Nonaka knowledge conversion model [24].

- Expert systems, knowledge maps and e-learning systems were tools with a reduced response rate. According the questionnaire answers, these tools that mainly support personal knowledge and tacit to explicit conversion of knowledge, they have a low dissemination in Portuguese organizations.
- Help-desk tools are also strongly referred in the survey. The relationship with customers and other external entities is crucial to obtain competitive advantages in a global economy. Thus, the result obtained shows the importance recognized to relationship capital and the need to satisfy the customer's needs.

It is interesting to note that none of the respondents have identified other categories of tools, beyond those mentioned in the questionnaire.

4.2 Intellectual Capital Measurement

Another issue addressed in this study was the metrics used by Portuguese organizations to measure IC. Table 4 summarizes the metrics mentioned by respondents, grouping them by IC component: human, relationship and structural capital.

Table 4. Summary of IC metrics survey

	Metrics	Metrics
Human Capital	▪ Training programs (days per year)	▪ Investment in training (per capita)
	▪ Employees in training plans (%)	▪ Employee turnover
	▪ Execution rate of annual training plan	▪ Value added per capita
	▪ Duration of training plans (average)	▪ Full-time employees (%)
	▪ Average level of academic degree	▪ Part-time employees (%)
	▪ Employees satisfaction (index)	▪ Specialized/expert employees (%)
	▪ Average duration of employees relationship	▪ Innovative employees (new ideas)
	▪ Age distribution of employees	▪ Employees with initiative (new ideas)
	▪ Absenteeism rate	▪ IT literacy skills (average)
	▪ Internships (number)	▪ Profits by employee
Relationship Capital	▪ Grow rate of customer's portfolio	▪ Investment Information Technologies
	▪ % of small, medium and large customers	▪ Investment in marketing
	▪ Profitability by costumer (average)	▪ New customers/customers lost (rate)
	▪ Bill per customer (average)	▪ Annual sales per customer
	▪ Customer's satisfaction index	▪ Market share in segment
	▪ Customers claims (number)	▪ Business alliances and partnerships
	▪ New contracts / proposals (rate)	▪ Average duration of customer relationship
	▪ Delay in delivery orders (average)	▪ Customers contacts (number)

Table 4. (*continued*)

	Metrics	Metrics
Structural Capital	▪ Quality certifications (number)	▪ New business generated
	▪ Processes in non-conformity	▪ Innovation and creativity capabilities (new products/services, upgrades)
	▪ Certified products (number)	▪ New products launched
	▪ Quality tests performed	▪ Revenue generated by new products / total revenue
	▪ Key-process documented	▪ Number of customers per employee
	▪ Continuous improvement projects	▪ Business partners (number)
	▪ Investment in developing new skills	▪ Computers per employee
	▪ Investment in training programs	▪ Knowledge management initiatives (#)
	▪ Investment in new products/services	▪ Protocols with Innovation entities (#)
	▪ Investment in new processes	▪ Suggestions from employees accepted by administration
	▪ Investment in Information Technologies	▪ Productivity index
	▪ Investment in R&D	▪ Time response to customers' requests
	▪ Administrative expense/employee	▪ Time to processing payments
	▪ Administrative expense/total revenues	

The results presented in the previous tables, allow us to draw the following considerations:

- The metrics focused on the human capital component are, mostly, related with the characterization of the employee's profile, and their effort in training programs. Few metrics are focused on measuring the value added by employees and their contribution to the organizational knowledge. Most of the metrics translate the effort invested in activities (training plans, for example) rather than the results obtained.
- The structural capital includes a few number of metrics for measuring research and development activities. The questionnaire answers do not include metrics for measure the innovative capability, for instance, the number of new products/services developed, the number of new ideas that generate new products or services and the number of new patents registered.
- Although Portuguese organizations define relationship capital as valuable relations with external entities, including customers, suppliers, partners, investors and other entities, they measure basically the customer capital. The metrics employed are almost all related with customers, ignoring the value of external relationships with, for instance, innovation entities, governmental departments, investors or business partners.

5 Conclusions

The KMS categories more mentioned by the respondents, namely business intelligence, document management systems (knowledge repositories), competence management and workflow systems, are focused on supporting explicit knowledge. Help-desk systems could offer support to both explicit and tacit knowledge. These results are in compliance with the IC metrics addressed in the survey questionnaire. Almost all metrics are aligned with the measurement of explicit knowledge, and they can be provided from the tools above mentioned. Almost all metrics referred by organizations to evaluate human capital, for example, could be found on competence management systems, which manage knowledge related with human competences and skills. On the other hand, a significant number of metrics pointed by organizations to measure relationship capital, could be obtained from explicit knowledge wrapped in help-desk or business intelligence systems. The results of the survey show that current practices relating KMS usage and IC measurement are in compliance, focusing mainly the management and the evaluation of the explicit knowledge, easier to represent, codify and share, than tacit knowledge.

The findings of this study do not allow us to obtain sustainable results. Unfortunately, a considerable number of organizations, in Portugal, still do not have a culture of manage their knowledge assets and haven't a knowledge management infrastructure based on KM tools. Future work comprises a more deeply study in some organizations that have answered to the questionnaire.

References

1. Bontis, N., Curado, C.: managing Intellectual capital: the MIC matrix. International Journal of Knowledge and Learning 3(2/3) (2007)
2. Leliaert, C., et al.: Identifying and managing IC: a new classification. Journal of Intellectual Capital 4(2), 202–214 (2003)
3. Ammann, E.: A Hierarchical Modelling Approach to Intellectual capital development. Electronic Journal of Knowledge Management 8(2), 181–192 (2010)
4. Pinto, M., Lopes, A., et al.: A Framework for Characterizing Knowledge Management Systems. In: 6th European Conference on Knowledge Management, Limerick, Ireland, pp. 442–450 (2005)
5. Cabrita, M., Vaz, J., Bontis, N.: Modelling the creation of value from intellectual capital: A Portuguese banking perspective. International Journal of Knowledge and Learning 3(2/3), 266–280 (2008)
6. Martí, J.: In Search of an Intellectual Capital General Theory. Electronic Journal of Knowledge Management 1(2), 213–226 (2003)
7. Alavi, M., Leidner, D.: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. MIS Quarterly 25(1), 107–136 (2001)
8. Fritzsche, A., Germany, C.: Implicit Evaluations of Intellectual Capital. Practical Decision Making 10(3), 236–243 (2012)
9. Edvinsson, L., Malone, M.: Intellectual Capital: Realizing your Company's True Value by finding its Hidden Brainpower. Harper Business, New York (1997)

10. Stewart, T.: Intellectual Capital - The New Wealth of Organization. Doubleday, New York (1997)
11. Brooking, A.: Intellectual Capital: Core Assets for the Third Millennium Enterprise. Thomson Business Press, London (1996)
12. Cabrita, M., Vaz, J.: Intellectual Capital and Value Creation: Evidence from the Portuguese Banking Industry. The Electronic Journal of Knowledge Management 4(1), 11–20 (2006)
13. Suciu, M., Piciorus, L., Imbriscã, I.: Intellectual Capital, trust, cultural traits and reputation in the Romanian education system. Electronic Journal of Knowledge Management 3, 223–235 (2012)
14. Leliaert, C., et al.: Identifying and managing IC: a new classification. Journal of Intellectual Capital 4(2), 202–214 (2004)
15. Pinto, M., Lopes, F., et al.: Knowledge Management Systems and Intellectual Capital Measurement. In: Proceedings of XVI ISPIM Annual Conference, Porto, Portugal (2005)
16. Roos, J., Dragonetti, N., et al.: Intellectual Capital: Navigating in the New Business Landscape. Macmillan, New York (1997)
17. Bontis, N., Keow, W., et al.: Intellectual capital and business performance in Malaysian industries. Journal of Intellectual Capital 1(1), 85–100 (2000)
18. Mouritzen, J.: Overview intellectual capital and the capital market: the Circulability of Intellectual Capital. Accounting, Auditing & Accountability Journal 16(1), 18–30 (2003)
19. Andriessen, D.: Making sense of intellectual capital designing a method for the valuation of intangibles. Butterworth-Heinemann, Oxford (2004)
20. Williams, M.: Is a company's intellectual capital performance and intellectual capital disclosure practices related? In: McMaster's Intellectual Capital Conference, Toronto, Canada (2000)
21. Lindvall, M., Rus, I., et al.: Software systems support for knowledge management. Journal of Knowledge Management 7(5), 137–150 (2003)
22. Baroni, R.: Aplicações de Softwares de Gestão do Conhecimento: tipologia e usos. Ciência da Informação. Blo. Horizonte, Brazil, UFMG University (2000)
23. Abdullah, R., Selamat, M.: A framework for knowledge management system implementation in collaborative environment for higher learning institution. Journal of Knowledge Management Practice 6 (2005)
24. Nonaka, I., Kono, N.: The Concept of BA: Building a Foundation for Knowledge Creation. California Management Review 40(3), 40–54 (1998)
25. Loureiro, J.: Gestão do Conhecimento. V. N. Famalicão, Centro Atlântico (2003)
26. Nantel, R.: Knowledge Management Tools and Technology 2004: 35 Systems to Maximize Your Organization´s Intellectual and Human Capital. Brandon- hall.com (2003)
27. Zhou, A., Fink, D.: The Intellectual Capital Web: a systematic linking of intellectual capital and knowledge management. Journal of Intellectual Capital 4(1), 34–48 (2003)
28. Fortin, M.: Le processus de la recherce: de la conception à la realization. Décarie Éditeur (1996)
29. Birch, M., Uccardi, M., et al.: Conduction Background Investigations. In: APTA, Standards Development Program, American Public Transportation Association (2011)
30. Bebensse, T., Helms, R., Spruit, M.: Exploring Web 2.0 Applications as a Mean of Bolstering up Knowledge Management. The Electronic Journal of Knowledge Management 9(1), 1–9 (2011)

Semantic Patent Information Retrieval and Management with OWL

Maria Bermudez-Edo, Manuel Noguera, José Luis Garrido, and María V. Hurtado

University of Granada. Department of Software Engineering, E.T.S.I.I.,
c/ Saucedo Aranda s/n, 18071 Granada, Spain
{mbe,mnoguera,jgarrido,mhurtado}@ugr.es

Abstract. Patent information is mainly represented and stored in databases containing large amounts of information about the inventions and metadata of patents such as the technological field to which they belong, which can be retrieved in standard formats such as CSV or XML. These, however, provide few semantics to enable further relationships among patents to be inferred for analysis purposes. Ontologies, mostly represented in the Web Ontology Language (OWL), are increasingly being developed to represent, manage and reason about data in information systems. Unfortunately, the current patent ontologies do not seem to fully capture the implicit hierarchies present in patent technology codes to exploit the information that can be derived from the formal representation of patent code classification hierarchies through logic reasoning. This paper presents an approach to automatically index hierarchical codes with ontological categories and enrich the information retrieved and knowledge management from different patent repositories with new relationships, properties and inferred information.

Keywords: Information and knowledge management, Ontology, OWL, XML, eXtensible Stylesheet Language Transformations (XSLT), Patent.

1 Introduction

Patents have a huge impact on national and international economies and represent a great part of all the scientific and technological knowledge worldwide [1]. Several studies have in fact used patents to measure the innovative capacity of firms [2] [3] and even technological trends [4] [5].

Patents are usually stored in large databases which belong to the different patent offices around the world, e.g. the European Patent Office (EPO) and the United States Patent and Trademark and Office (USPTO). While most of these databases have become available online in recent years, they exhibit different datasets and data structures for patent representation and this makes it difficult to automate their processing.

Patent metadata comprise different types of information such as the name of the patent applicant, publication date and the technological patent classification. Each patent database defines a set (or sets) of technological codes according to hierarchical

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 33–42.
DOI: 10.1007/978-3-642-36981-0_4

classifications which specify the technological fields a patent may pertain to or be associated with. These fields are widely used in database searches to discover, for instance, the field or fields in which a firm may infringe another company's industrial rights or existing gaps in a certain technology which a company could exploit [6].

When analysts need information about a firm's innovations (to find niche markets, to internationally extend its innovations, etc.), they can retrieve this information from patent databases in standard formats, such as Comma-Separated Values (CSV) or eXtensible Markup Language (XML) [7]. Since these formats lack formal semantics to enable the interrelation of patent information, it is also therefore difficult to share data from different databases. In this regard, an efficient retrieval and processing of patent information based on semantics could improve the information and knowledge management of patents [8].

Such semantics could be provided by ontology languages with a formal model-theoretic grounding. Ontologies have been increasingly developed to represent, manage and reason about information system data. Ontologies allow common vocabularies and relationships between domain entities to be defined [9]. The web ontology language (OWL) [10] with its formal semantics [11] based on description logics has become the de facto standard among ontology languages.

In the context of patent metadata, ontologies have already been developed [1] [12]. These ontologies have been populated by translating XML documents retrieved from patent databases into OWL ontologies. However, these approaches seem to disregard the hierarchical relationships between the technological fields into which a patent may be classified, thereby hindering the exploitation of information that can be derived from the formal representation of patent code classification hierarchies by means of logic reasoning. Another drawback is that they only represent patent data from a single patent office.

This paper proposes an approach which is explicitly intended or utilized for creating and processing knowledge about patents. This approach provides a practical mechanism to automatically build and populate patent metadata ontologies by indexing hierarchical codes, which can be retrieved from different patent repositories, and by defining ontological categories which enrich patent information management with new relationships, properties and enabling the inference of new knowledge.

An application study is presented in order to illustrate the applicability of the proposal in the information and knowledge management about firms' innovation, by means of a case study, that shows how to automatically infer information about the internationalization of environmental patents on the basis of the information provided in the metadata of patent documents.

The rest of the paper is organized in the following way: Section 2 studies related work; Section 3 shows the proposed method for translating hierarchical patent codes from XML files into hierarchies of concepts in OWL files, including the population of the ontology; Section 4 presents a case study that shows the benefits of arranging the patent technological field in a hierarchical manner in OWL; and Section 5 concludes the paper by discussing the contributions of the proposal.

2 Related Work

In the domain of patent ontologies, various efforts have been made to create ontologies with information retrieved from patent databases. The most relevant patent ontologies based on patent metadata and represented in OWL are Patexpert, which was created within the European Patexpert project [13], [1], and PatentOntology, which was developed at Stanford University [12]. Although these two ontologies are relevant and close to our field of interest, none of them reflect the structure of the technological codes and they do not allow fully exploiting the logical reasoning with technological codes.

Patexpert was created to homogeneously represent different patent information from several EPO databases and to provide this with semantic meaning. However, Patexpert does not merge information retrieved from different patent offices: the patent metadata ontology has been populated by XSLT (eXtensible Stylesheets Transformation Language)[14], stylesheets. Unfortunately, the public version of this ontology is not populated, but to the best of our knowledge this ontology does not automatically retrieve or represent the semantics of the hierarchy of technological codes.

PatentOntology was developed to avoid the limitations of Patexpert when integrating heterogeneous domains [8]. PatentOntology merges information from USPTO patent documents retrieved from the USPTO database with information from patent courts of USPTO retrieved from the LexisNexis database [15]. This ontology has been populated with a parser, but does not automatically retrieve the semantics of the hierarchy of technological codes neither merge information from different offices.

In other domains, there exist other proposals which have extracted OWL documents from XML documents by using XSLT, but none of them has dealt with the extraction and then indexing of hierarchical codes in ontological categories as our proposal does. While some of this work uses XML schema [16] [17], other work only creates the OWL model [18] and others the OWL model and instances [19] but none attempt to represent the code structure in an ontological categorization.

Other work has also been published on translating XML into OWL and the development of visual tools such as JXML2OWLMapper [20], [21] or the online XMLtoOWL tool [22]. These tools enable the visual assignment of XML labels of the XML instances or schema into OWL labels. However, neither of these tools combines information retrieved from different sources nor capture the semantics of the hierarchical organization of codes.

3 Semantic Information Retrieval from XML to OWL

The codes for patent technological fields exhibit a certain structure that needs to be identified before a transformation document is defined to translate patent metadata in XML into an OWL ontology. Furthermore, each patent office defines its own scheme

to represent patent metadata. In this section, we introduce an overview of the method proposed and describe the transformation process from XML to OWL by means of XSLT. We then propose that the stylesheets be customized in order to fully exploit the information gathered in the hierarchical codes. Finally, we show an example of the method for a particular hierarchical code.

3.1 Method for Transforming XML into OWL with Stylesheets

This paper presents a method for processing query results from different databases in XML format, and the XML files are converted into OWL by means of the corresponding XSL files (styleshhets) and an XSLT processor. Figure 1 shows an overview of the method proposed.

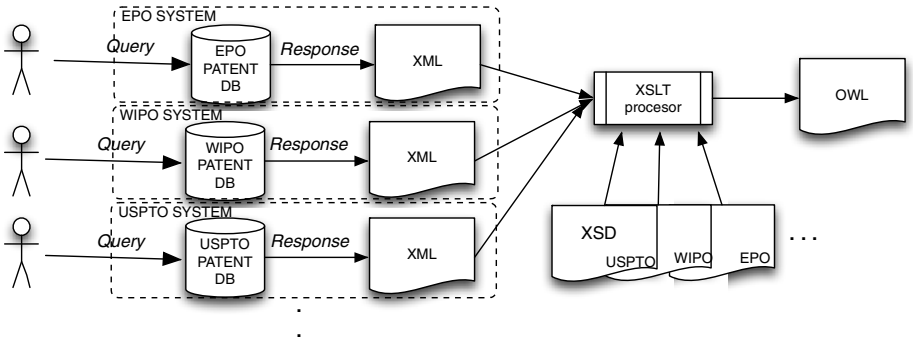


Fig. 1. Method Overview

The most important phase of this process is to create the stylesheet. Previous publications [12] [1] have used stylesheets to convert one node label of an XML document into a single class or property or instance of an OWL document.

In our proposal, we will automatically create an OWL model from the XML instance document for the technological fields of patents and OWL instances of this model. This method encompasses the translation of each instance of the technological code (a node label of the XML document) into a hierarchy of classes in the OWL document (several classes and one instance). For this purpose, it is necessary to divide the XML label into several parts and transform each into an OWL class or instance. The code is partitioned using XPath (XML Path language) [23].

The stylesheet developed in this work allows the ontology to be populated automatically with the hierarchical technological codes of the patents. Even when new codes appear, the stylesheet can create new classes and instances automatically without any further modification.

3.2 Structure of the Hierarchical Codes

Various patent technology codes exist such as the US classification or the European ECLA and ICO classifications or the International (IPC). The ICO and ECLA codes are based on the IPC codes and have the same underlying structure:

- Section, represented by one letter
- Class, represented by two digits
- Subclass, represented by one letter
- Main group, represented by one to three digits
- Group, represented by at least two digits

For instance, the ICO code Y02E10/20, which refers to energy generation, is made up of section (Y), class (02), subclass (E), main group (10) and group (20). As such, the complete code H04L25/02 means:

Y	new technological developments
Y02	technologies for mitigation of climate change
Y02E	reduction of greenhouse gases
Y02E10	energy generation for reduction of greenhouse gases
Y02E10/20	energy generation through renewable energy sources

Hierarchical codes have two main characteristics:

- Characteristic 1. Importance of the context of the code. The meaning of one part of the code depends on the meaning of the previous parts of the code. For example, the meaning of the part of the code "main group" (10) has different meanings if the previous part of the code is Y02E (Y02E10, energy generation for reduction of the greenhouse gases) or if the previous part of the code is Y02C (Y02C10, C02 capture or storage).
- Characteristic 2. Importance of parts of the code. The same representation of a part of the code (for example: 20) has different meanings depending on the part of the code to which it belongs. For example, the representation (20) has different meanings if it belongs to the "group" part of the code (Y02E10/20, energy generation through renewable energy sources) or to the "main group" part of the code (Y02E20, combustion technologies with mitigation potential).

Taking these two characteristics into account, the proposed method automatically creates an OWL ontology named HCOntology (Hierarchical Code Ontology) which represents the hierarchical codes with the full semantic meaning of the codes. The resulting HCOntology for the ICO codes with the tool Protégé [24] is shown in Figure 2. In OWL, since child classes inherit the meaning of the parent classes, HTCOntology therefore complies with the aforementioned Characteristic 1. Furthermore, HTCOntology represents different parts of the codes (for example the main group 10) with previous parts of the code (example ICO_Y02E10) and so HTCOntology fulfils the aforementioned Characteristic 2.

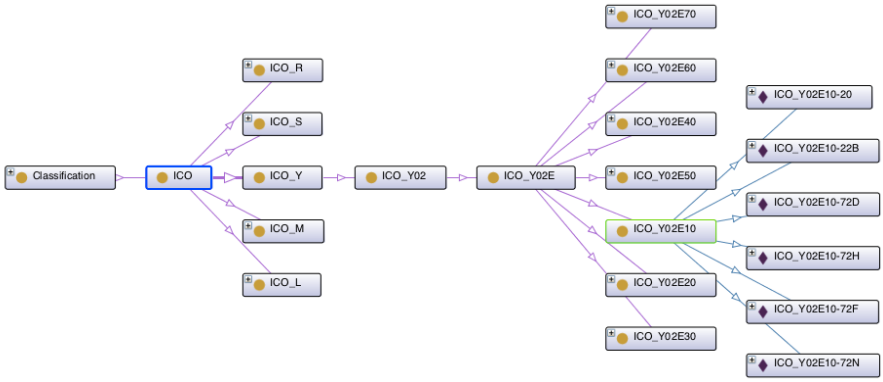


Fig. 2. HTCOntology focus on the ICO code Y02E10/20 in Protégé

3.3 Customization of the Stylesheets for Hierarchical Codes

In the method proposed, the customized stylesheets will allow the automatic translation of XML instance documents (where the instances are hierarchical codes) into an OWL ontology model and instances with the process shown in Figure 1. While other works translate one XML label (instance) into one single OWL label (class or instance), our work, on the other hand, proposes the translation of one XML label (instance) into several OWL labels (several classes and one instance) following an implicit hierarchy in the XML label.

The customization of the stylesheets proposed in this paper consists of the following steps:

1. Study the codes and define their parts, identifying the number of characters or digits in each part and whether they have separating characters
2. Study the structure of the labels in the XML and OWL files
   2.1. Location of the hierarchical codes in the XML file
   2.2. Detect the class from which implement HCOntology in the OWL file
3. Write the XSL file with its different parts:
   3.1. A header of the OWL file with the namespaces and if HTCOntology is implemented on top of an existing ontology, import the existing ontology
   3.2. Clean each label of the XML file, deleting unnecessary spaces, ensuring that each code is only written once in the OWL file, even if the code is repeated in several places in the XML file, etc.
   3.3. Create the hierarchical code structure
      3.3.1. Define the whole code except the last part as a subclass of the previous part of the code
      3.3.2. Repeat Step “3.3.1” for each part of the code until the first part of the code, and define the first part of the code as a subclass of the class of step “2.2”
      3.3.3. Insert the individuals (the whole code) in the corresponding created class
   3.4. Close the open labels

Table 1 shows an example for ICO codes. In which, we make use of the XPath functions related to substrings to split the hierarchical codes into its parts and create the corresponding ontology classes, subclasses and individuals for each code. This example shows the method steps to create an XSL file with the ICO codes (for clarity purposes we will show the instance of the ICO code in the XML file: Y02E10:20):

Table 1. Customizationof the stylesheets for ICO Codes

Step	Result
1	The structure of the ICO codes is shown in Section 3.2. The mark ":" separates the main group from the group in the XML file
2	Study the labels in XML and OWL file
2.1	<RESULT-LIST> <ROW> <ICO><p>
2.2	http://www.semanticweb.org/HCOntology#ICO
3	Write XSL file
3.1	The RDF namespace envelope and the ontology elements [25]
3.2	Clean the labels in the XML file that contains the hierarchical code
3.3	The code Y02E10:20 (figure 1) should have an OWL instance ICO_Y02E10-20, with the previous OWL hierarchy of classes (ICO_Y, ICO_Y02, ICO_Y02E, ICO_Y02E10). Figure 3 shows this step in the XSD file
3.3.1	In the XML instance Y02E10:20. Create the classes ICO_Y02E10 and (ICO_Y02E) and then define one as a subclass of the other (ICO_Y02E10 subclass of ICO_Y02E)
3.3.2	Create the class ICO_Y0E and define it as a subclass of ICO_Y02, create the class ICO_Y02 and define it as a subclass of ICO_Y, create the class ICO_Y and define it as a subclass of ICO
3.3.3	Add the instance ICO_Y02E10-20 to the class ICO_Y02E10
3.4	Close all the labels that remain opened

```
<xsl:variable name="var8" select="concat('ICO_',substring-before(normalize-space(.),':'))"/>
   <owl:Class rdf:about="{$var8}">
       <rdfs:subClassOf rdf:resource="{concat('ICO_',substring(normalize-space(.),1,4))}"/>
   </owl:Class>
   <owl:Class rdf:about="{concat('ICO_',substring(normalize-space(.),1,4))}">
        <rdfs:subClassOf rdf:resource="{concat('ICO_',substring(normalize-space(.),1,3))}"/>
   </owl:Class>

   <owl:Class rdf:about="{concat('ICO_',substring(normalize-space(.),1,3))}">
        <rdfs:subClassOf rdf:resource="{concat('ICO_',substring(normalize-space(.),1,1))}"/>
   </owl:Class>

   <owl:Class rdf:about="{concat('ICO_',substring(normalize-space(.),1,1))}">
        <rdfs:subClassOf rdf:resource=" http://www.semanticweb.org/HCOntology#ICO "/>
   </owl:Class>
   <owl:NamedIndividual rdf:about="{concat('ICO_',substring-before(normalize-space(.),':'),
                                           '-',substring-after(normalize-space(.),':'))}">
        <rdf:type rdf:resource="{concat('ICO_', substring-before(normalize-space(.),'/'))}"/>
   </owl:NamedIndividual>
```

Fig. 3. Excerpt of XSD file implementing step 3.3.1, 3.3.2 and 3.3

4 Reasoning through Ontology Hierarchy

A broader international scope of selected regions for the exploitation of patented environmental innovation provides existing patents a greater potential to influence

firms' financial performance [26]. Hence, our case study illustrates how the proposed method enables relationships to be created between concepts using the full semantics of the technological codes. In particular, it examines the patent portfolio of the German company and classifies its environmental patents (patents with an ICO code starting with Y02) as international patents if they have a family patent outside its region (Europe). In this example, the international patents of the family have been considered as the ones filed in the main patent offices outside Europe (the WIPO-WO- and the offices of China-CN-, USA-US-, Korea-KR- and Japan-JP-).

The XML file is a search result from the EPO database with the firms' patents. The OWL file is the result of applying the proposed methodology in Section 3 to the XML file. In order to exploit the reasoning over the resulting ontology OWL-DL expressivity is needed. Axiom 1 shows in description logic notation [11] the equivalent class created that classifies patents belonging to the class ICO_Y02 and that has one instance of a family patent with the representation WO, CN, KR, US or JP, in the class GlobalFamily. The representation of axiom 1 in Protégé is shown in the EquivalentClasses part of Figure 4.

The reasoner Pellet [27] has classified all the environmental patents with a family patent in one of the above patent offices as "globalPatent" (see the Menbers part of Figure 4). For example, Siemens patent PL356346A120040628, entitled "group of at least two windmills", with Y02E10/20 ICO code, and with family patents in WO and CN, among others, has been classified as GlobalPatent. Without HCOntology, these new relationships should not be possible and so HCOntology therefore enriches the information retrieval and management process and thus improves patent analysis.

$$
\begin{aligned}
&GlobalFamily \\
&\equiv PatentDocument \sqcap \exists hasICOCode.ICO\_Y02 \\
&\sqcap \big((\exists hasFamilyStates.\{WO\} \\
&\sqcup \exists hasFamilyStates.\{CN\} \\
&\sqcup \exists hasFamilyStates.\{US\} \\
&\sqcup \exists hasFamilyStates.\{KR\} \\
&\sqcup \exists hasFamilyStates.\{JP\})\big)
\end{aligned}
$$

Axiom 1: Definition of the equivalent class of GlobalFamily

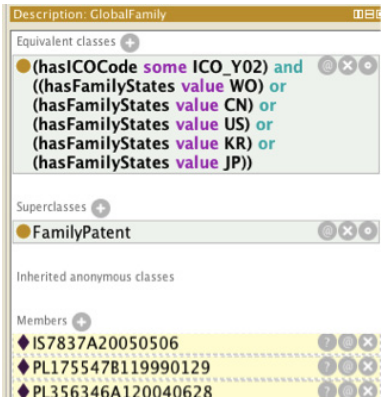


Fig. 4. Equivalent class GlobalFamily in Protégé

5 Conclusions

Information about patents is stored in large patent document databases with heterogeneous data representations. The technological field, although defined as a

single register in the databases, has ample implicit information about the classification of the technological field to which the patent belongs due to the hierarchical nature of the technological codes. However, this information is stored without any semantics and the information is not therefore explicitly represented to be automatically processed by computers. Patent documents can be retrieved from the databases in several standard formats, such as CSV or XML. Because of the limited semantics of such formats, additional relationships cannot be inferred between patents for analysis purposes, and makes hampers the sharing of heterogeneous data from different databases. Ontology languages such as OWL provide this semantics and have proved useful for representing and managing knowledge. Various efforts have been made to create patent ontologies in OWL, enabling new knowledge to be discovered and patent analysis improved, and these could be used by firms to optimize innovation management. Such ontologies have been populated by transforming XML instances into OWL instances. These translations are, however, limited to mirror hierarchies of patent codes and do not take further advantage of reasoning capabilities, and have not dealt with the heterogeneity of patent representation by different patent offices.

This article introduces an approach for automatically retrieving information from different patent repositories and for indexing hierarchical codes with ontological categories. This indexing enriches the information retrieval and management process with new relationships, properties and inferred information. The paper also discusses the importance and potential of this indexing.

More specifically, this paper provides a method for automatically translating XML instances from different data repositories into OWL classes and instances. This method is based on XSLT, and the XSD file is built with the help of XPath that splits the code into its structural parts. This methodology enables the future emerging codes to be translated automatically without the need for any reimplementation. The resulting hierarchy of concepts in OWL (which we called HCOntology) allows the exploitation of the information gathered in each part of the hierarchy.

We have also shown the potential of the proposal through a case study that creates relationships between concepts using the full semantics of the technological codes. In particular, we have classified as international environmental patents those patents that have family patents outside the region of the owner firm and are environmental patents. The detailed empirical comparison between our proposal and other methods is out of the scope of this paper because space restrictions, however we are implementing this analysis in a future paper that it is now under preparation.

Although this method has been applied in the domain of patents, it could equally be applied to any domain with hierarchical codes. The hierarchical codes would therefore be enriched with semantics, enabling the definition of new relationships, properties and inferred information.

References

1. Giereth, M., St, A., Rotard, M., Ertl, T.: Application of Semantic Technologies for Representing Patent Metadata. In: 1st International Workshop on Applications of Semantic Technologies AST 2006, vol. 94, pp. 297–305 (2006)
2. Almeida, P., Phene, A.: Subsidiaries and knowledge creation: the influence of the MNC and host country on innovation. Strategic Management Journal 25, 847–864 (2004)

3. Vries, D.: Leveraging Patents Financially. Gabler Verlag, Wiesbaden (2012)
4. Oltra, V., Kemp, R., Vries, F.: de: Patents as a measure of eco-innovation. International Journal of Environmental Technology and Management 13, 130–148 (2010)
5. Trappey, C.V., Wu, H.-Y., Taghaboni-Dutta, F., Trappey, A.J.C.: Using patent data for technology forecasting: China RFID patent analysis. Advanced Engineering Informatics 25, 53–64 (2011)
6. Foglia, P.: Patentability search strategies and the reformed IPC: A patent office perspective. World Patent Information 29, 33–53 (2007)
7. W3C: Extensible Markup Language (XML), 2nd edn. (2006)
8. Taduri, S., Lau, G.T., Law, K.H., Yu, H., Kesan, J.P.: Developing an Ontology for the U. S. Patent System. Text (2011)
9. Guarino, N., Giaretta, P.: Ontologies and Knowledge Bases: Towards a Terminological Clarification. In: Mars, N. (ed.) Towards Very Large Knowledge Bases Knowledge Building and Knowledge Sharing, pp. 25–32. IOS Press (1995)
10. W3C: OWL - Ontology Web Language (2009)
11. Baader, F.: The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press (2003)
12. Taduri, S., Lau, G.T., Law, K.H., Yu, H., Kesan, J.P.: An Ontology to Integrate Multiple Information Domains in the Patent System. In: 2011 IEEE International Symposium on Technology and Society ISTAS, pp. 23–25 (2011)
13. Wanner, L., Brügmann, S., Diallo, B., Giereth, M., Kompatsiaris, Y., Pianta, E., Rao, G., Schoester, P., Zervaki, V.: PATExpert/: Semantic Processing of Patent Documentation. Knowledge Creation Diffusion Utilization (2009)
14. W3C: XSL Transformations (XSLT), `http://www.w3.org/TR/xslt`
15. LexisNexis Website
16. Ferdinand, M., Zirpins, C., Trastour, D.: Lifting XML Schema to OWL. In: Koch, N., Fraternali, P., Wirsing, M. (eds.) ICWE 2004. LNCS, vol. 3140, pp. 354–358. Springer, Heidelberg (2004)
17. Bedini, I., Gardarin, G., Nguyen, B.: Deriving Ontologies from XML Schema. In: Proceedings EDA 2008 (2008)
18. Bosch, T., Mathiak, B.: XSLT transformation generating OWL ontologies automatically based on XML Schemas. IEEE (2011)
19. Lacoste, D., Sawant, K.P., Roy, S.: An efficient XML to OWL converter. In: Proceedings of the 4th India Software Engineering Conference 2011 ISEC 2011, pp. 145–154 (2011)
20. Rodrigues, T., Rosa, P., Cardoso, J.: Moving from syntactic to semantic organizations using JXML2OWL. Computers in Industry 59, 808–819 (2008)
21. Cardoso, J., Bussler, C.: Mapping between heterogeneous XML and OWL transaction representations in B2B integration. Data & Knowledge Engineering 70, 1046–1069 (2011)
22. Bohring, H., Auer, S.: Mapping XML to OWL Ontologies. Leipziger InformatikTage 72, 147–156 (2005)
23. Malhotra, A., Melton, J., Walsh, N.: XQuery 1.0 and XPath 2.0 Functions and Operators (2007)
24. Noy, N.F., Fergerson, R.W., Musen, M.A.: The knowledge model of Protégé-2000/: combining interoperability and flexibility. Knowledge Engineering and Knowledge Management Methods Models and Tools 1937, 1–20 (2000)
25. Smith, M.K., Welty, C., McGuinness, D.L.: OWL Web Ontology Language Guide (2004)
26. Kirca, A.H., Hult, G.T.M., Mena, J.A., Miller, J.C.: Firm-Specific Assets, Multinationality, and Financial Performance: A Meta-Analytic Review and Theoretical Integration. Academy of Management Journal 54, 47–72 (2011)
27. Sirin, E., Parsia, B., Grau, B., Kalyanpur, A., Katz, Y.: Pellet: A practical OWL-DL reasoner. Web Semantics Science Services and Agents on the World Wide Web 5, 51–53 (2007)

Multilevel Clustering of Induction Rules for Web Meta-knowledge

Amine Chemchem, Habiba Drias, and Youcef Djenouri

USTHB, LRIA
BP 32 El Alia Bab Ezzouar, Algiers, Algeria
aminechemchem@gmail.com, hdrias@usthb.dz,
youcef062009@hotmail.fr

Abstract. The current World Wide Web is featured by a huge mass of knowledge, making it difficult to exploit. One possible way to cope with this issue is to proceed to knowledge mining in a way that we could control its volume and hence make it manageable. This paper explores meta-knowledge discovery and in particular focuses on clustering induction rules for large knowledge sets. Such knowledge representation is considered for its expressive power and hence its wide use. Adapted data mining is proposed to extract meta-knowledge taking into account the knowledge representation which is more complex than simple data. Besides, a new clustering approach based on multilevel paradigm and called multilevel clustering is developed for the purpose of treating large scale knowledge sets. The approach invokes the k-means algorithm to cluster induction rules using new designed similarity measures. The developed algorithms have been implemented on four public benchmarks to test the effectiveness of the multilevel clustering approach. The numerical results have been compared to those of the simple k-means algorithm. As foreseeable, the multilevel clustering outperforms clearly the basic k-means on both the execution time and success rate that remains constant to 100 % while increasing the number of induction rules.

Keywords: Knowledge mining, meta-knowledge, multilevel paradigm, k-means, k-nearest neighbors, induction rules, genetic algorithm.

1 Introduction

To accelerate knowledge discovery [1][2], it would be interesting to exploit the knowledge currently present on the web and proceed to its mining. The purpose of this paper is to propose a knowledge mining process and to show how to adapt some data mining tasks to knowledge. Let us point out that the concept of knowledge mining is different from the one we found in the literature [3][4][14][15]. To be clear, we are interested in this work in mining knowledge instead of elementary data and the result of the desired task is therefore meta-knowledge.

The only similar study we found in the literature deals with frequent sequential patterns [5]. In fact in this paper, the authors focus on clustering sets of items and not

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 43–54.
DOI: 10.1007/978-3-642-36981-0_5

on simple items and more precisely sequential patterns. In our case we are interested in induction rules because of their closeness to the natural language. On the other hand, clustering is considered as a mining task for scalability concern. A new clustering approach based on multilevel paradigm is proposed.

The idea behind the multi-level approach is to be able to tackle very large scale knowledge sets. The reminder of the paper is organized as follows. The next section presents the concept of knowledge mining compared to data mining. Afterwards, induction rules representation and similarity measures are proposed according to the morphological aspect. Then a new clustering approach based on multilevel paradigm for induction rules is proposed in section3. The experimental evaluation presented in section 4 is performed on a benchmark including three different knowledge bases and another one containing hard SAT instances. Conclusions are finally summarized and some perspectives are suggested.

2 Knowledge Mining

In the literature [14] the knowledge mining concept is defined as an evaluation process of knowledge discovery, or as a selection process of interesting knowledge [15]. We can also cite many other works [3] [4] that define the concept of knowledge mining as a process of extracting knowledge from a tremendous database using prior knowledge. In other words, we illustrate the concept by the following scheme:

Prior Knowledge + data mining +goal → desired knowledge.

This definition deals with mining data not knowledge, it is different from the paradigm that we aim at introducing and studying.

We define knowledge mining as the process of extracting new knowledge from a knowledge set, such as a knowledge base or a knowledge warehouse. It can be then described as follows:

Knowledge + Prior Knowledge + Goal → Meta-Knowledge

2.1 Knowledge Representation

Knowledge representation consists in translating the natural knowledge into a symbolic formalism that can be processed by a machine. Knowledge is too vast and diverse to be represented and operated by a single formalism. Among the knowledge representations, the procedural and the declarative approaches have been investigated for a long time [6]. The procedural approach, invokes the simplicity and the ease of understanding reasoning, represented by algorithms simulating real behaviors. In addition, the procedural representation allows treating problems with algorithmic style which is fully analyzable and fully understandable. On the contrary, the declarative approach is more flexible because it provides heuristic expressions using statements. It allows to specify constraints and to learn independently methods of use. Fig. 1 shows the various knowledge representation formalisms from procedural which

is rigid and well structured to the declarative one which is on the contrary more open and free.

Clearly, induction rules are the closest to the natural language phrases. In addition, they represent an efficient framework to design reasoning systems and by producing meta-knowledge.

Another subsequent extension of induction rules mining is the design of super intelligent agents that are capable to reason on subsets of rules rather than on single rules. Lot of researches can be launched from this idea, such as knowledge represented by taxonomies or association rules.

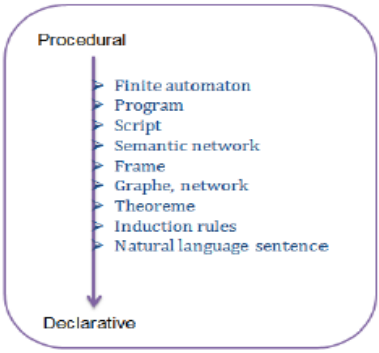


Fig. 1. Knowledge Representation

2.2 Induction Rules

An induction rule is a Boolean formula of the form: $R : X \rightarrow Y$, where X and Y are sets of clauses. X is called the premise of the rule and Y its consequence[3].

2.3 A Distance Measure for Knowledge Rules

In the previous work [7], a similarity measure between induction rules was proposed. It took into account the main items of the rule except the logical connectors in the design of the distance between two rules. This is a subtle way to evaluate the differences between rules. However, one important element was not considered at all, which is the position of the implication operator because this element distinguishes between the premise and consequent parts of the rule. In the present paper, we integrate this feature in designing similarity measures in order to be more precise and hopefully yield more accurate results. In addition to that fact, the clauses are taken as basic entities for measuring the disparity between rules. Moreover, if we consider the pair (variable, value) when ignoring the relational operator instead of the whole clause we obtain another measure with average bounded rationality. For a completely bounded rationality, we can take in account only the variables of the clauses to obtain a third measure. These three measures are summarized as follows:

Distance with Complete Rationality

$$Dist(C_i;C_j) = total(C_i;C_j) - shared\ (C_i;C_j) \quad (1)$$

$$Dist1(R_i;R_j) = Dist(antecedent(R_i),antecedent(R_j)) + Dist(consequent(R_i),consequent(R_j)) \quad (2)$$

where C_i and C_j are respectively two sets of clauses and the distance between them is equal to the number of clauses that are totally different from one clause to the other.

Antecedent(R) is the set of clauses appearing in the premise part of the rule R and consequent(R) is the set of clauses appearing in the consequence part of the rule R.

Distance with Average Bounded Rationality
In this case, the same formula is used except for $Dist(C_i;C_j)$ where we ignore the relational operators of the clauses.

Distance with Full Bounded Rationality
In this case, the same formula is used except for $Dist(C_i;C_j)$ where only the variables are considered. Of course we will obtain three kinds of clustering depending on which distance we select.

Let look at the following rules:

R1: If (temperature = hot) and (humidity = low) then (outlook=sunny),
R2: If (outlook=sunny) and (temperature = hot) and (wind = light) then (play_tennis=no),
dist(R1,R2)= total(R1,R2) - shared(R1,R2) = 7 - 4 = 3. computed by [7]
dist(R1,R2)= dist(antecedent (R1),antecedent(R2)) + dist(consequent(R1), consequent(R2)) = (5-1)+(2-0) = 6, computed by the complete rationality distance.

The defined distance is a reliable and valid metric measure across the whole of the induction rules, because we have demonstrated mathematically the four properties of a metric distance function, which are:

1- D is a function which is defined as follows:
$D : E\times E \rightarrow R$ // E is the whole of induction rules: $(X, Y) \rightarrow D(X, Y)$
2- $\forall x \in E : D(x, x) = 0$.
3- $\forall (x, y) \in E^2$: $D(x, y) = D(y, x)$.
4- $\forall (x, y, z) \in E^3 : D(x, y) = D(x, z) + D(z, y)$.

2.4 Centroids Computation

Another concept, which is necessary to perform clustering using k-means is the centroid. The latter, which is the central element of the cluster represents somehow all the objects of the whole cluster. Finding formula that computes exact centroid of a set of rules is not evident. One possible idea is to calculate the distance that separates each pair of rules, then for each rule associate the sum of the distances that separates it from the others. The centroids will correspond to the rule that has the near-average sum of distances because intuitively, it is the closest to all the other rules and hence the most similar to them.

3 Multilevel Induction Rules Clustering

3.1 Related Works of Multilevel Paradigm

In the literature, the multilevel paradigm was first proposed by [8], as a method of speeding up spectral bisection, and improved by generalizing it to encompass local refinement algorithms [9]. It has been made popular by [10], and since then

it was commonly used for graph partitioning problems, as the work [11]. Also the authors in [12] hybridize a multilevel approach with an ant-colony meta-heuristic for mesh-partitioning. And recently the multi-level approach was used for solving problems such as combining memetic algorithms with multi-level approach for solving the problem SAT[13]. The multilevel paradigm is a simple technique which at its core involves recursive coarsening to produce smaller and smaller problems that are easier to solve than the original one. This paradigm consists of three phases: coarsening, initial solution, and uncoarsening. The coarsening phase aims at merging the components associated with the problem to form clusters. The clusters are used in a recursive manner to construct a hierarchy of problems. Each level of this hierarchy represents the original problem but with fewer degrees of freedom. The coarsest level can then be used to compute an initial solution.

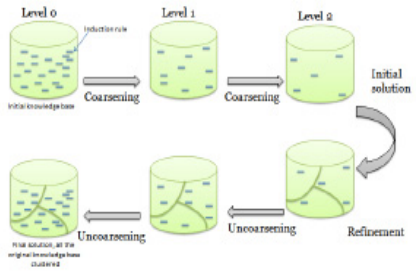


Fig. 2. Multilevel Induction Rules Clustering

This last, found at the coarsest level is extended to give another initial solution for the next level, and then improved using a chosen optimization algorithm.

3.2 The Proposed Algorithm

In this section, we develop a new general algorithm for induction rules clustering based on multilevel paradigm. First, it begins by eliminating the nearest induction rules from the original knowledge set in the coarsening step to keep only the most far away rules from each other. Thanks to a genetic algorithm that computes at the end of this stage a mini knowledge base that includes the different most distant induction rules as explained in fig.2. After this step, the k-means algorithm is launched on the small base and it is expectable that it will provide reliable centroids, because the selected rules are very distant from each other. Finally in the refinement step, the k-NN algorithm adapted to induction rules is invoked on the result of k-means of the previous step. This operation is repeated until obtaining the initial knowledge set. The purpose of using k-NN is to complete inserting the remaining rules in the corresponding classes.

3.3 Coarsening Step

The purpose of this step is to reduce the induction rules number. We use a genetic algorithm for this aim while ensuring a good sweeping of all the knowledge base. The rules are not removed randomly but in a way to keep only the rules that appear very far from each other.

A Chromosome (Individual)
It is an encoded solution in our case represented by an array of integers varying between 0 and the number of induction rules of the knowledge set, each gene representing a rule. Therefore the length of the chromosome represents the size of the small knowledge base. For instance, if we have the following chromosome (19,201,1987,2012,3033,4200), its size is equal to 6, the number of rules, which can be located as follows: The first rule is at position 19 in the set, the second one at position 201 and so on...

Fitness Function
The notion of fitness is fundamental to the application of genetic algorithms. It is a numerical value that expresses the performance of an individual (solution), so that different individuals can be compared. We propose as a fitness function the sum of the distances between each of the genes of the individual, that is:
$\text{fitness}(\text{chromosome}) = MAX\ (\sum_{i=1}^{n-1} \sum_{j=1}^{n} Dist(i,j))$, for all i different from j with i and j belonging to chromosome.

Coarsening Algorithm

```
Genetic Algorithm for Coarsening Step
Begin
   Generate randomly an initial population of solutions;
   Compute fitness for each solution of the population;
    For ( i:= 1 to maxIter)do
      Choose 2 individuals (or solutions);
        Generat Rc randomly RC, RM : a random numbers between [0 and
        length of chromosome];
          If (Rc < rate of crossover) then
               Apply the crossover;
          end_If
          If (Rm < rate of mutation) then
               Apply the mutation on one chromosome selected
                  randomly;
          end_If
     Evaluate fitness(S')
               /*where S' is the new individual*/
          If (fitness (S') > fitness (S)) then
               Remove S from the population and insert S';
          end_if
    end_for
   Consider the best found solution;
end
```

3.4 The Initial Solution

After the coarsening step, we obtain a solution that maximizes the distance between its induction rules. So it is sour that the last one sweeps all the areas of the knowledge set. In this step, we apply a simple k-means algorithm to cluster these rules as explained in Fig.3 and the reliable gravity centers will be achieved. The result of this clustering constitutes the initial solution.

3.5 The Refinement Step

After the clustering in the initial step, the different clusters are obtained with the rules belonging to the initial solution. In order to rebuild our initial knowledge set, we suggest applying the k-nearest-neighbors algorithm for the rest of the induction rules, as shown in the following algorithm.

Refinement Step Algorithm based on k-NN

```
Begin
rule_list := all rules (knowledge base) - rules (initial
 solution);
   while (rule_list is not empty) then
      Current_rule  :=  get_rule (rule_list);
         for  (each  already  classified  rule  Ri)  do
           Calculate  the  distance  Dist_Clauses  (curr-
           nt-rule,Ri);
         end_for
      Compute  k_nearest_neighbor(current-rule);
          for  (each  rule  belonging  to  k-NN)  do
           Calculate  the  number  of  frequency  of  each
           class;
         end_for
      Attribute  to  "current_rule"  the  most  frequent
      class;
   end_while
End.
```

4 Evaluation and Experimentation

One of the hardest problems in comparing different clustering algorithms is finding an algorithm to evaluate the quality of the clusters. Our main idea is to construct the induction rules set from three different benchmarks, and with fixing the parameter “k” of the clustering algorithms at 3, we can calculate the clustering success rate as explained in fig.3. And in a second step and only after having validated the clustering approach we develop it on a public SAT benchmark.

4.1 Benchmark Construction

Data Benchmark Adaptation
In this part, we build the knowledge set from three public different benchmarks. It includes 11000 induction rules, after transforming the data sets to induction rule sets, as follows:

If ($attribute_1 = value_1$) and ($attribute_2 = value_2$) and ... then ($attribute_n = value_n$)

Where n is the last attribute of the data set.

The first one is originally a big data set known as **Chess (King-Rook vs.King) Data Set**[1], which contains 28056 instances with 7 attributes.

The second benchmark is known as **Abalone Data Set**[2]. It contains 4177 instances with 9 attributes. The third data set is known as **Car Evaluation Data Set**[3]. It contains 1728 instances, with 7 attributes.

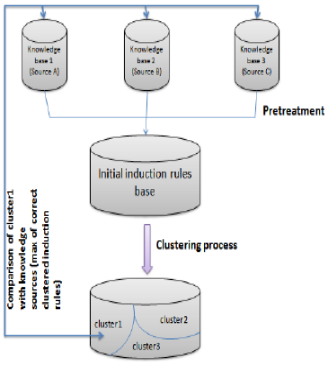


Fig. 3. Clustering Evaluation

SAT Benchmark Adaptation

We used the public SAT benchmark known as SATLIB[4], it contains 1000 SAT instances with 20 variables and 91 clauses. A SAT Benchmark is the set clause-conjunctions of literals. Each instance SAT can be modified as follows:

Let $C_i = \neg\alpha 1 \vee \alpha 2 \vee \neg\alpha 3$ be a clause, then

As $a \Rightarrow b \equiv \neg\alpha 1 \vee \beta$ then: $\neg\alpha 1 \vee \alpha 2 \vee \neg\alpha 3 \Rightarrow (\alpha 2 \vee \neg\alpha 3)$...(3)

Using Eq.(3), we can consider the following rule: if $\alpha 1 = 1$ then $\alpha 2 = 1$ or $\alpha 3 = 0$. Generally, each clause C is transformed as Eq.(3).

After that, for each variable,$\alpha_i \in C$, if α_i is true, then it set to 1, otherwise to 0.

4.2 Evaluation Pattern

Fig.3 shows the evaluation process of the clustering, when fixing the parameter k of the clustering algorithms to 3. Then after the clustering step, the three obtained clusters are compared to the initial rule bases. If the clustering process is efficient, the respective clusters should be identical. The success rate is calculated using the following formula:

$$\text{success rate} = \frac{ncr}{npr} = \frac{ncr1}{npr1} + \frac{ncr2}{npr2} + \frac{ncr3}{npr3} \quad \ldots \qquad (2)$$

Where, ncri=number of correct rules for knowledge base i.
npri=number of pertinent rules from knowledge base i (total number of its rules).

The number of correct rules of KB_i = max(number rules common (KB_i;C_j)) for all j in [1,3].

[1] http://archive.ics.uci.edu/ml/datasets/Chess+%28King-Rook+vs.+King%29
[2] http://archive.ics.uci.edu/ml/datasets/Abalone
[3] http://archive.ics.uci.edu/ml/datasets/Car+Evaluation
[4] http://www.satlib.org/

4.3 Experimentation

Performance Comparison

Fig.4 compares the performance of several variants of the designed clustering approaches. We remark that the results computed by the simple k-means and the multi-level k-means are very different. While the success rate of the simple k-means algorithms is reduced when increasing the number of rules to cluster, the multilevel algorithms success rate remain constant to 100%.

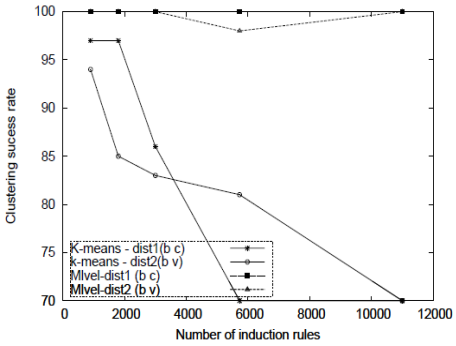


Fig. 4. Success rate of clustering approaches

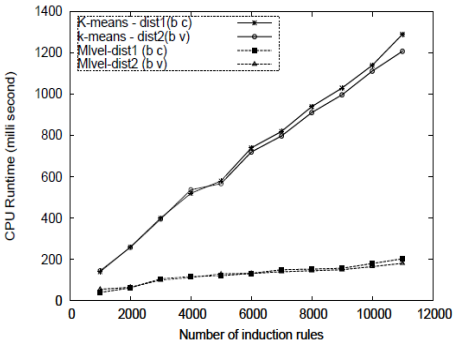


Fig. 5. CPU Time of clustering approaches

CPU Runtime Comparison

The execution times for the different clustering approaches are compared in Figure 5, when increasing the number of induction rules. We observe that the multilevel clustering algorithms are faster than the simple clustering. Moreover, when the simple clustering approaches cluster a number of induction rules varying between 100 and 11000, the execution time increases from 180 to 1270 milliseconds whereas the execution time of multilevel approaches increases just between 26 and 200 millisecond.

Application of Clustering Approaches on SAT Benchmark

Fig.6 shows how the execution time augments with the increase of number of sat instances for the four clustering approaches. According to this figure, we remark that the simple clustering approaches are slower than the two approaches based on multilevel paradigm, like in the previous subsection2. This is due to the fact that the simple clustering approaches repeat a computation of new centroids, until the stability of all induction rules in clusters, however, in the multilevel clustering approaches the gravity centers are computed only once. Execution time of the multilevel approaches do not exceed 35milliseconds, when the number of sat instances is 1000 rules, even though the simple clustering approaches execution time is 200 ms when the number rules is 1000.

In figure 7 the results of comparing the clustering approaches while increasing the parameter 'k' are shown. When the number of SAT instances is fixed at 1000 rules,

and with the increase of the number of clusters from 3 to 40 clusters, we remark that the CPU runtime of the simple clustering approaches increase between 125 to 290 milliseconds. However the CPU runtime of the clustering approaches based on multilevel paradigm increases just from 40ms to 90 milliseconds.

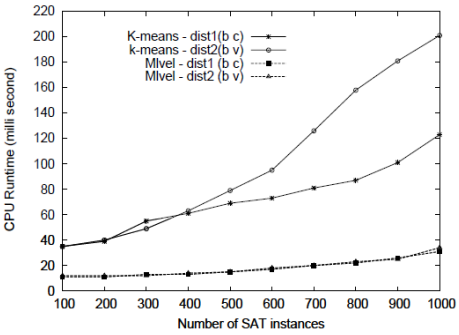


Fig. 6. CPU Runtime of clustering approaches for SAT instances

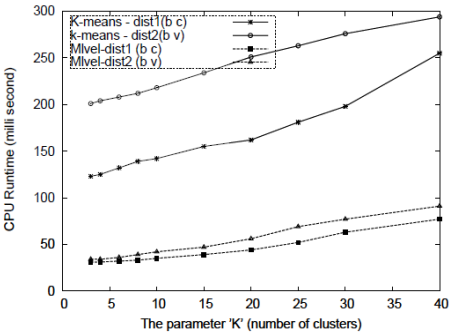


Fig. 7. CPU Time of clustering SAT instances Vs number of clusters

General Comparison

When comparing the four clustering approaches, we note that the approaches based on multilevel clustering are more efficient on both solution quality and runtime criteria than the simple clustering approaches, so they take until 83% less time than the simple clustering approaches to treat the same rule base.

Furthermore, even on the quality criterion expressed through the success rate, their performance reaches 100% whatever the number of rules whereas the simple clustering approaches success rate varies between 97% and 70%. Finally with applying the four approaches on SAT benchmark, and with increasing the number of clusters, the difference between the runtime of the approaches is very important. Therefore when increasing the number of clusters from 3 to 40 clusters, the runtime for the simple clustering approaches increase from 120 to 290 milliseconds, while the runtime of the multilevel clustering approaches remain constant between 40 and 90 milliseconds.

5 Conclusion

In this paper we presented a new multilevel k-means algorithm for clustering induction rules. The algorithm combines a traditional clustering technique in the initial solution step, a genetic algorithm in the coarsening step and a supervised classification especially the k-nearest-neighbors algorithm in the refinement step. As our experimental results proved, the best clustering solutions were produced by the multilevel clustering approaches when comparing to the simple k-means. Furthermore, the algorithm has the additional advantage of being extremely fast, as

it is based on genetic algorithm. For instance, the amount of time required by the proposed algorithm ranges from 26 milliseconds for a knowledge base with 900 induction rules to 115 milliseconds for a knowledge base with 11000 induction rules, on a I5 core PC. We believe that this paper presents the first attempt for developing a robust framework for a large scale clustering approaches. However, a number of key questions remain to be addressed, in particular the best way to design the different components of the multi-level paradigm. We can imagine in a future work, an application of the multilevel clustering of induction rules on a rule base of an agent intelligent, in order to speed up its reasoning and also to discover the new meta-knowledge.

References

1. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques, 3rd edn. Elsevier (2011)
2. Mariscal, G., Marbn, Fernndez, C.: A survey of data mining and knowledge discovery process models and methodologies. The Knowledge Engineering Review 25, 137–166 (2010)
3. Kaufman, K.A., Michalski, R.S.: From Data Mining to Knowledge Mining. In: Rao, C.R., Solka, J.L., Wegman, E.J. (eds.) Handbook in Statistics. Data Mining and Data Visualization, vol. 24, pp. 47–75. Elsevier/North Holland (2005)
4. Michalski, R.S.: Knowledge mining: A proposed new direction, School of Computational Sciences George Mason University and Institute for Computer Science Polish Academy of Sciences (2003)
5. Saneifar, H., Bringay, S., Laurent, A.: S2MP: Similarity Measure for Sequential Patterns. In: Proceeding of the 7th Australian Data Mining Conference AusDM 2008, Adelaide, Australia, November 27-28, pp. 95–104 (2008)
6. Tuomi, I.: Data is More Than Knowledge Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory. Journal of Management Information Systems 16(3), 107–121 (fall 1999)
7. Drias, H., Aouichat, A., Boutorh, A.: Towards Incremental Knowledge Warehousing and Mining. In: DCAI 2012, pp. 501–510 (2012)
8. Barnard, S.T., Simon, H.D.: A fast multilevel implementation of recursive spectral bisection for partitioning unstructured problems. Concurrency: Practice and Experience 6, 101–117 (1994)
9. Hendrickson, B., Leland, R.: A multilevel algorithm for partitioning graphs. In: Proceedings of the Supercomputing 1995 (1995)
10. Karypis, G., Aggarwal, R., Kumar, V.: A fast and high quality multilevel scheme for partitioning irregular graphs. SIAM J. Sci. Comput. 20(1), 359–392 (1999)
11. Dhillon, S., Guan, Y., Kulis, B.: Weighted Graph Cuts without Eigenvectors: A Multilevel Approach. IEEE Transactions on Patterns Analysis and Machine Intelligence 29(11) (2007)
12. Korosec, P., Silc, J., Robic, B.: A Multi-level Ant-Colony-Optimization: Algorithm for MESH Partitioning, Computer Systems Department, Jozef Stefan Institute, Ljubljana, Slovenia. IEEE 2003 Conference Publication (2003)

13. Bouhmala, N.: A Multilevel Approach Applied to Sat-Encoded Problems, Vestfold University College Norway. VLSI Design (2012) ISBN: 978-953-307-884-7
14. Poongothai, K., Sathiyabama, S.: Integration of Clustering and Rule Induction Mining Framework for Evaluation of Web Usage Knowledge Discovery System. Journal of Applied Sciences 12, 1495–1500 (2012)
15. Poongothai, K., Sathiyabama, S.: Efficient Web Usage Miner Using Decisive Induction Rules. Journal of Computer Science 8(6), 835–840 (2012)

Knowledge-Based Risk Management: Survey on Brazilian Software Development Enterprises

Sandra Miranda Neves[1,3], Carlos Eduardo Sanches da Silva[2], Valério Antonio Pamplona Salomon[3], and André Leonardo Almeida Santos[2]

[1] UNIFEI, Federal University of Itajuba, Irmã Ivone Drummond Street, 200 Industrial District, 35903-087, Itabira, Minas Gerais State, Brazil
sandraneves@unifei.edu.br
[2] UNIFEI, Federal University of Itajuba, BPS Ave. 1303, Pinheirinho, 37500-903, Itajuba, Minas Gerais State, Brazil
sanches@unifei.edu.br, andre.leonardo.cco@gmail.com
[3] UNESP, Sao Paulo State University, Ariberto Pereira da Cunha Ave. 333, 12516-410, Guaratingueta, Sao Paulo State, Brazil
salomon@feg.unesp.br

Abstract. A risks management, carried on in an effective way, leads the software development to success and may influence on the organization. The knowledge takes part of such a process as a way to help taking decisions. This research aimed to analyze the use of Knowledge Management techniques to the Risk Management in software projects development and the possible influence on the enterprise revenue. It had, as its main studying subject, Brazilian incubated and graduated software developing enterprises. The chosen research method was the Survey type. Multivariate statistical methods were used for the treatment and analysis of the obtained results, this way identifying the most significant factors, that is, enterprise´s achievement constraining factors and those outcome achievement ones. Among the latter we highlight the knowledge methodology, the time of existence of the enterprise, the amount of employees and the knowledge externalization. The results encourage contributing actions to the increasing of financial revenue.

Keywords: Software development, Knowledge Management, Risks Management, Incubated and Graduated Technological Based Enterprises.

1 Introduction

All projects have risks. Once a project management aims to increase the success rate of projects, it is, in its essence, the risk management [1]. In this context, the software projects are part of them and, especially susceptible to faults [2]. The high failure rates, associated with the information system projects, suggest that organizations need

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 55–65.
DOI: 10.1007/978-3-642-36981-0_6

to improve not only their ability to identify, but also manage the risks linked to those projects [3].

An organization cannot effectively manage the risks provided it doesn´t manage its knowledge [4]. However, it is not enough to have the knowledge in some part of the organization; it needs to be accessible to be used and to serve as a basis to new knowledge to be created [5]. Farias et al. [6] consider that projects managers may make mistakes from the past simply for not knowing the mitigation actions that have been successfully applied, once an applied ineffective risk Knowledge Management contributes to maximize such a problem. A research, carried on by Wong e Aspinwall [7], from a survey with 72 small and medium information technological enterprises, identified that only 26 enterprises used Knowledge Management formally. The main reason for not using it was due to the uncertainty as to the potential benefits (45.7%). For Alhawari et al. [8], to use Knowledge Management processes to improve the applications of processes, linked to Risk Management, is a recent and important research area. And yet, this research area is not so addressed. According to Massingham [9] the Knowledge Management and Risk Management relationship is an academic research emergent field. So, this research has as its aim to analyze the use of Knowledge Management (KM) techniques to the Risk Management (RM) in software projects development and the possible influence on the enterprise revenue.

The article is structured as follows. Section 1 presents justifications and the research aim. Section 2 discusses the theoretical foundation on Technological Incubated and Graduated Based Enterprises and the relationship between RM and KM. Section 3 contemplates the research method and Section 4 the resulting presentation and analysis. Finally, Section 5 presents conclusions and suggestions for future research.

2 Literature Review

2.1 Incubators and Technological Incubated and Graduated Based Enterprises

Enterprises incubators have become a phenomenon in many parts of the world and are seen as a tool to promote the growing of technological based enterprises development [10]. Dahlstrand [11] defines a Technological Based Enterprise (TBE) as that one depending on technology for its development, not meaning necessarily, in most of the cases, it has to be new or in innovation. The enterprises incubators supply technical support, networking capacity, infrastructure, shared services and facilitates access to capital, making it vital to business development in their early stages [12].

At Graduated Enterprises the information management and its procedures have been identified as being more aware and structured. In addition, incubated enterprises presented an organization environment with better conditions for knowledge creation [13].

2.2 Risk Management and the Relationship with Knowledge Management

Risks in software projects are a series of factors or conditions that may represent a serious threat for the success of the project achievement [14] and they imply to quantify the importance of a risk, assessing and its possible impact on the project, as well as in the strategies development to control it [15]. Despite the improvements already achieved, many software development projects still use more sources than planned, take longer to be finished and supply less quality and functionality than expected [16].

For Davenport e Prusak [5], KM is composed by set of processes that seek to support in the organizational environment the knowledge generation, their register and their transfer. Anantatmula e Kanungo [17] state that "knowledge is recognized as a critical resource to acquire and keep up competitive advantage in business". So, several enterprises have expectations that KM, if accomplished the right way, may transform knowledge into competitive advantage [18].

Normally micro and small enterprises use knowledge more than other traditional resources to compete. However, a micro and small enterprises (MSE) significant majority is not employing KM techniques [19]. These techniques, or practices, help enterprises to empower their knowledge generation capacity. According to Nonaka and Takeuchi [20], the knowledge generation goes on information interactions and its effective transformation occurs in 4 conversion modes, the so called SECI model: Socialization (S) - knowledge conversion from tacit into tacit; Externalization (O) - articulation process of the tacit knowledge into explicit concepts.; Combination (C) - the explicit knowledge conversion into explicit; Internalization (I) - incorporation of the explicit knowledge into tacit one.

The relationship between KM and RM is also approached by Farias et al. [6], who describe a software project risk planning approach, based on the reuse of organizational risk knowledge. Verhaegen [21] shows a KM as a tool to decision makers and management capacity improvement, especially regarding to risk related issues. Karadsheh et al. [22] present the KM as a strategic resource for organizations, and can have a major influence on risk reduction. Massingham [9] proposes and tests a Knowledge Risk Management (KRM). This author also discusses the application of KM tools and techniques for the organizational KM. Jafari et al. [23] elaborate and apply a model for Risk Management of knowledge loss at the projects management. Recently, Alhawari et al. [8] presented a structure of a Knowledge-Based Risk Management (KBRM) for information technology projects.

3 Research Method

It has been planned a Survey type research for obtaining empirical evidences. The steps will be carried out according to the sequence established in the works of Forza [24] e Bryman e Bell [25].

The objects of this study are Brazilian incubated and graduated software development enterprises. The choice of these enterprises is justified by researches as the one by Radas and Bozic [26], where Small and Medium Enterprises are

considered as economic growing propellers, as well as of employment generation, once, due to such an importance, the developed and developing countries are interest to meet ways so these enterprises accomplish innovations. Besides that, incubated enterprises present an organizational environment with better conditions for knowledge generation [13].

The population has been defined as of 89% (eighty nine) of Incubated and Graduate software Development Enterprises belonging to an Incubators, Technologic Parks and Tecnopolis association. The size of the sample, according to Malhotra [27] must be defined according to the type of the study to be carried on. The used sample was the population itself, chosen for the research, that is, all the 89 software development enterprises. This was not a random and intentional sample. For the collection of data it has been used a questionnaire validated in the research carried on by Neves [28] and made up for 31 (thirsty one) questions on different kinds of scale. The data were collected in the period from January till June 2012. There were performed eight rounds for sending the enterprises the questionnaire. At the end of the eight rounds, the total was 23 (twenty three) answers. The return rate was 26%, which, according to Forza [24] is an acceptable one.

4 Results Obtained and Analyzes

4.1 Obtained Results

The internal validation was carried out through Cronbach´s Alpha. It has been estimate the value of Cronbach´s Alpha through the Minitab 15® software, being the smallest result 0.6682 which, according to Malhotra [27] is regarded as acceptable. The external validity was obtained by reliability of the respondents (96% in charge of management and had more than two years at the enterprise), which assures validity to the obtained data. The respondents average age is 30% (thirty), whereas, considering the latter completed course, 57% concluded graduation, 30% specialization and 13% Bachelor. The average time of enterprises existence is of 4 (four) years, and the oldest is 13.5 (thirteen and a half) years old and it is a graduated enterprise, and the youngest is only 8 (eight) months old, and it an incubated enterprise. The enterprises have an average of 10 (ten) employees and the average projects time is of 8 (eight) months. They have an average annual billing of R$ 70 thousand.

Based mainly on the approach by Nonaka and Takeuchi [20], there have been identified the main techniques that contribute to the enterprises to maximize their capacity to generate knowledge. These techniques were associated with the SECI model. Fig. 1 presents the results with relation to the KM Techniques in the enterprise, in accordance with criteria (0) Never used, (1) Rarely used, (2) Little used, (3) Used and (4) Frequently used. It is worth mentioning, as a KM technique "Never used" by the enterprises, the Database Skills (43%, n=10) and as "Frequently used" the observation, imitation and practice (52%, n=12). Among the most cited we find the Telephone/computer network (65%, n=15).

Table 1 presents the evaluation of KM frequency use techniques (Fig. 1), associated to the conversion modes proposed through SECI model [20].

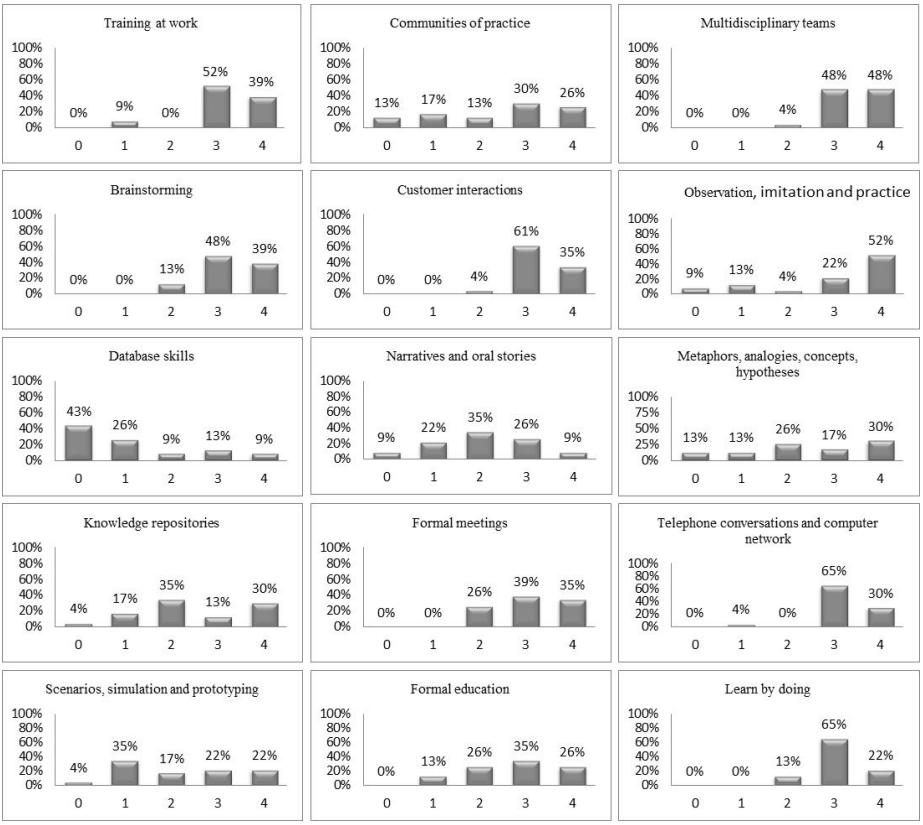


Fig. 1. KM Techniques use frequency

Table 1. Use of KM techniques frequency according to the conversion mode

Frequency	Socialization	Externalization	Combination	Internalization
0 – Never	9%	9%	1%	0%
1 - Rarely used	9%	17%	13%	0%
2 - Little used	7%	32%	17%	13%
3 - Used	39%	19%	40%	65%
4 - Frequently used	35%	23%	28%	22%

There are two aspects in this table: the "Externalization" conversion mode as "Little used" (32%) and the "Internalization" conversion mode as "Used" (65%). Most enterprises have an average level of awareness for RM (43%). However, a methodological formal use is bigger to RM (48%, n=11) than to KM (39%, n=9). As to RM, enterprises employ, in its major part, the Probability versus Impact analysis. Table 2 presents the same evaluation considering the separation of the enterprises in incubated (65%, n=15) and graduated (35%, n=8).

Table 2. Formal RM and KM methodologies use by Incubated and Graduated enterprises

Frequency	Incubated		Graduated	
	Yes	No	Yes	No
RM Formal Methodology	33% (n = 5)	67% (n = 10)	50% (n = 4)	50% (n = 4)
KM Formal Methodology	40% (n = 6)	60% (n = 9)	63% (n = 5)	38% (n = 3)

One notices the Graduated enterprises had the highest indexes regarding the use of formal methodology for RM (50%, n=4) and for KM (63%, n=5). The results highlight the research by Vick et al. [13]. According to the authors, the information management and its procedures were identified as being more aware and structured in Graduated companies. The obtained data, which were discrete variables, were transformed into continuous values, employing the Minitab l5® software and, later on, calculated their correlation [27]. The correlated factors are identified in Table 3.

Table 3. Correlation found in data analyses

Hypothesis	P-Value
There is a relationship between the enterprise annual billing and its employees number	**0.000**
There is a relationship between the enterprise annual billing and the existence time of the enterprise	**0.049**
There is a relationship between the enterprise annual billing and the project average time	0.455
There is a relationship between the enterprise annual billing and the existence of an RM methodology	0.234
There is a relationship between the enterprise annual billing and the existence of a KM methodology	**0.008**
There is a relationship between the existence of a KM methodology and the existence of an RM methodology	**0.002**
There is a relationship between the number of employees and the existence of an RM methodology	0.255
There is a relationship between the number of employees and the existence of a KM methodology	0.128
There is a relationship between the project average time and the existence of a KM methodology	0.542
There is a relationship between the project average time and the existence of an RM methodology	0.104
There is a relationship between the existing time of an enterprise and the existence of an RM methodology	0.483
There is a relationship between the existing time of an enterprise and the existence of a KM methodology	0.492
There is a relationship between the Board of Directors´ awareness level and the importance of an RM projects average time	**0.039**
There is a relationship between the enterprise annual billing and the Board of Directors´ awareness level as to the RM importance.	0.411

Observing the results (Table 3) it turns out that the hypotheses, statistically considered validated, with a significance of 5%, are the existence of a relationship among: (I) The enterprise annual billing and its number of employees, this relationship is directly pro rata, once this number is a key factor for the production amount, influencing this way the billing; (II) The annual billing and the enterprise existing time are two proportionally values, which makes sense, once a enterprise with a longer existence time tends to have a certain know-how, so having a larger project number and in a more effective way; III) a methodology existence for KM and its association with the billing, it corroborates with the statements by Anantatmula e Kanungo [17]. According to the authors, knowledge is recognized as a critical weapon to acquire and keep up competitive knowledge in business; IV) Relationship between a KM methodology and the existence of an RM methodology, the establishment of such a relationship strengthens and quantifies statements by Karadsheh et al. [22], where they present the KM processes as a strategic resource for organizations, being able to have a great influence on the risk reduction; (V) The Board of Directors awareness level as to the importance of the RM and the projects average time, the risks incorporation at projects events turns out to be feasibly timed for the projects, these times scheduled on a more realist way, considering the threats and opportunities or even bigger or smaller projects can be abandoned or encouraged, as well like decisions related to resources allocation.

4.2 Data Analysis through Partial Least Square (PLS)

The Partial Least Square regression method applies when there are: one or multiple dependent variables; highly correlated predictors; more predictors than observations [29], [30].

The calculations for analysis were performed in the Minitab15® software. It has been considered as the result the annual billing (Y), once, based on this number, one can evaluate the enterprise success in its line, once the goal is to maximize the billing of all the enterprises. The predictors variable (X) are: 1) the number of employees; (2) the enterprise existing time; (3) project average time; (4) the Board of Directors as to the importance of RM; (5) RM methodology; (6) KM methodology; (7) Socialization; (8) Externalization; 9) Combination; (10) Internalization. 5 (five) main components were selected, and the PLS mode presented explanation rates of 83.44% and P-Value of 0.000 (trust intervals of 95%). First, a PLS model was generated, when one checked the normalcy of the resulting waste, considered to be normal.

Fig. 2 presents a chart of the result of a PLS model that used the first 5 (five) main components. One used the PLS Std Coefficient Plot, which patterns allow the coefficients to be compared.

Negative coefficients indicate that the factor restrain the getting the results and positive coefficients indicate that factor contributes to achieve the result. This way, Table 4 presents the results of contributing factors, neutral and the billing restraining ones.

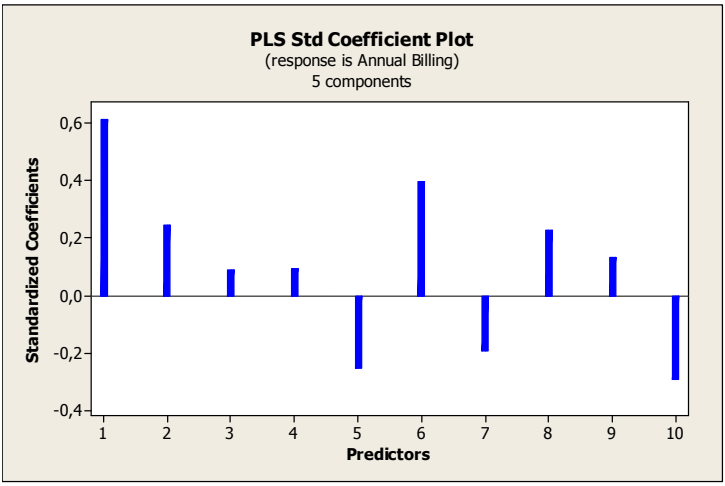


Fig. 2. PLS for the "Annual billing" result - deployed to the KM application

Table 4. Contributing factors to achieve results

Result: annual billing	10 factors (deployed the KM application)
Enhancing factors	Number of employees (1) KM methodology (6) Existing time of enterprise (2) Externalization (8)
Neutral factors	Combination (9) The Board of Directors awareness level as to the importance of RM (4) Projects average time (3)
Restraining factors	Internalization (10) RM methodology (5) Socialization (7)

As to the obtained results from the analysis of contributing factors for getting results, one highlights the existing number of employees at the enterprise (1) and use of a methodology for KM (6). Another factor that potentiates is associated to the "Externalization" (8) conversion mode, which, according to Nonaka and Takeuchi [20] has been neglected at organizations and should be taken into consideration. On the other hand, the financial results may be restricting for the use of a RM methodology (5). Considering this specified result and its relationship with the billing item, this restriction may be due to time issues and possible investment, like human resources and materials for the RM deployment.

5 Conclusions

This research aimed to analyze the use of KM techniques to the RM in software projects development and the possible influence on the enterprise revenue. It was obtained, as factors that contribute to achieving the financial results, the number of employees, the KM methodology and the enterprise existing time. It has been highlighted the "Externalization" conversion methodology as one of the factors to be observed by the enterprise, once it influences on the revenue, as a way of articulation of tacit knowledge for explicit ones. Such an articulation would allow the use of this knowledge for organization benefit, contributing especially with the help for taking decisions related to the risks. This is why it is important the existence of an environment that stimulates the knowledge sharing. The KM and methodology use techniques are still on an initial stage at the evaluated enterprises. It had been seen the many of them use KM techniques in their processes, but they haven´t formalized them so far as being KM. This attitude can make enterprises not to have the benefits the KM practice can provide, especially related to the help of taking decisions.

Taking into account that 53% of the respondents still doesn´t prioritize KM, it is still necessary to increase the awareness level as to the importance of this process for high tech enterprises, allowing the creation a proper environment for improving the organizational performance. Here is a reflection on "if and how far" these enterprises are performing a transfer strategy of their risks for the funding bodies. As the research main contribution it has been highlighted the fact of presenting, in a quantitative way, the benefits of the joint use of KM techniques and methodologies for RM, an item not very much exploited item in literature. The obtained results may show the software developing incubated and graduated enterprises managers the development of politics and strategies aiming to maximize the enterprise profits.

It is suggested, as for future research, to evaluate the efficiency of Incubated and Graduated Technological Based enterprises, through the use of Multiobjective Programming techniques, like the Goal Programming and Data Envelopment Analysis (GPDEA), having as the main variables the methodologies use for KM and RM and its relationship with the enterprises efficiency.

Acknowledgments. The authors need to express their acknowledgments to two Brazilian research agencies: the CAPES Foundation (Grant No. PE024/2008) and FAPEMIG (Grant No. PPM-00586), and especially all interviewees.

References

1. Kendrick, T.: Identifying and managing project risk: essential tools for failure-proofing your project. Amacom, NY (2003)
2. Bannerman, P.L.: Risk and risk management in software projects: A reassessment. The Journal of Systems and Software 81, 2118–2133 (2008)
3. Jiang, J.J., Klein, G., Discenza, R.: Information system success as impacted by risks and development strategies. IEEE Transactions on Engineering Management 48, 46–55 (2001)

4. Neef, D.: Managing corporate risk through better knowledge management. The Learning Organization 12, 112–124 (2005)
5. Davenport, T.H., Prusak, L.: Working Knowledge: How organizations Manage what they know. Harvard Business School Press (1998)
6. Farias, L.L., Travassos, G.H., Rocha, A.R.: Managing Organizational Risk Knowledge. Journal of Universal Computer Science 9, 670–681 (2003)
7. Wong, K.Y., Aspinwall, E.: An empirical study of the important factors for knowledge-management adoption in the SME sector. Journal of Knowledge Management 9, 64–82 (2005)
8. Alhawari, S., Karadsheh, L., Talet, A.N., Mansour, E.: Knowledge-Based Risk Management framework for Information Technology Project. International Journal of Information Management 32, 50–65 (2012)
9. Massingham, P.: Knowledge risk management: a framework. Journal of Knowledge Management 14, 464–485 (2010)
10. Bergek, A., Norrman, C.: Incubator best practice: A framework. Technovation 28, 20–28 (2008)
11. Dahlstrand, A.L.: Technology-based entrepreneurship and regional development: the case of Sweden. European Business Review 19, 373–386 (2007)
12. Bollingtoft, A.: The bottom-up business incubator: Leverage to networking and cooperation practices in a self-generated, entrepreneurial-enabled environment. Technovation 32, 304–315 (2012)
13. Vick, T.E., Nagano, M.S., Santos, F.C.A.: Identifying the information management process and knowledge creation in technology-based companies: a Brazilian comparative case study. Knowledge Management Research & Practice, http://www.palgrave-journals.com/kmrp/
14. Wallace, L., Keil, M., Rai, A.: Understanding software project risk: a cluster analysis. Information & Management 42, 115–125 (2004)
15. Huang, S.-J., Han, W.-M.: Exploring the relationship between software project duration and risk exposure: A cluster analysis. Information & Management 45, 175–182 (2008)
16. Barros, M.O., Werner, C.M.L., Travassos, G.H.: Supporting risks in software Project management. The Journal of Systems and Software 70, 21–35 (2004)
17. Anantatmula, V., Kanungo, S.: Modeling Enablers for Successful KM Implementation. Journal of Knowledge Management 14, 100–113 (2010)
18. Fan, Z.-P., Feng, B., Sun, Y.-H., Ou, W.: Evaluating knowledge management capability of organizations: A fuzzy linguistic method. Expert Systems with Applications 36, 3346–3354 (2009)
19. Maguire, S., Koh, S.C.L., Magrys, A.: The adoption of e-business and knowledge management in SMEs. Benchmarking: An International Journal 14, 37–58 (2007)
20. Nonaka, I., Takeuchi, H.: The knowledge-creating company: how Japanese companies create the dynamics of innovation. Oxford University Press, New York (1995)
21. Verhaegen, T.: Knowledge makes risks manageable. Business Insurance: Industry Focus 3, 16–17 (2005)
22. Karadsheh, L., Alhawari, S., El-Bathy, N., Hadi, W.: Incorporating knowledge management and risk management as a single process. In: Proceedings of GBDI, USA, pp. 207–214 (2008)
23. Jafari, M., Rezaeenour, J., Mazdeh, M.M., Hooshmandi, A.: Development and evaluation of a knowledge risk management model for project-based organizations A multi-stage study. Management Decision 49, 309–329 (2011)

24. Forza, C.: Survey research in operations management: a process-based perspective. International Journal of Operations & Production Management 22, 152–194 (2002)
25. Bryman, A., Bell, E.: Business research methods. Oxford University Press, NewYork (2007)
26. Radas, S., Bozic, L.: The antecedents of SME innovativeness in an emerging transition economy. Technovation 29, 438–450 (2009)
27. Malhotra, N.K.: Pesquisa de Marketing: uma orientação aplicada. Bookman, Porto Alegre (2006)
28. Neves, S.M.: Análise de riscos em projetos de desenvolvimento de software por meio de técnicas de gestão do conhecimento. Federal University of Itajuba (UNIFEI), Itajuba (2010)
29. Yacoub, F., Macgregor, J.F.: Product Optimization and Control in the Latent Variable Space of Nonlinear PLS Models. Chemometrics and Intelligent Laboratory Systems 70, 63–74 (2004)
30. Helland, I.S.: On the Structure of Partial Least Saquares Regression. Communications in Statistics, Part B Simulation and Computations 17, 581–607 (1988)

Leveraging Knowledge from Different Communities Using Ontologies

Herlina Jayadianti[1,2], Carlos Sousa Pinto[1], Lukito Edi Nugroho[2], Paulus Insap Santosa[2], and Wahyu Widayat[2]

[1] Departemento de Sistemas da Informacao, Universidade do Minho, Campus de Azurem, Guimaraes, Portugal
[2] Electrical Engineering and Information Technology, Gadjah Mada University, Yogyakarta, Indonesia
[3] Economic Development, Gadjah Mada University, Yogyakarta, Indonesia
herlinajayadianti@gmail.com, csp@dsi.uminho.pt, {Lukito,Insap}@mti.ugm.ac.id, wahyu@mep.ugm.ac.id

Abstract. The purpose of this paper is to provide research based understanding of leveraging knowledge and managing knowledge within and across several communities using the poverty domain as a case study. We hypothesize that leveraging knowledge with a good taxonomy and a good integration process are good approaches to organize and share knowledge. Problems appear when a group of people in different communities share data and collaborate using different perceptions, different concepts, different terms (terminologies), and different semantics to represent the same reality. In this paper we present an approach to solve this problem. We will generate a common set of terms based on the terms of several different storage devices, used by different communities, in order to make data retrieval independent of the different perceptions and terminologies used by those communities. We use ontologies to represent the particular knowledge of each community and discuss the use of mapping and integration techniques to find correspondences between the concepts used in those ontologies.

Keywords: Leveraging Knowledge, Knowledge Management, Common Ontology, Perception, Terminology, Ontology.

1 Introduction – Leveraging Knowledge

Information technology has lead many institutions or communities to imagine a new world of leverage knowledge. Internet made it possible for professionals allowing them to draw on the latest thinking of their peers no matter where they are located. As a result many communities are rethinking how works gets done, linking people to electronic media so they can leverage each other's knowledge. Knowledge is different from information and sharing it requires a different set of concepts and tools. Four characteristics of knowledge distinguish it from information: [1]–[3]

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 67–76.
DOI: 10.1007/978-3-642-36981-0_7

- Knowing is a human act; whereas information is an object that can be filed, stored and moved around.
- Knowledge is created in present moment; whereas information is fully made and can sit in storage. To share knowledge we need to think about the current situation.
- Knowledge belongs to communities.
- Knowledge circulates through communities in many ways.

Leveraging knowledge involves a unique combination of human and information systems [3], [4]. Leveraging knowledge also allows us to tap into many innovations to access diverse knowledge bases and integrate them to create new competencies with multi-dimensional concepts [5], [6]. Ironically to leverage knowledge we need to focus on the community that owns it and the people who use it, not the knowledge itself. Six implications for leveraging knowledge are: [3]

- Different communities. Focus on knowledge important to both the business and the people. People naturally seek help, share insights and build knowledge in areas they care about.
- Sharing information. The ways to share knowledge should be as multidimensional as knowledge itself. Most corporate knowledge sharing efforts revolve around tools.
- Let the community decide what to share and how to share it. Knowledge needs to have an "owner" who cares. It is tempting to create organization-wide systems for sharing knowledge.
- Community support structure. Communities are held together by people who care about the community.
- Use the community's terms for organizing knowledge.
- Integrate sharing knowledge into the natural work flow.

Knowledge and learning are the only capabilities that can provide sustained competitive advantage. 'Knowledge' is the content of learning. 'Learning' is the process of gaining new knowledge, so that the firm is constantly accumulating and assimilating knowledge and this becomes the basis for creating and improving organizational routines [7]. Knowledge is a critical resource that warrants much more attention. If we are serious about managing knowledge, then we need to embrace the concepts associated with knowledge management [3], [6], [8]. Since knowledge is the sense we make of information, then the way information is organized is also a sense making device. A good taxonomy should be intuitive for those who use it.

To be "intuitive" it needs to tell the story of the key distinctions of the field, reflecting the natural way discipline members think about the field [3]. There are great temptations to make all systems for organizing knowledge the same, such as formatting information – to make it easily transferred, and having the same metadata – to make it easy searched, indexed and used in different context. However beyond that, the system for organizing information should be the community's. If a community of people sharing knowledge spans several disciplines, then such thing of terms and structures should be the *common* among those communities [3].

Having some *common* ground, among those communities, either within an application area or for some high-level general concepts, this could alleviate the problem of integrating knowledge [9]–[12]. Based on the presented reasons, we believe that ontologies with common terms and common concepts are very important in a knowledge sharing process. In this paper we describe an approach of leveraging knowledge using a common set of terms derived from several different ontologies. This paper is organized as follows: (1) Introduction; (2) Knowledge and Common ontology; (3) Implementation of the solution; (4) Conclusions.

2 Knowledge and Common Ontology

Figure 1 shows the relation between knowledge and ontologies.

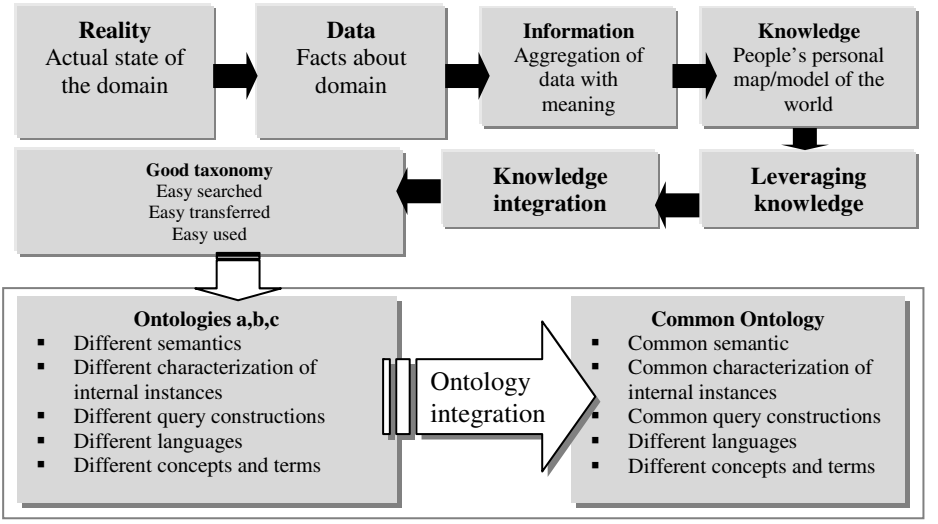


Fig. 1. Managing Knowledge, and Leveraging Knowledge with Ontologies

At the level "Reality" we represent the actual state of a particular domain. At this level we can find lots of data. Data are facts in the context of a domain of discourse. At the next level, establishing relationships between data, it is possible to derive information and expand it beyond the limits of understanding of each person. Knowledge is obtained by adding experience, reflection and reasoning to information. If different information is discussed by people, it is easy to understand what is inside their minds, either by arguments or communication, but what happens if those differences exist at the machine level? We need to combine information so that machines can "think" and understand the concepts we can find inside human brains. To do that, we can use ontologies to represent data and information of the several communities. Ontology is some formal description of a domain of discourse. However, ontology is not enough to make computers understand what is necessary. Scattered ontologies should then be incorporated and integrated into a new ontology,

a Common Ontology (See Fig.2). Ontology integration is one way to solve the problem of data, information, and semantic heterogeneity. Semantic heterogeneity on naming includes problems with synonyms (same concept with different terms) and homonyms (same term with different meanings). Semantic heterogeneity occurs when the same reality is modeled by two or more different people or systems [13].

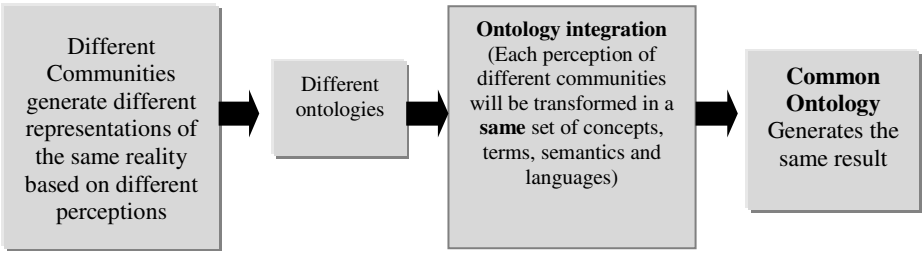


Fig. 2. Towards a Solution of Different Perceptions

The goal of ontology integration is to derive a more general domain ontology (Common Ontology) from other several ontologies in that domain. Every person has his/her own knowledge. They can justify everything based on their thoughts, perceptions and conceptualizations. Conceptualization is an abstraction of the external world inside an individual mind. It can be used to construct one or several concepts and also to interpret some reality in a conceptual way [14].

3 Implementing the Solution

Ontology integration is one way to solve the problem of semantic heterogeneity and it can be done using several approaches. For example, merging, matching or mapping. The integration of ontologies creates a new ontology by reusing other available ontologies through assembling [15]–[17], extending [18] , or specializing operations [19]. In integration processes the source ontologies and the resultant ontology can have different amounts of information [14]. Ontology process integration implies several steps. According to Noy [20] there are some specific challenges in the ontology integration process:

- Finding similarities and differences between ontologies in an automatic and semi-automatic way;
- Defining mappings between ontologies;
- Developing an ontology integration architecture;
- Composing mappings across different ontologies;
- Representing uncertainty and imprecision in mappings.

Particularly, in ontology integration, some tasks should be performed to eliminate differences and conflicts between those ontologies [20]. Ontology integration is used to find similarities and differences between ontologies. The goal of ontology

integration is to derive a more general domain ontology (Common Ontology) from several other ontologies in the same domain, into a consistent unit. The domain of both the integrated and the resulting ontologies is the same.

3.1 How to Get Common Terms – Common Ontology (CO)

Groups of people in different communities will probably have a different way of view to the same reality. Different view of each set of users is then called user view (UV) and can be implemented using an ontology. Common Ontology is expected to overcome the differences that exist in the several user views (UVs). CO will contain common terms which will then be equated with each term in the UVs. Common term is a common word recognized and used with the same meaning by different communities. To get the CO terms we use WordNet[1] , Thesaurus[2] and Swoogle[3] (See Table 1). Wordnet is a large lexical database or electronic dictionary for English [21], [22]. WordNet implements measure of similarity and relatedness among terms. Measures of similarity use information found in an *is–a* hierarchy of concepts, and quantify how much concept A is similar to concept B [23]. Thesaurus is a reference work that lists words grouped together according to similarity of meaning. Swoogle is the first Web search engine dedicated to online semantic data. Its development was partially supported by DARPA and NFS (National Science Foundation).

Table 1. Equivalences for some terms related to poverty from different applications

Search string	Synonym		
	Wordnet 2.1 (Noun)	Thesaurus (Noun)	Swoogle (Terms)
Hospital	Infirmary, medical institution	Clinic, emergency room, health service, hospice, infirmary, rest home.	Hospital, hospital,
Clinic	Medical institution, Session, Medical building, health facility, healthcare facility	Emergency room, hospice, infirmary, nursing home, rest home.	Clinic, Clinical, ClinicalTreatment

There are two senses for the term hospital in Wordnet (version 2.1).

Sense 1. *hospital, infirmary -- (a health facility where patients receive treatment)*
=> Medical building, health facility, healthcare facility -- (building where medicine is practiced)

Sense 2. *hospital -- (a medical institution where sick or injured people are given medical or surgical care)*
=> Medical institution -- (an institution created for the practice of medicine)

[1] http://wordnet.princeton.edu/
[2] http://thesaurus.com/
[3] http://swoogle.umbc.edu/
Swoogle was the first search engine dedicated to online semantic data. Its development was partially supported by DARPA and NFS (National Science Foundation).

There are three senses for the term clinic in Wordnet (version 2.1).

Sense 1. *clinic -- (a medical establishment run by a group of medical specialists)*
=> Medical institution -- (an institution created for the practice of medicine)

Sense 2. *clinic -- (meeting for diagnosis of problems and instruction or remedial work in a particular activity)*
=> Session -- (a meeting for execution of a group's functions)

Sense 3. *clinic -- (a healthcare facility for outpatient care)*
=> Medical building, health facility, healthcare facility -- (building where medicine is practiced)

3.2 Implementation

Let's consider the referred case study of poverty and two ontologies or user views both representing that domain. By using WordNet, Thesaurus and Swoogle we can find synonyms and similarities that were chosen from ontology UV1 and Ontology UV2. The next step is to find the number of each synonym with Google and Swoogle. The results provided by Google and Swoogle are different due to the number of documents that are available in each system. Google provides more documents than Swoogle. Currently, Swoogle only indexes some metadata about Semantic Web documents[4].

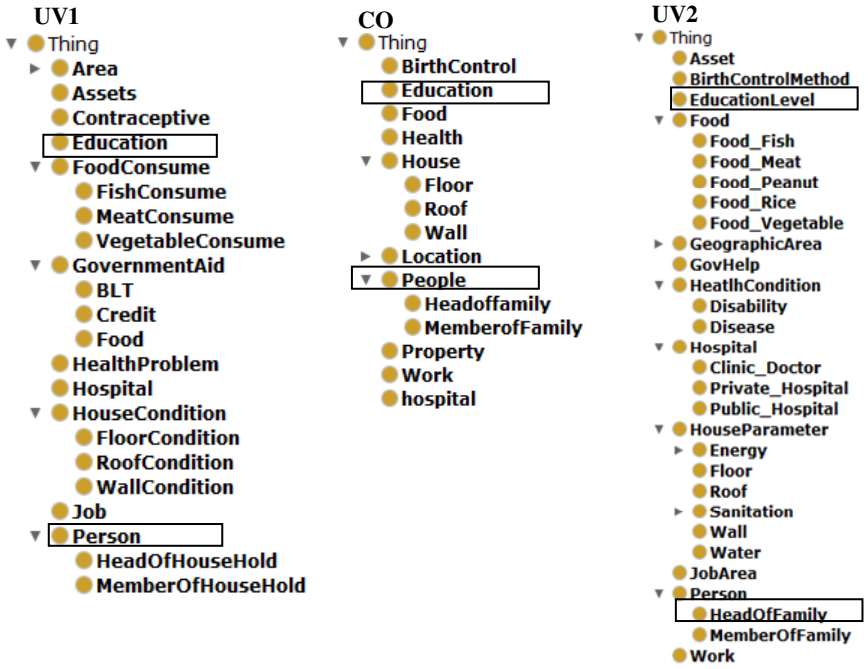


Fig. 3. Terms in ontology UV1, UV2 and CO

[4] `http://swoogle.umbc.edu/index.php?option=com_swoogle_manual&manual=faq`

Based on data from Swoogle and Google, then we selected the term with highest references. We assume that the number of references reflects a widely and commonly usage of the term by users. We use a common term as a term in CO. For example: term `Person` is more specific than `People` (See Fig.3). We use Wordnet, Swoogle and Google not only for comparing the number of result to get common terms but also to find common ObjectProperties. For example: ObjectProperties `hasFloorMadeFrom` (CO) is more general than `hasHouseFloorMadeFrom` (UV2) and `hasLargestFloorAreaMAdeFrom` (UV1) (See Fig. 4). Poverty is not the focus of our research. We just use that case as a real scenario that allows us to demonstrate our approach. We combine different existing terminologies about the same reality (poverty in this case) used by different communities in order to get a common set of terms that can be transparently used by those communities, while maintaining the original terms in the data sources. We use Indonesia as the country for the example because in that country there are several communities in charge of dealing with poverty data, generating problems due to differences in the criteria used to make their surveys, even considering that the semantics of these different criteria are the same. For example, let's consider the two communities, BKKBN[5] and BPS[6],that are responsible for collecting data on poverty. Each community has a different system and use different sets of terms to describe the same domain and

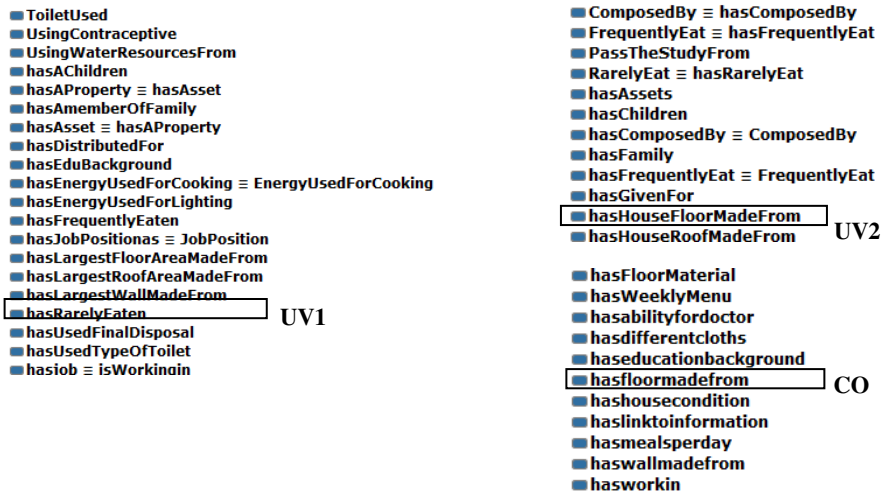


Fig. 4. ObjectProperties in ontology UV1, UV2 and CO

[5] Badan Keluarga Berencana Nasional (BKKBN) or National Population and Family Planning Board is a governmental agencies that appointed to conduct a survey of poverty in Indonesia. `www.bkkbn.go.id`

[6] Badan Pusat Statistik (BPS) or Central Berau of Statistic is a non departmental government institution directly responsible to the President of Indonesia. `www.bps.go.id`

different criteria to classify people as poor or not. To be similar *(≅)* or not equal *(≠)* depend on several factors, such as the programmer's interpretation, the needs of the system itself, and last but not least the domain/area that we are talking about. Currently, both communities are working separately to collect and manage data on poverty. Each community sends data to the government based on its perception.

3.3 Testing Queries

SPARQL[7] is a query for Resource Description Framework (RDF)[8]. SPARQL can be used to express queries across diverse data sources whether the data is stored natively as RDF or viewed as RDF via middleware. SPARQL contains capabilities for querying and also supports extensible values for testing and constraining queries [24]. SPARQL commands from our work shows below.

```
Prefix :<http://www.semanticweb.org/CO.owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?People ?Food ?Floor ?Location
WHERE { ?People :hasWeeklyMenu?Food. ?People :hasFloorMaterial?Floor.
?People :islivinginvillage?Location.
?Food :FoodName ?value1. ?Floor :Material ?value2. ?Location
:VillageName ?value3.
FILTER (?value1 ='Vegetable' && ?value2 ='Soil'  && ?value3
='Widodomartani')
}
Prefix : <http://www.semanticweb.org/UV1.owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?Person ?Food ?Job ?Floor ?Area
WHERE {?Person :RarelyEat ?Food. ?Person :hasJobPositionAs ?Job. ?Person
:hasHouseFloorMadeFrom ?Floor. ?Person :isLivinginVillage ?Area.
?Food :FoodName ?value1. ?Job :JobName ?value2. ?Floor :TypeOfFloor
?value3. ?Area :hasName ?value4.
FILTER (?value1 ='Chicken' && ?value2 ='Farmer' && ?value3 ='Soil' &&
?value4 ='Widodomartani')
}
```

ObjectProperties `hasHouseFloorMadeFrom` in UV1 is equivalent to ObjectProperties `hasFloorMaterial` in CO. ObjectProperties `hasFloorMaterial` is more common than ObjectProperties has `HouseFloorMadeFrom`. In this work, we found the same result these queries (See Fig.5).Our future work will include

[7] `http://www.w3.org/TR/rdf-sparql-query/`
[8] RDF is a standard model for data interchange on the web. `http://www.w3.org/RDF/`

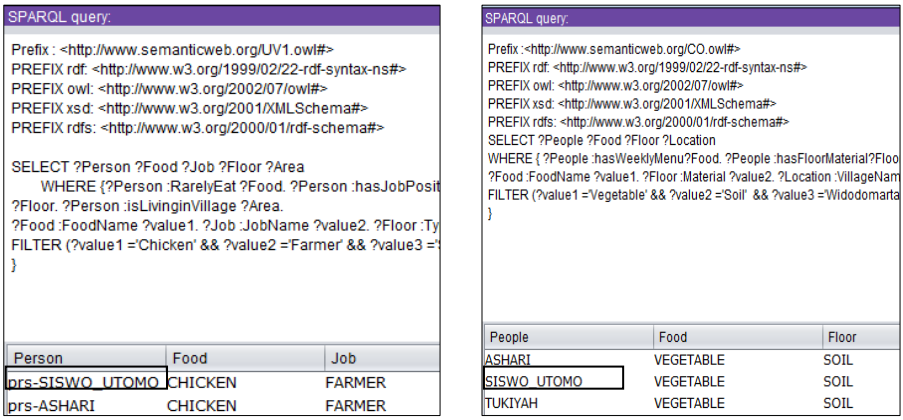


Fig. 5. The same result between ontology UV1 and CO

functionalities that will allow users ask queries using JSP[9] (Java Server Pages) and JENA[10] ontology API against OWL/RDF files. Through the ontology API, JENA provides a consistent programming interface for ontology applications.

4 Conclusions

In this research we try to leveraging knowledge by using an ontology integration as a process to create a new ontology (Common Ontology). Using this approach it is possible to share different conceptualizations, different terminologies, and different meanings between different systems. We believe that ontology integration is one of the best approaches to solve the problem of data and semantic heterogeneity.

Acknowledgment. We would like to acknowledge the support of the Erasmus Mundus EuroAsia program for the research foundation of this research, and also to acknowledge Universidade do Minho and Universitas Gadjah Mada for the collaboration.

References

[1] McDermott, R.: Knowing in community. IHRIM, 19 (2000)
[2] McDermott, R.: Knowing is a human act, Informatik-Informatique. Zeitschrift der Schweizerischen Informatikorganisationen 1, 7–9 (2002)
[3] McDermott, R.: Why information technology inspired but cannot deliver knowledge management. Knowledge and Communities 41(4), 21–35 (2000)

[9] Javaserver Pages is a technology provides a simplified, fast way to create dynamic web content.
http://www.oracle.com/technetwork/java/javaee/jsp/index.html
[10] http://jena.apache.org/

[4] Lesser, E.L., Fontaine, M.A., Slusher, J.A.: Knowledge and Communities. Routledge (2000)
[5] Hatteland, C.J.: Commentary on leveraging knowledge based resources: The role of contracts. Journal of Business Research 65(2), 162–163 (2012)
[6] Mudambi, R., Swift, T.: Leveraging knowledge and competencies across space: the next frontier in international business. Journal of International Management 17(3), 186–189 (2011)
[7] Carayannis, E.G.: Knowledge-driven creative destruction, or leveraging knowledge for competitive advantage strategic knowledge arbitrage and serendipity as real options drivers triggered by co-opetition, co-evolution and co-specialization. Industry and Higher Education 22(6), 343–353 (2008)
[8] Quinn, J.B.: Strategic outsourcing: leveraging knowledge capabilities. Sloan Management Review 40(4), 9–21 (1999)
[9] Euzenat, J., Shvaiko, P.: Ontology matching. Springer-Verlag New York Inc. (2007)
[10] Gangemi, A., Pisanelli, D., Steve, G.: Ontology integration: Experiences with medical terminologies. In: Formal Ontology in Information Systems, vol. 46, pp. 98–94 (1998)
[11] Noy, N.F.: Semantic integration: a survey of ontology-based approaches. ACM Sigmod Record 33(4), 65–70 (2004)
[12] Wache, H., Voegele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Hübner, S.: Ontology-based integration of information-a survey of existing approaches. In: IJCAI 2001 Workshop: Ontologies and Information Sharing, vol. 2001, pp. 108–117 (2001)
[13] Bouzeghoub, M., Goble, C.A., Kashyap, V., Spaccapietra, S. (eds.): ICSNW 2004. LNCS, vol. 3226. Springer, Heidelberg (2004)
[14] Xue, Y.: Ontological View-driven Semantic Integration in Open Environments. The University of Western Ontario (2010)
[15] Pinto, H.S., Martins, J.P.: A methodology for ontology integration. In: Proceedings of the 1st International Conference on Knowledge Capture, pp. 131–138 (2001)
[16] Kim, K.Y., Manley, D.G., Yang, H.: Ontology-based assembly design and information sharing for collaborative product development. Computer-Aided Design 38(12), 1233–1250 (2006)
[17] Farquhar, A., Fikes, R., Rice, J.: Tools for assembling modular ontologies in Ontolingua. In: Proc. of Fourteenth American Association for Artificial Intelligence Conference (AAAI 1997), pp. 436–441 (1997)
[18] Heflin, J., Hendler, J.: Dynamic ontologies on the web. In: Proceedings of the National Conference on Artificial Intelligence, pp. 443–449 (2000)
[19] Zimmermann, A., Krotzsch, M., Euzenat, J., Hitzler, P.: Formalizing ontology alignment and its operations with category theory. Frontiers in Artificial Intelligence and Applications 150, 277 (2006)
[20] Noy, N.F., Crubézy, M., Fergerson, R.W., Knublauch, H., Tu, S.W., Vendetti, J., Musen, M.A.: Protégé-2000: An Open-Source Ontology-Development and Knowledge-Acquisition Environment: AMIA 2003 Open Source Expo. In: AMIA Annual Symposium Proceedings, vol. 2003, p. 953 (2003)
[21] Fellbaum, C.: WordNet. Theory and Applications of Ontology: Computer Applications, 231–243 (2010)
[22] Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J.: Introduction to wordnet: An on-line lexical database*. International Journal of Lexicography 3(4), 235–244 (1990)
[23] Pedersen, T., Patwardhan, S., Michelizzi, J.: WordNet: Similarity: measuring the relatedness of concepts. In: Demonstration Papers at HLT-NAACL 2004, pp. 38–41 (2004)
[24] Pérez, J., Arenas, M., Gutierrez, C.: Semantics and Complexity of SPARQL. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 30–43. Springer, Heidelberg (2006)

An Approach for Deriving Semantically Related Category Hierarchies from Wikipedia Category Graphs

Khaled A. Hejazy and Samhaa R. El-Beltagy

Center for Informatics Science, Nile University, Cairo, Egypt
khaledhejazy86@yahoo.com, samhaa@computer.org

Abstract. Wikipedia is the largest online encyclopedia known to date. Its rich content and semi-structured nature has made it into a very valuable research tool used for classification, information extraction, and semantic annotation, among others. Many applications can benefit from the presence of a topic hierarchy in Wikipedia. However, what Wikipedia currently offers is a category graph built through hierarchical category links the semantics of which are undefined. Because of this lack of semantics, a sub-category in Wikipedia does not necessarily comply with the concept of a sub-category in a hierarchy. Instead, all it signifies is that there is some sort of relationship between the parent category and its sub-category. As a result, traversing the category links of any given category can often result in surprising results. For example, following the category of "Computing" down its sub-category links, the totally unrelated category of "Theology" appears. In this paper, we introduce a novel algorithm that through measuring the semantic relatedness between any given Wikipedia category and nodes in its sub-graph is capable of extracting a category hierarchy containing only nodes that are relevant to the parent category. The algorithm has been evaluated by comparing its output with a gold standard data set. The experimental setup and results are presented.

Keywords: Wikipedia, Semantic relatedness, Semantic similarity, Graph analysis, Category hierarchy, Hierarchy extraction.

1 Introduction

Wikipedia is an online encyclopedia that has more than 23,000,000 articles in which, more than 4 Millions articles are in English covering a wide variety of topics. Articles are maintained by more than 100,000 active volunteer contributors. As Wikipedia is written collaboratively by anonymous volunteers, anyone can write and change Wikipedia articles. It is assumed that contributors will follow a set of policies and guidelines developed by the Wikipedia community. However, there is nothing in place to enforce editing policies before or during contributing[1] which means that breaches to Wikipedia's policies and guidelines are being conducted by its community, greatly affecting its quality.

[1] http://en.wikipedia.org/wiki/Wikipedia

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 77–86.
DOI: 10.1007/978-3-642-36981-0_8

Just like articles, Wikipedia's categories are socially annotated. When creating new categories and relating them to previously created ones, there is no strict enforcement of which higher-level categories a child sub-category can belong to; thus, Wikipedia's category structure is not a tree, but a graph in which links between nodes, have loosely defined semantics.

Consequently a sub-category in Wikipedia does not necessarily comply with the concept of a sub-category in a hierarchy. A category label in Wikipedia is simply intended as a way for users to navigate among articles, and only signifies that there is some sort of a relationship between the parent category and its sub-category that is not necessarily of the type "is-a" which is expected in a hierarchical Knowledge Organization System. This problem causes irregularity in semantics between categories that is amplified in deeper levels. For example, following the category of "Computing" down its sub-category links, the totally unrelated category of "Theology" appears. Also, the graph nature of the Wikipedia category structure means that following the sub-category links of any given category, can eventually lead back to the same category. Detecting and eliminating cycles is a minor issue. Detecting sub-categories that should be considered as belonging to any given category is the main challenge addressed by this work. To address this challenge, an approach for measuring lexical semantic relatedness between Wikipedia's categories and nodes in their sub-graphs and using this as an indicator for relatedness, was developed.

In this paper, we introduce this new approach for deriving semantically related category hierarchies from Wikipedia category graphs and extracting a category hierarchy containing only sub-categories that are relevant to the parent category.

The rest of the paper is organized as follows; firstly, related work is presented in section (2), the proposed approach is described in section (3), the procedure followed for evaluating our approach and the experimental results are presented in section (4). Analysis of the results is discussed in section (5). And finally section (6) concludes this paper.

2 Related Work

Since its inception, Wikipedia has undergone tremendous growth, and today it is the largest online encyclopedia known to date. Wikipedia has been widely used as a huge resource of concepts and relationships for text mining tasks; like classification, information extraction, and computing semantic relatedness of natural language texts, among others. Most research works that make use of Wikipedia have used Wikipedia's concepts and relationships as is, except for some preprocessing and slight modifications. No previous research (as far as the authors are aware) addressed semantic irregularity between categories in Wikipedia's categorization system.

Wikipedia's categories' growth has previously been analyzed in [1], where an algorithm that semantically maps articles by calculating an aggregate topic distribution through the articles' category links to the 11 top Wikipedia categories (manually selected). Semantic relatedness for category nodes is then calculated through link distance metrics, such as the length of the shortest path between two nodes.

The evolution of Wikipedia's category structure over time has been studied in [2]. Results of this research have shown that the Wikipedia category structure is relatively stable for a bottom-up evolved system. However, the work did not address the accuracy of the category structure.

Wikipedia has been used for measuring lexical semantic relatedness between words or text passages. Explicit Semantic Relatedness (ESA) [3] has been shown as a successful measure for semantic relatedness. It treats each Wikipedia article as a dimension in a vector space. Texts are then compared by projecting them into the Wikipedia articles' space, then measuring the similarity between vectors using conventional metrics like cosine similarity. Because this work relies mostly on individual articles, the category structure of Wikipedia was not an issue.

Wikipedia has been used to compute semantic relatedness by taking the categorization system of Wikipedia as a semantic network [4].

Wikipedia Link-based Measure [5] also measures the semantic similarity of two Wikipedia pages by comparing their incoming and outgoing links. The score is determined using several weighting strategies applied to the overlap score of the articles' links.

In this paper, we propose an approach for deriving semantically related category hierarchies from Wikipedia category graphs. Our approach is somehow similar to ESA, except the fact that we are measuring semantic relatedness between categories instead of articles or words. Also, we use a key-phrase extraction for dimensionality reduction.

3 Methodology

Detecting semantically related categories based on measuring lexical semantic relatedness between them requires an efficient representation for each category. A TF-IDF scheme [6] has been used to assign weights to the feature vectors representing Wikipedia categories. In the following subsections, we start with the pre-processing step; in which we discuss the data sources with their components and the pre-processing steps conducted before these data are used, and then we discuss the steps of generating the feature vectors of Wikipedia categories.

3.1 Pre-processing

Wikipedia's backups are created regularly by the Wikimedia Foundation[2]. These dumps are publicly available. We have used Wikipedia's XML dump release 02-05-2012, which contains all Wikipedia article pages. The size of the uncompressed dump is around 38 GB.

Pages in this xml dump are represented by multiple tags. From those our system uses the page's unique ID, page's title, page's time stamp, and page's text.

[2] www.wikimedia.org

In order to handle this large XML dump file, apache Solr [7] has been used to index it through its Data Import Handler, which also facilitated the searching processes required by our followed approach.

We also acquired some relevant SQL files from the same source in order to allow us to re-construct the categorization graph of Wikipedia. One of these files is the enwiki-20120502-category.sql.gz which is an SQL file containing metadata for each category in Wikipedia; its category ID (differs from the page ID), its title, and the number of its pages and subcategories. The other is 20120502-categorylinks.sql.gz which is an SQL file has been acquired, and used for building Wikipedia's categorization graph. The SQL file contains the page IDs of any page defined as a category member, the page title of the category's description page, the time stamp of the approximate addition time of the link, the category link type that determines whether the page ID is a page, a sub-category or a file, along with some other attributes for sorting and for defining the collation of the category links.

Pages in Wikipedia are not only articles; categories are special pages that are used to group articles together, and their titles start with the namespace "Category:". Also, there are administrative pages with different namespaces that are not used to share encyclopedic information, but rather to preserve policies created and implemented

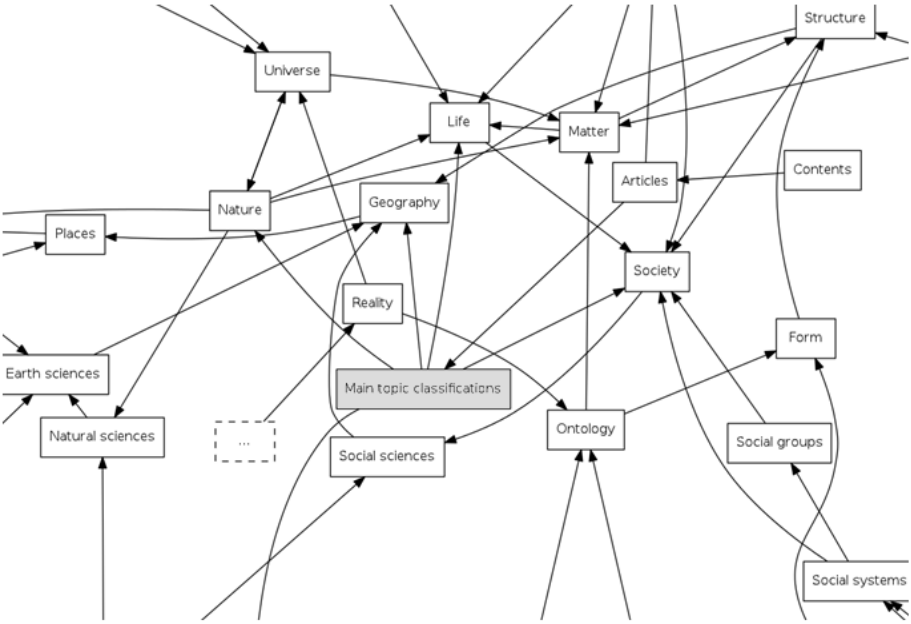


Fig. 1. A sub-graph in Wikipedia showing the category "Main topic classifications" and some of its super-categories and sub-categories

by user consensus for editing Wikipedia, such as User, User talk, File, File talk, Template, Template talk, Help, Help talk, Portal, Portal talk, etc.[3]. A cleanup step has been conducted for removing all administrative pages from the XML dump, leaving behind only article and category pages.

The Wikipedia category system does not offer top level categories. So, the category "Main topic classifications" was selected as our root category, the sub-categories of which are topical categories, indirectly contain almost all Wikipedia articles. Figure 1 shows an example of the category "Main topic classifications" and some of its parent and children categories. The parents, as well as the sub-categories, have many other children that are not shown in the figure.

3.2 Algorithm

The approach that is followed in this work is based on the observation that each domain or category has its own unique vocabulary. For example, the terms {north, region, island, earth surface, river, geography, south, geographical, urban, landscape, map, spatial, earth, sea, etc} collectively represent a subset of terms that are often used in the context of geography. Sub-categories that are in fact the hierarchical descendants of any given category are highly likely to share the same vocabulary and use the same terminology. Direct sub-categories are more likely to be related to their parent categories than their 2nd level or other deeper level descendants.

Basically, the relationship between a root category and its descendents grows weaker as we go deeper down the tree. Having said that, even first level sub-categories in Wikipedia can sometimes be un-related to their parent category in the hierarchical sense. The assumption made by this work, is the categories that are directly related to some parent category, will most likely share a reasonable part of its vocabulary. A category that has an entirely different vocabulary is not likely to be a hierarchical decedent of its parent category, even if some other relationship exists between the two. To build on this observation, two steps had to be followed:

1. Modeling the vocabulary for a given category
2. Measuring semantic relatedness between categories

Modeling the Vocabulary of a Category

In order to make use of a category's vocabulary, the first step that needed to be carried out was to capture the vocabulary of Wikipedia categories. In Wikipedia, each category has both a set of direct pages and a set of direct descendents or sub-categories. The set of pages is sometimes very small, or non-existent, making it very difficult to model the vocabulary of the category based entirely on these. First level categories, while often related to a category, also often have some noise. In the proposed approach, both a category's pages and its immediate sub-categories count towards building the category's vocabulary by extracting key-phrases from both. In the context of this work, key-phrases are defined as a list of terms each of which is made up of one or more words and that describe the "sub-category" with which they are associated.

[3] http://en.wikipedia.org/wiki/Wikipedia:Administration

The KP-Miner system [8] was the system used to extract key-phrases. When applied to a group of Agricultural documents, 66% of the key-phrases extracted from those, were found to correspond to concepts in the Agricultural Ontology known as AGROVOC4 and 90% were found to be directly related to the field of Agriculture [8]. The system which was designed to be generic relies on term frequency information gathered from a large corpus of random pages. To make it more relevant to Wikipedia, the system was re-trained using Wikipedia's articles, in the sense that term frequency information was obtained based on term occurrences within Wikipedia. The system was then used to extract the top n-key-phrases from a given text in the manner described below.

In order to model the vocabulary of a category, a preliminary step of extracting n key-phrases from each of Wikipedia categories' direct pages (if possible) is conducted. To extract the key-phrases, all pages are concatenated and treated as a single document; extracting n key-phrases referred to as "*Pages_Keyphrases*", which has been indexed and stored in a multi-valued text field in Solr. The number of key-phrases extracted to from the *Pages_Keyphrases* of each of Wikipedia categories is fixed as a constant $\boldsymbol{n}$ ($\boldsymbol{n}$ = 300), regardless of the size of a category, to prevent larger categories from biasing the model of their parent category (in the future, we intend to experiment with different values of $\boldsymbol{n}$).

In order to build the representative feature vector for each of Wikipedia's category (referred to as "*Category_Keyphrases*"); direct pages of each category are considered as the most important resource for representing the category; that's why *Pages_Keyphrases* are fully included in the *Category_Keyphrases*. Also the direct sub-categories' *Pages_Keyphrases* were used for constructing a Category's *Category_Keyphrases* vector in a way that amplifies the common concepts among sub-categories, and excludes noise that can appear in any of them.

Each key-phrase obtained from each sub-category's *Pages_Keyphrases* can actually be thought of as a single vote for this key-phrase. Only key-phrases with votes greater than some value $\boldsymbol{m}$ (obtained from all 1st level subcategories), are included in the *Category_Keyphrases* vector of the parent category along with those of its *Pages_Keyphrases*, which then serves as a representative for that category.

Table 1 and Table 2 show the extracted *Pages_Keyphrases*, and *Category_Keyphrases* for the category "Islands". To calculate the weight of each key-phrase, both its frequency and its IDF factor are used.

Table 1. Sample of the key-phrases obtained for the *Pages_Keyphrases* of the category "Islands"

Stemmed Key-phrases		
island	unsinkable aircraft carrier	islet
floate island	island ecosystem	new zealand
coral reef	island restoration	reef
artificial island	private island	high island
unsinkable aircraft	low island	oceanic island

Table 2. Sample of key-phrases of the *Category_Keyphrases* for the category "Islands"

Stemmed Key-phrases		
island	isle	south
sea	archipelago	pacific
area	pacific ocean	coral reef
population	reef	indian ocean
map	ocean	sea level

In classical information retrieval models, the frequency of a term is calculated as the number of times it appears in a document. This is often normalized by dividing that number by the total number of terms that appear in the same document. In our proposed approach, a category is treated as a single document. The frequency of a key-phrase is calculated as the total number of times that this key-phrase has occurred across its sub-categories and its pages, and the weight is determined by multiplying this value with the IDF value obtained across all obtained key-phrases from Wikipedia.

Measuring Semantic Relatedness between Categories

After obtaining all feature vectors for all Wikipedia categories, building a hierarchical tree for any category becomes possible. To build such a tree for any category, its subcategories are traversed in a depth first fashion in order to accept or reject as hierarchical descendents of the category in question.

A subcategory is said to be accepted if the cosine similarity of its vector and that of the main category under consideration is greater than an empirically calculated threshold Ω.

4 Evaluation

4.1 Building the Evaluation Dataset

Humans have the natural ability to disambiguate topics and judge their relatedness. In order to evaluate our algorithm, a test dataset of 1000 categories has been randomly collected from the sub-graph of the category "Geography" in Wikipedia. Each instance represents a Wikipedia category that may or may not be considered as a semantically related sub-category to the main category being tested ("Geography" in our case). The test dataset was then manually annotated by 3 different human judges; determining whether or not semantic relatedness exists between each of the testing sub-categories and the main category being tested. The final manual annotation for each instance was determined by taking the consensus annotation represented by having the majority votes of the 3 judges. The resulting dataset[4] was used as a gold standard.

[4] The dataset is available upon request, and it will be available shortly on our website `http://tmrg.nileu.edu.eg/`

4.2 Results

The developed system was used to derive the hierarchical tree of the Geography category and the results were compared with the gold standard dataset described in the previous section. Table 3 shows the different results of the algorithm when compared against the gold standard dataset using different values for the semantic relatedness threshold Ω.

Table 3. Evaluation of the algorithm with different thresholds for Ω

#	Ω	Precision	Recall	F-Score
1	0.076	0.51078167115903	0.844097995545657	0.636439966414778
2	0.086	0.519662921348315	0.824053452115813	0.637381567614126
3	0.096	0.535871156661786	0.815144766146993	0.646643109540636
4	0.11	0.544753086419753	0.78619153674833	0.643573381950775
5	0.12	0.547049441786284	0.763919821826281	0.637546468401487
6	0.13	0.552238805970149	0.741648106904232	0.633079847908745
7	0.14	0.573476702508961	0.712694877505568	0.635551142005958
8	0.146	0.5893536121673	0.690423162583519	0.635897435897436

It was found that setting the threshold Ω to be 0.096 gives the highest F-score value. Thus, the following analysis section focuses on analyzing the results while setting Ω to be 0.096.

5 Analysis

As shown in table 3, there is a tradeoff between the precision and the recall; increasing the threshold results in increased precision and decreased recall, and vice versa.

Tables 4, and 5, show a sample of discrepancies between results of the presented system and manually annotated data, while tables 6, and 7 show samples of agreement between the two. The term "ACCEPTED" is used for indicating that the developed system has concluded that the sub-category in question is semantically related to the category being tested ("Geography" in this case) and that it should be part of its subtree, while the term "REJECTED" is used when there is no semantic relatedness.

Looking at table 4, and taking the category "Mountaineering" [5] as an example, it is easy to see why this category was mistakenly accepted. Mountaineering is a sport, or a hobby of mountain climbing, however, there is an overlap between the "Mountaineering" concept and the geographical concepts like "climbing mount Everest that is located in somewhere between China and Nepal". Taking another example of category "Paços de Ferreira" [6] from table 5, to examine why this category was rejected, when it should have been accepted, we find that this particular category does not have any sub-categories and only 3 pages the textual content of which is too poor to extract meaningful and sufficient key-phrases from.

[5] http://en.wikipedia.org/wiki/Category:Mountaineering
[6] http://en.wikipedia.org/wiki/Category:Paços_de_Ferreira

Table 4. Examples of categories detected as ACCEPTED while they are manually annotated as REJECTED

Category	Depth	Cosine Similarity
Mountaineering	5	0.2246
Women who reached the Poles	4	0.177
Lists of buildings and structures	3	0.154
Telecommunications infrastructure	4	0.17956
Baltimore City College	5	0.099

Table 5. Examples of categories detected as REJECTED while they are manually annotated as ACCEPTED

Category	Depth	Cosine Similarity
Underground cities	4	0.056
Ramsar sites in Israel	5	0.07657
Protected areas of the Republic of the Congo	5	0.0477
Paços de Ferreira	5	0.0154
Kronoberg County	5	0.0604

Table 6. Examples of categories detected as ACCEPTED and manually annotated as ACCEPTED

Category	Depth	Cosine Similarity
Geography of Austria	5	0.16888
Pas-de-Calais	5	0.156
Barnsley	5	0.14856
Brighton and Hove	5	0.1453
Algarve	5	0.15865

Table 7. Examples of categories detected as REJECTED and manually annotated as REJECTED

Category	Depth	Cosine Similarity
The Chronicles of Narnia music	5	0.0372
Yorkville University	5	0.0339
People from the Azores	5	0.0716
Science and technology in Uganda	5	0.04715
Health in Cyprus	5	0.01122

6 Conclusion

This paper presented a novel approach for deriving semantically related category hierarchies from Wikipedia category graphs. Future work will focus on refining the developed methodology so as to improve both precision and recall.

This approach is being applied within an ongoing project to generate a semantically related category hierarchy for collecting statistics on Wikipedia categories (where a category refers to an entire hierarchy) based on their number of pages, language instances, in-links, and out-links, among others. The statistics generated based on this hierarchy are supposedly more real than those generated from Wikipedia's category system.

Acknowledgements. This work was partially funded by Microsoft's Advanced Technology Lab in Cairo, grant number CIS-001-1011. Special thanks to members of NU's Text Mining research group for their insightful comments and help.

References

[1] Kittur, A., Chi, E.H., Suh, B.: What's in Wikipedia?: mapping topics and conflict using socially annotated category structure. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1509–1512 (2009)
[2] Suchecki, K., Salah, A.A.A., Gao, C., Scharnhorst, A.: Evolution of Wikipedia's Category Structure. Advances in Complex Systems 15 (2012)
[3] Gabrilovich, E., Markovitch, S.: Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence, pp. 1606–1611 (2007)
[4] Strube, M., Ponzetto, S.P.: WikiRelate! Computing Semantic Relatedness Using Wikipedia. In: Proceedings of the 21st National Conference on Artificial Intelligence, vol. 2, pp. 1419–1424 (2006)
[5] Milne, D., Witten, I.H.: An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. Association for the Advancement of Artificial Intelligence (2008)
[6] Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Information Processing & Management 24(5), 513–523 (1988)
[7] Apache Solr (2012), `http://lucene.apache.org/solr/`
[8] El-Beltagy, S.R., Rafea, A.: KP-Miner: A keyphrase extraction system for English and Arabic documents. Information Systems 34(1), 132–144 (2009)

Privacy Policies in Web Sites of Portuguese Municipalities: An Empirical Study

Gonçalo Paiva Dias[1,3], Hélder Gomes[1,4], and André Zúquete[2,4]

[1] School of Technology and Management of Águeda, University of Aveiro,
Rua Comandante Pinho e Freitas, 28,
3750-127 Águeda, Portugal
{gpd,helder.gomes}@ua.pt
[2] Department of Eletronics, Telecomunications and Informatics,
University of Aveiro, Campus Universitário de Santiago,
3810-193 Aveiro, Portugal
andre.zuquete@ua.pt
[3] Research Unit on Governance, Competitiveness and Public Policies,
University of Aveiro, Campus Universitário de Santiago,
3810-193 Aveiro, Portugal
[4] Research Institute of Electronics and Telematics Engineering of Aveiro,
Campus Universitário de Santiago,
3810-193 Aveiro, Portugal

Abstract. In this article we present an empirical study of privacy policies at local government web sites in Portugal. Web sites of all the 308 Portuguese municipalities where surveyed for privacy policy manifests and a content analysis was performed on those documents. We concluded that privacy of citizens is clearly not a priority in Portuguese local government, since only 26% of the municipalities include privacy policy statements in their web sites and many of those documents exhibit significant gaps considering the information that a user would expect to find in a privacy policy. By raising the awareness to the need of improving privacy in local government web sites, our conclusions are useful for both academics and practitioners in the areas of e-government, privacy and local government

Keywords: privacy, security, e-government, local government, Portugal.

1 Introduction

Privacy is an important and growing concern when accessing information and services in the Internet. More and more personal data is collected everyday by web sites and service providers. This data can potentially be aggregated and used to build personal profiles, raising the fears that it can be used for unauthorized purposes that may affect its owners.

In 1995, the European Parliament and Council launched the Directive 95/46/CE on the protection of individuals with regard to the processing of personal data and on free

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 87–96.
DOI: 10.1007/978-3-642-36981-0_9

movement of such data [1]. In 1996, this directive was transposed to the Portuguese legislation by means of the personal data protection law [2]. Among other aspects, this law imposes that the gathering, processing and communication of personal data to third parties must be known by its owner and subject to his/her explicit authorization.

Although not mandatory by law, privacy policies became a common practice to inform web site users about the use that is made of their personal data. A privacy policy is a method for increasing users' trust on the privacy practices of an organization [3]. It is a written description of how privacy is handled by the organization it pertains to. It arises as a mean to overcome the difficulty on gaining consumers' confidence on commercial electronic transactions (e-commerce). Other methods exist, like third-party seals, but the use privacy policies become the most widely used. The simple presence and visibility of a link to a privacy policy statement may positively impact users' perception of trust, even when readers don't read it [4].

In 2007 Portugal was rated 79 out of 100 in the privacy policy feature of the Global e-Government study [5], which addressed the existence of privacy policies' statements in central government web sites, among other aspects. In this article we present an empirical study of the privacy policies advertised by Portuguese municipalities' web sites.

Privacy is a main concern when e-government is at stake [6-9]. This is somewhat a consequence of the trust in government being a key factor for e-government adoption [10-12]. Taking this into consideration, we argue that the adoption and correct advertising of privacy policies in government web sites, including at the local levels of government, is fundamental to sustain e-government use and development.

The scientific study of local e-government in Portugal is still in its infancy. In effect, just a few articles have been published internationally on the subject [13-21], none of which has directly addressed the adoption of privacy policies and their advertising in municipal web sites. To our knowledge, this article is the first one to address this subject in Portugal. With the present study we aim to contribute to raise the awareness of privacy policies in web sites of local authorities in Portugal by establishing the current state of affairs concerning their adoption and advertising. The conclusions of the study are useful for both academics and practitioners in the areas of e-government, privacy and local government.

There are 308 municipalities in Portugal, with an average population of 38,000 inhabitants (ranging from 500 to 470,000) and an average area of 400 Km^2 (ranging from 8 to 1,700 Km^2). Municipalities have political, administrative and financial autonomy and are governed by an assembly and a city council (which is led by a mayor). Excluding the Azores and Madeira archipelagos (which have regional governments), municipalities are in Portugal the single relevant form of political representation apart from central government.

The remaining of this article is organized as follows: in Section 2 we present similar studies performed by other authors; in Section 3 we present the methods used; in Section 4 we present and discuss the results; and in Section 5 we draw the general conclusions and future work.

2 Related Work

As already mentioned, Portugal achieved a high score (79%) in the ranking of privacy policy feature of the Global E-Government study [5]. Besides the existence of a privacy policy, the following particular aspects of the privacy policy were assessed: if the privacy statement prohibits commercial marketing of visitors' information; the use of cookies or individual profiles of visitors; disclosure of visitors' personal information without the prior consent of visitors, or disclosure of visitors' information to law enforcement agents. However, this study focused on the central administration web sites and excluded the local administration. This raises the question if the local administration is following the example of good privacy practices that comes from the central administration in Portugal.

On a 2007 assessment of European municipal web sites, which analyzed the web sites of the 80 most populated European cities, it was found that only 55% of them included a privacy policy [22]. However, this number cannot be extrapolated to the Portuguese reality because only one Portuguese city (Lisbon) was considered in that study. In the UK, a 2009 study found that a large number of web sites, in a random sample of 54 gathered from the "direct.gov.uk" domain (which also includes municipalities), didn't have a privacy policy and that the existing privacy policies were not consistent among them [23]. Another study, from Saudi Arabia and published in 2012, found that only 28% of the e-government sites of Saudi-Arabia presented a privacy policy and that 60% of these had a "strong status" on the privacy statement [24]. On this study, the privacy statement status was evaluated according to the United States Federal Trade Commission's Fair Information Practical Principles (FIPs).

The studies [25-27] analyze the content of privacy policies from a mix of private and public sectors, but not municipalities. In all three studies the results were focused on three main topics: general characteristics of the policy, data collection and data sharing. Since the studies were from different countries the results differ, but they share an important conclusion: there is a lot to do either in terms of companies' compliance with their policies and in terms of informing an effective and readable policy.

Based on the Dutch Personal Data Protection Act (Dutch implementation of the European Parliament and Council Directive 95/46/CE) privacy policies in Dutch municipal web sites were analyzed to find answers for two research questions: "*What are the guarantees contained in the privacy statements of Dutch web sites?*" and "*How easy (or difficult) is to find the privacy statements on Dutch municipal web sites?*" [28]. The results demonstrated that privacy statements are not well aligned with the Dutch Personal Data Protection Act and that municipalities do not pay much attention to making privacy statements observable in their web sites.

In Table 1 we compare the privacy policy features accessed in three of the above mentioned studies: the study on privacy policies in Saudi Arabia public administration web sites [24], the study on privacy policies in generic organizations web sites [25] and the study on privacy policies in Dutch municipalities [28]. As we can observe from the table, the privacy policies' features accessed by the three studies are mainly

the same, but with different levels of detail: The presence/visibility of the privacy policy; whether or not the privacy policy informs the user that data is collected, which data is collected, the reason for the collection and the processing it will be subject to; whether or not the collected data will be disclosed to third parties and if the user can oppose to that disclosure; whether the user has access to his/her data; and whether or not it assures the existence of measures to promote data security. This set of features will be used as a basis for our methodology.

Table 1. Privacy policy features accessed in three selected studies

Type of analysis	[Alhomod]	[McRobb]	[Beldad]
Presence/Visibility	Policy privacy present		Are privacy policies available?
		Prominence	and findable?
Content Analysis	Notice/Awareness of data collection	Is personal data collected?	Collection of data related to web site visit
		What data?	Notification of the purposes for data collection
		Why?	Storage and retention of collected data
			Data processing and use
	Choice/consent to disclosure	Is personal data disclosed?	Disclosure of personal data to third-parties
		Why?	
		User can oppose?	
	Access/ Participation	Check/Amend procedure stated?	User's right of access to their data
	Integrity / Security	Data protection methods stated?	Security of personal data
		Cookies used?	Notification of usage of cookies
		Children issues addressed?	
		Further privacy advice offered?	Contact possibilities for enquiries regarding the privacy policies
			Revisions in the privacy statements
			Catching user trust right at the start
			Collection of special personal data

3 Methods

We used the observation of web sites and conceptual analysis, a subcategory of content analysis, to collect empirical data. Content analysis and specifically web content analysis is a known method in e-government research [29]. More specifically, a two-step method was developed and used:

- Web sites were observed to identify the existence of privacy policy manifests and how prominent they are;
- When privacy policy manifests were available, conceptual analysis was used to categorize the information present in those manifests.

In the first step, the following items were observed and registered:

- Availability of a privacy policy manifest;
- The link or links to the manifest page include the word 'privacy';
- Pages that link to the manifest (all pages; homepage; others than homepage);

In the second step, the presence of the following concepts (categories) in the privacy policy manifests was observed and registered:

- *Information collection* – refers to the existence of references to users' personal information being collected by the web site, whether explicitly or not;
- *Collected information* – refers to the information items that are collected from the user by the web site;
- *Information use* – refers to the use that is made of the information collected from the user;
- *Information disclosure* – refers to whether or not the information is disclosed to third parties;
- *Disclosure purpose* – refers to the purpose of the disclosure of information to third parties;
- *Opt-out option* – refers to the availability of an option for the user to oppose to the disclosure of information;
- *User access* – refers to the mechanisms available for the user to access and manipulate his/her own information;
- *Security measures* – refers to the measures that are taken to prevent unauthorized access to the user's information;
- *Use of cookies* – refers to the use of cookies by the web site to track users along several interactions;
- *Additional information* – refers to the existence of a contact where the user can ask for additional information concerning the privacy policy;
- *Child privacy* – refers to the existence of specific information concerning child privacy.

For *Information collection*, *Information disclosure*, *Opt-out option* and *Use of cookies*, the number of affirmative answers was also recorded. Additionally, specific cases were documented to illustrate typical or abnormal situations concerning the content of privacy policy manifests. These cases were used to enrich the discussion in Section 4.

Data was collected in 2012, from March to May. The observations were made by six different teams. NUTS III (third level of Nomenclature of Units for Territorial Statistics – geocode standard for referencing the subdivisions of countries for statistical purposes in the European Union) was used to assign the municipalities analyzed by each team. Three to seven NUTS III were assigned to each team (26 to 67 municipalities). A spreadsheet was filled out for each municipality. Harmonization

meetings were organized with all the teams and the data was verified by an independent team to ensure reliability. Data was later aggregated into a single database in order to calculate global results.

4 Results and Discussion

According to the results of the empirical study, only 26% (80 out of 308) of the Portuguese municipalities advertise their privacy policies by means of a web page in their web site (see Figure 1). Note that all Portuguese municipalities have functional web sites. This is a very poor result, when comparing, for example, with the result of 77% for the Dutch municipalities [28], and demonstrates that privacy of citizens is not a priority in Portuguese local government. Furthermore, we found that 5% (17) of the municipal web sites have links pointing to inexistent privacy policies (and, consequently, were not considered as having a privacy policy statement), and that 20% (16) of the total of links pointing to privacy statements do not include the privacy word.

Considering where the privacy policy page is linked from (broken links included) links appear in all the pages of the web site for 72 municipalities (74% of the municipalities having multiple pages) and in pages inside the web site other than the home page for 25 municipalities (26%), as depicted in Figure 2.

Fig. 1. Web sites with privacy manifest; without privacy manifests; with the word "privacy" in the link to manifest; and with broken links to manifest

Fig. 2. Where the privacy manifest is referred from in the web site

Table 2 resumes the results of the conceptual analysis. Besides the occurrence of the selected categories in privacy policy manifests, affirmative answers relating the scope of those categories are also presented, whenever applicable. From the results, it is noticeable that only 95% of the manifests refer the gathering of users' personal information by the web site, which is precisely the primary reason to justify the need for a privacy policy. Typically the other 5% have pages with Privacy Policy title but with contents related to subjects other than privacy, like copyrights, for example.

Table 2. Occurrence of content analysis' categories in the 80 web sites with privacy policy statements; and correspondent affirmative answers whenever applicable

Category	occurrences	percentage of occurrences	affirmative answers	percentage of affirmative answers
Information collection	76	95%		
Collected information	13	16%		
Information use	51	64%		
Information disclosure	65	81%	7	11%
Disclosure purpose	7	100%	4	57%
Opt-out option	5	71%	5	100%
User access	49	61%		
Security measures	48	60%		
Use of cookies	16	20%	16	100%
Child privacy	8	10%		

Regarding the collection of users' data, practically all of them warn users about the data that is necessary to apply for some service, with only a few exceptions referring the collection of data from the navigation in the pages of the web site, like IP addresses, pages visited, etc. Furthermore, only 64% refer the use that is made of the information gathered and only 16% refer the information items that are collected. This reinforces the previous conclusion of privacy not being a priority for Portuguese municipalities, since many of the manifests do not address some of the most important topics that users would expect to find in a privacy policy. One strange example that illustrates the low importance devoted to this subject is a privacy policy where the name of the municipality is never mentioned, appearing the name of a fictional town instead (*Vila Feliz* – Happy Town).

Besides *Information collection*, the second most relevant category is *information disclosure*: 81% of the manifests refer to whether or not the information they collect is disclosed to third parties. Concerning this aspect, it is worth noting that only 7 of the 65 municipalities that refer this category state that information may be disclosed and, of those, only 4 refer the disclosure purpose (in one case, one unique privacy policy was written for the municipality and the company that manages the web site, rather than the company appearing as a third-party which may have access to data). The opt-out option is referred only by 5 of the 7 municipalities that state their disclosure of information to third parties, but how the user can express his opposition is not indicated. The small number of municipalities that disclose personal information may be related to the fact that online service delivery in local government remains substantially underdeveloped in Portugal [19], or at least that interoperability is seldom used [17], since integrated delivery of e-government services is expected to imply the disclosure of information to other government bodies. It's interesting to note the difference to the Dutch case where only 11% of the municipalities state that personal data will not be disclosed, which is in complete opposition to the Portuguese case.

The procedures that must be followed for users to access and manipulate their data and the security measures that are taken to prevent unauthorized access to the user information, both important pieces of information in a security policy, are mentioned in only 61% and 60% of the manifests, respectively. None of the municipalities gives a concrete indication regarding specific security measures that are taken.

The use of cookies is referred in only 20% of the manifests, always to state that they are used in the web site. Nevertheless, we identified several cases of municipalities that do not include this topic in their manifest but whose web sites use cookies or have add-ons that explore them. References to the existence of specific privacy policies concerning child privacy where observed in only 10% of the manifests.

5 Conclusions and Future Work

In this article we presented an empirical study of privacy policy manifests of local government web sites in Portugal. The web sites of all the 308 Portuguese municipalities were surveyed for privacy policy manifests and a conceptual analysis was performed on the content of those manifests.

The general conclusion of the study is that users' privacy is not a priority for Portuguese municipalities, or at least privacy policies are not considered relevant for users' trust, as it is referred in the literature [4]. This conclusion is sustained in the observations that only 26% of the Portuguese municipalities have publicly available privacy policy manifests; and that even when those manifests are available, they miss important information (e.g. only 16% of the manifests, representing 4% of the Portuguese municipalities indicate which information is collected from the users by the web site, either implicitly or explicitly).

The absence of privacy manifests in the local government web sites is especially significant since trust in government is considered in the literature as an important factor for e-government adoption [10-12]. Thus, our results for privacy policy manifests corroborate previous conclusions that local e-government is substantially underdeveloped in Portugal [19]. The fact that a small number of municipalities refer an effective disclosure of personal information to third parties may also constitute evidence of this underdevelopment, since, besides depending on the collection of users' personal information, integrated delivery of e-government services is expected to imply the disclosure of information, at least to other government bodies [17].

In terms of international comparison, Portugal has a lower percentage of municipalities with privacy manifests than Nederland [28], where 77% have privacy manifests. The same is true when comparing to the 80 most populated European cities [22], among which 55% have privacy manifests. Although it is not possible to establish a quantitative comparison for the content and prominence of the manifests, Portugal share some of the problems identified in the case of Dutch municipalities [28]: privacy statements are not well aligned with the European Directive on the protection of individuals with regard to the processing of personal data and on free

movement of such data [1]; and municipalities do not pay much attention to making privacy statements observable in their web sites.

This study contributes to knowledge by establishing the current situation in terms of adoption and advertising of privacy policies in Portuguese local e-government. The knowledge of the current situation may be useful to researchers as basis for further studies, including investigating the reasons behind the current non satisfactory state of affairs, and to practitioners, by raising the awareness of the need for adequate privacy policies and their advertisement in the municipalities' web sites.

Our investigation addressed the privacy policy manifests that are advertised through the local government web sites. Although the conclusions are relevant, this approach may not entirely capture the complexity of the phenomenon. Future studies should use different methodologies to help complete this vision, for example by focusing on users and public officials perceptions relating the importance of privacy policies, or the constraints of their adoption, and by focusing on the effective practices of the municipality to determine if they indeed comply with the announced privacy policies.

References

1. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995. Official Journal of the European Communities, L 281, November 23, pp. 31–50 (1995)
2. Lei n.º 67/98, Lei da Proteção de Dados Pessoais. Diário da República, 247, Série A, October 26, pp. 5536–5546 (1998)
3. Peterson, D., Meinert, D., Criswell II, J., Crossland, M.: Consumer trust: privacy policies and third-party seals. Journal of Small Business and Enterprise Development 14(4), 654–669 (2007)
4. Arcand, M., Nantel, J., Arles-Dufour, M., Vincent, A.: The impact of reading a web site's privacy statement on perceived control over privacy and perceived trust. Online Information Review 31(5), 661–681 (2007)
5. West, D.M.: Global e-Government 2007. Center for Public Policy, Brown University (2007)
6. Moon, M.J.: The evolution of e-government among municipalities: Rhetoric or reality? Public Administration Review 62(4), 424–433 (2002)
7. Edmiston, K.D.: State and Local e-Government: Prospects and Challenges. The American Review of Public Administration 33, 20–45 (2003)
8. Dias, G.P., Rafael, J.A.: A Simple Model and a Distributed Architecture for Realizing One-Stop e-Government. Electronic Commerce Research and Applications 6(1), 81–90 (2007)
9. Otjacques, B., Hitzelberger, P., Feltz, F.: Interoperability of e-Government Information Systems: Issues of Identification and Data Sharing. Journal of Management Information Systems 23(4), 29–51 (2007)
10. Warkentin, M., Gefen, D., Pavlou, P.A., Rose, G.M.: Encouraging Citizen Adoption of e-Government by Building Trust. Electronic Markets 12, 157–162 (2002)
11. Carter, L., Bélanger, F.: The Utilization of e-Government Services: Citizen Trust, Innovation and Acceptance Factors. Information Systems Journal 15, 5–25 (2005)
12. Thompson, S.H.T., Srivastava, S.C., Jiang, L.: Trust and Electronic Government Success: An Empirical Study. Journal of Management Information Systems 25, 99–132 (2008)

13. Dias, G.P., Moreira, J.M.: Transparency, corruption and ICT (illustrated with Portuguese cases). In: Transparency, Information and Communication Technology: Social Responsibility and Accountability in Business and Education, pp. 151–162. Philosophy Documentation Center, Charlottesville (2008)
14. Simões, M.J., Santos, D.: Challenges in the Digital Cities and Regions in Portugal. In: Collaboration and Knowledge Economy: Issues, Applications, Case Studies. IOS Press, Amsterdam (2008)
15. Tavares, M.: The use of Cybergenres as a Communication and Knowledge Management Process in Local Government. In: Proc. of the 26th ACM International Conference on Design of Communication, pp. 283–284. ACM, New York (2008)
16. Gomes, R., Ribeiro, J.: Use of Web Services in E-Government Information Systems - A Case Study. In: Proc. of the 2009 International Conference on Information Management and Engineering, pp. 475–480. IEEE Comuter Soc., Los Alamitos (2009)
17. Dias, G.P., Narciso, T.: Analysis of the potential for organizational interoperability improvement in local government. In: Proc. of the 5th Iberian Conference on Information Systems and Technologies, pp. 167–172. AISTI, Rio Tinto (2010)
18. Santinha, G., Castro, E.A.: Creating More Intelligent Cities: The Role of ICT in Promoting Territorial Governance. Journal of Urban Technology 17(2), 77–98 (2010)
19. Dias, G.P.: Local e-Government Information and Service Delivery: A survey of municipal websites in Portugal. In: Proc. of the 6th Iberian Information Systems and Technologies Conference, vol. 1, pp. 98–103. AISTI, Rio Tinto (2011)
20. Aleixo, C., Nunes, M., Isaias, P.: Usability and Digital Inclusion: Standards and Guidelines. International Journal of Public Administration 35(3), 221–239 (2012)
21. Almeida, F.B., Tomé, P.R.P.: e-Government Local Analysis and Evaluation of Municipalities in the district of Viseu. In: Proc. of the 7th Iberian Conference on Information Systems and Technologies, pp. 253–258. AISTI, Rio Tinto (2012)
22. Miranda, F.J., Sanguino, R., Bañegil, T.M.: Quantitative assessment of European municipal web sites: Development and use of an evaluation tool. Internet Research 19(4), 425–441 (2009)
23. Tolley, A., Mundy, D.: Towards workable privacy for UK eGovernment on the web. International Journal of Electronic Governance 2(1), 74–88 (2009)
24. Alhomod, S.M., Shafi, M.M.: Privacy Policy in E Government Websites: A Case Study of Saudi Arabia. Computer and Information Science 5(2), 88–93 (2012)
25. McRobb, S., Rogerson, S.: Are they really listening?: An investigation into published online privacy policies at the beginning of the third millennium. Information Technology & People 17(4), 442–461 (2004)
26. Pollach, I.: Privacy Statements as a Means of Uncertainty Reduction in WWW Interactions. In: Clarke, S. (ed.) End-User Computing: Concepts, Methodologies, Tools, and Applications, pp. 188–208 (2008)
27. Robles-Estrada, C., Vargas-Barraza, J.A., del C. Sepúlveda-Núñez, M.D.: Are Privacy Issues Important in Mexican Online Markets? An Empirical Investigation into Published Online Privacy Statements of Mexican Web Sites. In: BLED 2006 Proceedings (2006)
28. Beldad, A.D., de Jon, M., Steehouder, M.F.: When the bureaucrat promises to safeguard your online privacy: Dissecting the contents of privacy statements on Dutch municipal websites. Government Information Quarterly 26(4), 559–566 (2009)
29. Heeks, R., Bailur, S.: Analyzing e-government research: Perspectives, philosophies, theories, methods, and practice. Government Information Quarterly 24, 243–265 (2007)

Knowledge Integration in Problem Solving Processes

Maria José Sousa

CIEO – Research Centre for Spatial and Organizational Dynamics
UAB – Open University of Lisbon
Isla Campus Lisboa
Portugal
mjsousa.pt@gmail.com

Abstract. This paper analyses the knowledge integration in problem-solving situations which requires a high level of interaction and trust among workers. The literature review explores the main barriers associated to knowledge integration and use; even if most problem situations are solved in an unconscious way, automatically and in a few seconds, others situations requires more time, effort, teamwork, collaboration and extensive abstract knowledge. This research goal is to analyse the perceptions of the workers from Alpha Organisation. The research findings allow us to conclude that, depending on the complexity of the workstation, the Operator decides if he has the knowledge and the tools to solve the problem or if he needs help from Managers. The use and share of employees' knowledge is an important factor to solve problems and strengthen performance. However, several organisational and individual barriers condition the process.

Keywords: knowledge integration, problem solving, organisations.

1 Introduction

The main goal of this research was to analyse the perceptions of organisational actors regarding knowledge integration through problem solving processes in the activities and practices of organisations. The research was conducted in one Portuguese industrial company and involved operators, technicians and managers in separate group recall sessions where they shared experiences, ideas and gave suggestions about the knowledge integration processes.

Five employees participated in each group recall session - totalling 30 persons from the company. For data analysis we used analysis grids based on employees' quotations. We did not want to find percentages or values, and it was not our goal to make any kind of measurements, but we sought to understand the processes for knowledge sharing and integration in the company in problem solving situations.

After the introduction, the literature on knowledge sharing processes is reviewed, followed by the research methodology and the findings from the research.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 97–109.
DOI: 10.1007/978-3-642-36981-0_10

2 Knowledge Integration Process

2.1 Knowledge Concept

The aim of this research is to analyse the knowledge integration processes in problem solving situations. To analyse this association it is important to support our research in the knowledge management literature, and it show us that knowledge can be an enabler or a disabler of problem situations because individual knowledge transfer and use is a very complex, social interaction process [1], [2].

According to this idea, it is important to analyse the concept of knowledge, as the nuclear element of this study. Thus, [3] states that "knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information". From another perspective, [4] associates knowledge to action, saying that "knowledge is the ability to act". [5] explain that knowledge is created by the flow of information associated with the beliefs and commitment of those who possess it. In their perspective, knowledge is created within the company to make it more successful, to keep it on the market, to increase competitiveness and to keep it ahead of its rivals.

To make a broader analysis of the concept, [6] remarks that today's organisations do not compete in terms of products, services or technology but in terms of know-how, processes and values.

This concept analysis leads us to conclude that the immaterial nature of knowledge makes the sharing and integration process difficult for new products, practices or even in problem situation solving processes.

2.2 Knowledge Sharing and Integration

To answer the main question of the research, it became important to understand how employees use their knowledge to develop their work routines and embed their knowledge in a problem situation or into new products and services leading to a sustained, competitive advantage. The literature is profuse in examples of barriers associated to knowledge sharing processes: a) individual barriers, grounded in the participants of the knowledge sharing process (e.g. the fear to lose a personal competitive advantage and to be misunderstood and misinterpreted, group thinking, preference to one's own ideas instead of somebody else's, etc. [7]; b) infrastructural barriers, determined by organisational structure, system of communications and organisational culture [8], [9]; c) ontological barriers, dealing with the knowledge itself and arising from the tacit knowledge transfer problems [10], as well as from perceived value of knowledge [11] that is often not recognized at all by the knowledge sharing participants [9]. Some of these barriers we also found later on in the field research, conditioning knowledge sharing.

These barriers also influence the integration and the effective use of knowledge. In addition, an important dimension for knowledge integration is the need for a coherent, company-wide social identity, instead of a multiple community or group-based social

identity, in order to promote effective knowledge integration in organisations [12]. Trust is also a nuclear element and [13] pointed out its importance: they suggest that competence and trust among individuals in an organization influences the link between them and the effective use of knowledge. [14] also refer that the lack of trust among employees is another key barriers to knowledge-sharing activities. To overcome that kind of obstacle, organisational studies point out the importance of a democratic and participative leadership as the main factor to enable a culture of knowledge sharing [15].

To overcome the barriers to knowledge integration and use, organisations create problem solving routines. According to [16], problem solving involves three phases: preparation (understanding the problem); production (developing different alternative solutions) and judgment (selecting a solution). [17] suggest a fourth phase: review and reflective assessment of both outcomes and processes.

Even with those kinds of routines implemented to help the organisation respond to problem situations or challenges, knowledge integration is also conditioned by its complexity and because there are factors depending on the source/receiver of knowledge that affects the effective use and integration of knowledge. In a culture of knowledge sharing, the transfer and integration of knowledge may also be influenced by the use of a common and shared vocabulary. [18] reinforces that idea, considering the influence of structural diversity on work group performance, meaning that when members of diverse work groups are capable of sharing external knowledge to the group, their performance improves.

3 Methodology

3.1 Research Methodology

The methodological approach was qualitative, and it was applied to the method of cases or intensive analysis. To accomplish the empirical work, we used the Action Research methodology; the main technique to collect data was group recall sessions with all levels of employees of the organization.

Action Research (AR) was the applied methodology for this study because we did not want to discover generalizations, but contextual findings and rich insights; AR allows a deeper analysis and a different understanding of complex organizational problems [15].

Data was collected through interviews to top management employing group recall techniques, which were also used for knowledge sharing among the researcher and the organizational actors. The group recall technique can be framed in social research, and the process is similar to a focus group process. It gives the researcher the opportunity to hear detailed revelations about people's thoughts, ideas and experiences. It has potential to illuminate workers' contrasting opinions and experiences and to help them get to know the organization better, while sharing their experiences and work practices with colleagues.

The combined contributions of participants in the group recall might point to new directions and questions changing the trajectory of the session. There is then potential for the researcher to explore their remembrances allowing space for participants to articulate their own ideas.

The researcher's key responsibilities associated to group recall technique are listed in Table1.

Table 1. Researcher's Group Recall Responsibilities

Role	*Responsibility*
Researcher - organises the group recall sessions, focuses on communications between and among group recall participants. This is accomplished in informal, face-to-face sessions.	- Clarify communications - Draw out the reticent - Ensure that dissenting points of view are heard and understood - Keep discussions on the topic - Reconcile opposing points of view - Ensure and articulate a valid purpose for the group recall realization - Stimulate interest in the group recall - Organize face-to-face group recall sessions - Stimulate organisational actors' participation - Arrange for communications support - Obtain official support when appropriate - Communicate the contributions of the group recall sessions to the formal organisation

The practical issues of group recall sessions are mainly the logistics of getting all participants together at the same time in the same room, prepared to talk on the same subject, and the concerns that some contributions might become lost in the wider debate.

One main constraint about the use of this technique is the danger of censoring: individuals hold back the contributions they wish to make, conforming instead to an apparent consensus or to the opinions of a self-appointed "expert" within the group. Another concern is the possibility for participants to exaggerate their accounts in order to "impress" others. However, and because several participants could belong to the same work unit, this was unlikely to happen since the other members of the group would function as control elements.

3.2 Data Analysis

To analyses data we will first consider how the theoretical concepts discussed in the literature (i.e., knowledge sharing and knowledge integration) can be analysed according to the workers' perceptions related to problem solving situations.

The procedure will include the analysis of the data collected from the several group recall sessions and mapping the contexts, practices and processes used by workers when a problem needs to be studied and solved.

Quotations from all levels of workers will be considered and analysed in the light of theoretical concepts and also the organization's practices.

4 Research Findings

4.1 Problem Solving in Alpha Organisation

Alpha Organisation uses collaborative problem-solving techniques. The problem-solving situations encountered were consistent with [19] distinction between routine and non-routine problems.

a) *Routine problems* involve situations that have been experienced before:

- *Simple, repetitive and well-understood situations*, which are handled in a tacit mode, with very little conscious thought. The simple nature of these situations allows for easy explication. This happens with Alpha Organisation quality problems.
- *Routine situations within a different context*, when workers face problems that are similar but are not exact repetitions of previous experiences. As routine problems become more complex, the capacity to address these situations depends on the ability to quickly recognise and diagnose the problem. In Alpha Organisation, when workers face this kind of problem and it seems to be impossible for them to solve them on their own, the problem is reported to the Manager.

b) *Non-routine problems* need workers' knowledge to solve novel problems that may represent their most valuable contribution.

- *Solving novel problems* need workers' ability to define the problem and to work collaboratively with others employees from different sections to find a solution. In Alpha Organisation, when a novel problem occurs, a team from different sections tries to find the solution.
- *Emergent problems* can be described as workers proactively identifying problems to explore or process improvement or new work situations. In Alpha Organisation, workers use the suggestion boxes to solve emergent problems.
- *Solving problems outside of expertise:* these are problems that are unique and outside of their existing domains of experience and expertise. In Alpha Organisation, we can find these problems related to some equipment that may need the supplier's expertise.

All these problems are discussed in communication corners along with all other aspects of organisational life. This not only promotes the sharing of ideas and opinions, but it also illustrates the importance of developing effective communication and consultation processes to overcome cultural and attitudinal obstacles to change. The need for all organisational actors to be informed about the need and purpose of the organisational innovation process is emphasized in this case.

4.2 Knowledge Integration in Problem Solving Processes

Problem solving is aided by the quality and availability of the knowledge used to handle situations. According to [20], it helps to "decide what to do, innovate, act and evaluate the implications of approaches and action".

According to the ALPHA Organisation context, we will try to understand the perceptions of the workers according to the procedures used in problem solving situations.

In respect to the problem solving methodology, we acknowledge that it is framed by the Production System, and it has established rules that define the autonomy and the complexity of the problems that can be solved by each level of workers:

> "ALPHA Organisation has a good system of problems resolution. It is part of the Production System and the new methodology of work." (Group recall – Middle Managers)

> "Workers have autonomy to solve less complex problems, and problems and solutions are registered in a database that can be consulted when a problem occurs, facilitating the use of knowledge." (Group recall – Department Managers)

Workers have an important role in problem-solving situations. Their individual knowledge is the critical factor to identify the problem and the possible solutions. Production Managers who work directly with the workers have identified workers with two different attitudes:

> "a) Workers that don't show any concern about the problems. b) Workers that try to help on an individual basis; when they can't solve the problem, they communicate it to the shift Manager." (Group recall – Production Managers)

ALPHA Organisation uses temporary workers when necessary; when the contract of some of these workers is near its end, they assume a contentious attitude and do not show any concern the achievement of production goals or for the product quality. However, it seems that the majority of workers have a strong link to ALPHA Organisation showing involvement and participation in Production System implementation.

Nevertheless, it is easier for workers to try to solve technical problems than organisational ones:

> "Some problems are mere anomalies that employees can identify, and they have an easy solution, especially when we are dealing with technical problems and not with organisational ones (for these they don't have the necessary knowledge). This has been an everyday battle, with systematic procedures thought to make all the employees involved." (Group recall – Production Managers)

The majority of the knowledge in organisations is dynamic because it is concentrated on workers, but some of that knowledge is static (documental information, for example). It is essential that the dynamic knowledge can be stored in repositories which, over a period of time, will become a substantial source of relevant information and expertise.

> "Each workplace has one level of autonomy associated, in respect to decision making and problem resolution." (Group recall – Production Managers)

> "If I have a simple problem in my machine, normally I know how to fix it. It is only when the problem seems to be very complex that I consult my shift Manager" (Group recall – Operators)

Knowledge can be a criterion for autonomy and decision-making. The more knowledgeable workers are, the more potential they have and the more autonomous they can be, unlike other workers that are less knowledgeable. When the worker's range of knowledge is broader, his contribution is greater, and he is in a position to make some kind of technical decision.

> "It is possible to seek a description of the problem's resolution, and access a set of quality tools: analysis, diagnosis, information and research." (Group recall – Middle Managers)

During the workday, workers face several problems, and they solve most of them unconsciously (in a tacit way), automatically and in a few seconds. Other situations require more time, effort, teamwork and collaboration. Situations can vary widely: some are well known and require routine – even automated – knowledge, while others are more complex and require extensive, abstract knowledge.

> "In ALPHA Organisation, when there is a problem, we have some technical procedures that we have to follow. If it is a simple problem that the Operator knows how to solve, he can do it alone. If he cannot discover a solution, he then informs the Team Manager and together they try to find a solution. If it is a very complex problem, a team with several Operators and Technicians is created to analyse the problem. The Operator that finds the problem also participates in this team that meets one time per week to decide the situations that appear and to define the corrective solutions." (Group recall – Department Managers)

> "We have problems that can be easily solved and others that are more difficult, but we have procedures defined for each of them." (Group recall – Middle Managers)

It is important to point out the alignment of perception in every hierarchical position according to problem resolution procedures. During Technicians' group recall, they described an identical procedure or routine when a problem occurs to Department Managers and Middle Managers. They said that when a problem emerges:

> "Depending on the complexity of the workstation, the Operator decides if he has the knowledge and the tools to solve the problem by him or if he needs help from the shift Manager. If the problem is too complex, he does not have the autonomy to decide the solution to the problem and then he informs the Shift Manager that evaluates the type of problem, like if it is a quality problem or if it assumes some other form." (Group recall – Technicians)

Operators also have a similar perception of problem resolution:

> "If it is a problem in a machine, the evaluation is made by me. I have autonomy to make the first evaluation. If the problem is very complex, we have an internal system that initiates with an intervention order send to maintenance and it is also communicated to the shift Manager."

> "If it is a quality problem, all the production stops and we quickly analyse the problem, trying to identify the phase where it has initiated. Sometimes the problem started in the previous shift."

> "In the Welding section the procedures are the same: we analyse the problem and if we can, we solve it. The remaining problems are registered in proper documentation." (Group recall – Operators)

Using [21] distinction of problem types as either routine or non-routine, [19] identified routine problems as the ones "requiring individuals to expend little conscious or effortful thinking" (p. 22). Routine problems are addressed through a process called assimilation, that is, the ability to act gained through repeated practice, without conscious thought. Solving routine problems reinforces and refines existing knowledge.

> "For instance, if it is a quality problem, we have some procedures that we need to follow according to the Quality Manual, and the problems need to be registered as well as their own specific solution. The people involved in the problem and in the solution are also identified, so that if another problem like that occurs in another area of the plant, all employees have access to the problems and solutions database." (Group recall – Department Managers)

Non-routine, or novel problems, require "extensive conscious thinking" [19] and extended knowledge through accommodation [21]. Learning occurs when one encounters a new task or challenge. Solving novel problems enables workers to identify and close gaps in knowledge and learn new models, clues and cues on how to proceed [19].

> "Solving new problems gathers the involved people in the discussion of the solution. They discuss the problem, identify it and implement several actions according to the problem resolution." (Group recall – Middle Managers)

ALPHA Organisation routine in creating and using knowledge in the problem-solving process began with the problem-finding phase; the problem is then analysed by the Operator and/or the shift Manager. If they cannot solve the problem, they consult the quality database where they store all problems and solutions. If the problem is too complex, they create a team to solve it, and, when they find the solution, they implement it and register the problem and its solution in the database.

However, ALPHA Organisation is always looking for new ways to improve their practices and routines. Middle Managers focus on a particular issue and are determined to implement a more efficient methodology of problem solving.

> "There is going to be implemented a more rigorous, standard and detailed methodology, not only in the production lines, but also in the other sections of the organisation. Production System is going to organize what already is a good practice, making it even more efficient." (Group recall – Middle Managers)

> "Problem resolution is the priority; correcting problems is something that we think about constantly and whenever the machines are working. Our priority is keeping a continuous production process." (Group recall – Middle Managers)

> "We are already well equipped to facilitate problem solving situations, and the plant organisation is an important factor, but we are creating standards that will help to solve problems more quickly." (Group recall – Middle Managers)

According to [16], problem solving involves three phases: preparation (understanding the problem), production (developing different, alternative solutions) and judgment (selecting a solution). [17] suggest a fourth phase: review and reflective assessment of both outcomes and processes.

Even if it is important to have tools, procedures and routines to help the organisation respond to problem situations or challenges, these kinds of factors can sometimes be a barrier to new knowledge development and even to knowledge use. One of the Middle Managers that participated in the group recall session was very concerned with standardization and "routinization":

> "The question of standardization is 'sexy'. It is 'fashionable', but we do not always have a reason for it. In many cases, the profit does not compensate, and the goal of standardization is lost.
>
> Our biggest fear is to create documents or systems that become impossible to use either because the information is already clear and accessible or because information is difficult to access. (Group recall – Middle Managers)"

> "For example, a quality database is a powerful tool, but we don't have the competencies to use it. It is necessary to create an external tool that allows database access.
>
> Creating a database was an imposition in order to improve the quality system, since it was necessary to register any problem. Now that we have a good database, we do not have enough knowledge to make queries and statistics. Not everything is pink, we have things to improve." (Middle Manager)

However, even if ALPHA Organisation had several routines to create and share knowledge, they would also need the space for informal sharing and its development. This is reflected in the Middle Managers' words:

> "We don't have resolution procedures defined for all kinds of problems. What we have is formal and informal actions. Normally, organisational problems are decided in an informal way and we have a trend to extend these actions to technical problems. The impulse is not to appeal to the defined procedure, but to decide the problem ongoing. However, with the organisation maturity, the procedures are more adapted to the reality." (Group recall – Middle Managers)

One of main concerns of Alpha Organisation's management is to store and codify rules and procedures in a simple format so that employees can easily access and understand them. If rules and procedures are not stored and written clearly, each employee is likely to follow his/her own interpretation of the rules. When rules and procedures are clearly marked down, there is far less ambiguity in understanding and interpreting those rules and procedures. The Production System process of automation and standardization of tasks and schedules is a mean of handling this situation.

On the other hand, a problem-solving approach creates a high level of interaction, and the closeness and the trust among workers is the key to the degree of tacit knowledge shared. Most problem situations are solved in an unconscious way, automatically and in a few seconds. Other situations require more time, effort, teamwork and collaboration, as well as extensive, abstract knowledge.

In Alpha Organisation, depending on the complexity of the workstation, the Operator decides if he has the knowledge and the tools to solve the problem or if he needs help from Managers. If the problem is too complex and he does not have autonomy to decide about the problem solution, he then informs the Shift Manager who evaluates the nature of the problem.

Alpha Organisation's routine in creating and using knowledge in problem-solving processes (Figure 1) began with the problem-finding phase; then the problem is analysed by the Operator and/or the shift Manager. If they cannot solve the problem, they consult the quality database where they store all problems and solutions. If the problem is too complex, they create a team to solve it; when they find the solution, they implement it and register the problem and the corresponding solution in the database.

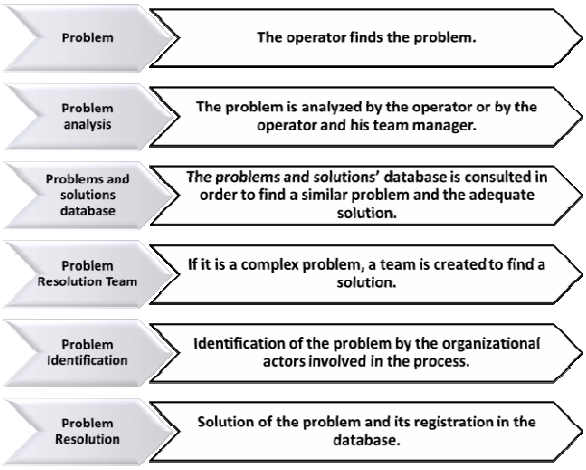


Fig. 1. Problem-solving process: routine in knowledge creation and use

Alpha Organisation has several routines to create and share knowledge, and the Production System imposes a constant creation of new knowledge, especially regarding the organisational innovation process.

The innovation process is a key factor because of the importance of implementing new ways of production and new organisational processes to accomplish higher efficiency. Involving workers in this process requires the use of management tools such as communication and the promotion of workers' involvement and participation. Alpha Organisation uses several mechanisms to promote knowledge sharing and develop new ideas. It is important to point out the suggestions system (mainly used to make production improvements), the workshops on innovations and new products, and the knowledge networks (especially the informal ones).

Looking at another perspective, we can say that Alpha Organisation is a learning space at a technical and organisational level. One of the most effective tools to create and disseminate knowledge is though workshops with people from different sections or people from only one section.

Consumers and external specialists often participate in the workshops and help the discussion and the creation of new knowledge that helps implement new practices, tools or technology.

The workshops in Alpha Organisation can be seen as knowledge creation and sharing processes, like the communities of practice or other processes of linking workers to others with expertise. Relational competences are a key to the capture, use and creation of new knowledge and learning within Alpha Organisation.

The participation of all organisational actors in the Production System process helps to develop a more consistent knowledge-sharing culture. Employees share ideas and insights naturally and not as something they are forced to do. There is a connection between sharing knowledge and achieving the business goals or solving practical problems.

In Alpha Organisation, the knowledge sharing process among sections and workers is very peculiar. They implement a new practice, process or technology in one specific workstation according to the Operator's openness to change. When it is working perfectly, and new and better results are achieved, they share this new knowledge to other workers and transfer it to their workstations, disseminating the new knowledge throughout the plant.

5 Conclusions

This research contributes to the fundamental understanding of problem solving in organisations. The integration and use of workers' knowledge in those situations are the nuclear element of the research developed in Alpha Organisation.

This organisation has an implemented routine of problem solving that includes several phases: it begins with the problem-finding phase, then the problem is analysed by the Operator and/or the Manager. If they cannot solve the problem, they consult the quality database where they store all problems and solutions. If the problem is too

complex, they create a team to solve it; when they find the solution, they implement it and register the problem and the corresponding solution in the database.

All phases can be accomplished by the disperse knowledge that belongs to the workers of the organisation. Each worker has autonomy to solve problems associated to their level; depending on the complexity of the problem, they decide if they have the knowledge required or if they have to go to another level of analysis.

The research methodology adopted was crucial for the success of the research and the amplitude of the knowledge shared along the process. The sharing was not only done among organisational actors, but also between them and the researcher. In the group recall sessions, organisational actors become more aware of practices and processes used in other sections/departments of the organisation, and they also shared individual and group experiences. This means that the group recall sessions functioned as a space for knowledge sharing, as well as learning.

For future studies, another kind of research could be undertaken:

- Studies on knowledge integration across organisational functions and in other types of organisations.
- Studies that develop and test a theoretical framework that relates knowledge integration mechanisms, situational characteristics and organisational outcomes.

Furthermore, future studies are required to analyse the capabilities of employees' informal networks in order to achieve efficient integration of knowledge into work practices and problem-solving situations.

Acknowledgments. This article was partially funded by the Foundation for Science and Technology.

References

1. McAdam, R., McCreedy, S.: The process of knowledge management within organisations: a critical assessment of both theory and practice. Knowledge and Process Management 6(2), 101–112 (1999)
2. Nonaka, I., et al.: SECI, Ba and leadership: a unified model of dynamic knowledge creation. Long Range Planning 33, 5–34 (2000)
3. Davenport, H.T., Prusak, L.: Working knowledge: how Organisations manage what they know. Harvard Business School Press, Boston (2000)
4. Polanyi, M.: Personal Knowledge. Towards a Post Critical Philosophy, 428 pages. Routledge, London (1958)
5. Nonaka, I., Takeuchi, H.: The knowledge-creating company: how Japanese companies create the dynamics of innovation. Oxford University Press, New York (1995)
6. Coulson-Thomas, C.: Transforming the company, 2nd edn. Kogan Page, London (2002)
7. Husted, K., Michailova, S.: Diagnosing and fighting knowledge sharing hostility. Organisational Dynamics 31(1), 60–73 (2002)
8. Bock, et al.: Behavioral intention formation knowledge sharing: examining roles of extrinsic motivators, social-psychological forces, and organisational climate. MIS Quarterly 29(1), 87–111 (2005)

9. Hall, W.P.: A technology architecture for managing explicit knowledge over the entire lifecycle of large and complex projects. Ark Group Intranet Content Management, 20–21 (2002)
10. Nonaka, I.: The knowledge-creating company. Harvard Business Review 69(6), 96–111 (1991)
11. Ford, D.P., Staples, D.S.: Perceived value of knowledge: shall I give uou my gem, my coal? In: de Edição, L. (ed.) 38th Annual Hawaii International Conference for Systems Sciences (HICSS-38), Waikoloa, Hawaii. Awarded Best Paper Award for Organisational Systems and Technology: Knowledge Management Track (January 2005)
12. Willem, et al.: Impact of coherent versus multiple identities on knowledge integration. Journal of Information Science 34, 370–386 (2008)
13. Levin, Cross: Mediating Role of Trust in Effective Knowledge Transfer. Management Science 50(11), 1477–1490 (2004)
14. Lee, H., Choi, B.: Knowledge management enablers, processes, and organisational performance: an integrative view and empirical examination. Journal of Management Information Systems 20(1), 179–228 (2003)
15. Sousa, M.: Dynamic knowledge: An Action Research Project. The International Journal of Knowledge, Culture and Change Management 10(1) (2010)
16. Johnson, D.M.: The Psychology of Thought and Judgment. Harper, New York (1955)
17. Argyris, C., Schön, D.: Organizational learning II: Theory, method and practice. Addison Wesley, Reading (1996)
18. Cummings, J.N.: Work groups, structural diversity, and knowledge sharing in a global organization. Management Science 50(3), 352–364 (2004)
19. Billett, S.: Learning throughout working life: Interdependencies at work. Studies in Continuing Education 23(1), 19–35 (2001)
20. Wiig, K.M.: Knowledge management has many facets (2003), http://www.krii.com/downloads/fourkmfacets.pdf
21. Piaget, J.: Autobiographie, Cahiers Vilfredo Pareto, vol. 10, pp. 129–159. Droz, Génève (1996)

Collaborative Elicitation of Conceptual Representations: A Corpus-Based Approach

Cristóvão Sousa[1,2], Carla Pereira[1,2], and António Soares[1,3]

[1] CIICESI - ESTGF, Polytechnic Institute of Porto,
Rua do Curral - Casa do Curral-Margaride, 4610-156, Felgueiras-Portugal
[2] INESC Porto, Rua Dr. Roberto Frias, s/n 4200, Porto-Portugal
[3] Department of Informatics Engineering, Faculty of Engineering,
University of Porto, Rua Dr. Roberto Frias, s/n 4200, Porto-Portugal
{cds,csp}@estgf.ipp.pt, als@fe.up.pt

Abstract. Knowledge is an important resource for organisations, and being able to manage it is a key factor for success. New information management and knowledge sharing approaches should be able to cope not only with possible variations in business situations and contexts, but also with various and sometimes discordant viewpoints, which are inherent to collaborative environments. Developing, reusing and maintaining common interpretations of available information is crucial to support real-word organisational activities. The particularly challenging problem of knowledge elicitation is tackled here combining terminological and knowledge representation views. A corpus-based conceptual modelling architecture was designed and discussed together with the workflows for real-time context retrieval and lexical pattern discovery. The practical implementation and validation of this work are accomplished on the ConceptME system, a platform developed as part of this research line, providing knowledge and terminological tools and resources to support activities that involve collaborative conceptualisation processes according to the ColBlend method.

Keywords: Knowledge sharing**,** Collaborative modeling, Knowledge Representation, Terminology.

1 Introduction

Organisations in general have been intensifying collaborative practices to improve their performance and to find new opportunities. The concept of collaborative networked organisations (CNO) emerged from this paradigm shift to cope with the challenges posed by the everyday business activities, which are becoming more and more knowledge and collaboration intensive [5], [6]. These new complex and demanding environments, which involve different social contexts, require not only an effective organisation and information retrieval, but also social interaction based on information management, as a way to effectively share and use knowledge. These demands call for a new approaches to information and knowledge management (I&KM), addressing the social construction of structures of organising

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 111–124.
DOI: 10.1007/978-3-642-36981-0_11

knowledge[23]. Semantic theories should underlie the development of artifacts to share information [10]. Following a design science research [9], we have been studying the conceptualisation process as the process that ends with a conceptual representation (CR). The designed approach addresses CR as designed pragmatic artifacts capable of collecting and processing domain concepts, whose notions are context-dependent. This assumption led to the design of the ColBlend method [17], a collaborative conceptualisation based solution at group level, in an inter-organisational context. This solution encompasses two components: the formalisation of a socio-semantic network - whose nodes are social actors and concept maps (or informal ontologies) - and the instantiation of this model in a conceptual negotiation process inspired by the conceptual blending theory. Considering specific CNO needs and empirical findings from a case study [18], the ColBlend method led to the development of the ConceptME, a "conceptual Modelling Environment" where groups of specialists can find tools and resources to collaboratively develop conceptual representations (such as concept maps), organise them in libraries, share them with other colleagues and reuse them when necessary. This tool was the instrument used to confirm the relevance and the validity of the approach. In this paper the emphasis is on supporting the construction of the blend space (as the ColBlend terminology) using techniques to retrieve immediate contexts of terms, and assisting users on their conceptualisation activities - more specifically concept elicitation and concept discussion. Without disregarding the socio-semantic approach on which this research is based, this particular approach encloses a double theoretical perspective combining terminology and knowledge representation to support our conceptual modelling approach. Furthermore, an experiment was conducted to evaluate the usefulness of the approach to a group of users building a conceptual representation of a domain. The paper is concluded with a brief description of the implementation of the designed process and architecture in the ConceptME.

2 Related Work

The work described in this paper is based on collaborative conceptualization tasks, but focusing on a particular approach about the symbiosis between corpus-analysis and the tasks to represent a domain knowledge model, which is the main concern of the Conceptualisation Process (CP)[1]. This work follows the ColBlend method [17] as the very process of supporting the collaborative conceptualisation process (CCP)[2].

[1] In relation to an individual, a conceptualisation process of a given piece of reality is a collection of ordered cognitive activities that has information and knowledge as inputs internally or externally accessible to the individual, and an internal or external conceptual representation as output [18].

[2] A "collaborative conceptualisation process" (CCP) is a conceptualisation process that involves more than one individual producing an agreed conceptual representation. In addition to an individual CP, the CCP involves social activities that include the meaning negotiation and practical management activities for the collaborative process. In this paper "knowledge representation process" is also used to refer, in practical terms, to a CP.

ColBlend was outlined based on the conceptual blending theory whose goal was to achieve an agreed set of conceptual models using a process involving explanation, discussion and negotiation. The process runs through a set of virtual spaces: a) the input spaces - private to each party involved in the CP where the knowledge models proposals are built; b) the blend space - which contains the proposal resulting from the analysis of the input spaces and presented for discussion. Moreover, the process proposes new concepts (originally not identified) from a global analysis of the current content of the spaces and; c) the generic space - which contains the common domain knowledge model composed of all parts of the proposals that were accepted by all and "published" on this shared space. Another aspect related to the work described here is related to the interplay between terminology and knowledge representation. The literature revealed ontology engineering as the main research topics in that field. There is a significant amount of research addressing the construction of ontologies based on terminology [8], [24], [4]. In [8] and [7], terminology plays the role of term system provider which acts as input to create the ontology. The authors propose an automatic process to identify a tree of lexical related terms, which constitute a candidate conceptual structure. In [24], the author presents a three step process based on extraction techniques and textual corpus analysis, comprising the following points: i) recognising terminology in text (using statistical analysis and association rules created using TextAnalyst software, plus semantic network analysis in order to overcome the problem of ignoring the terms which occur less frequently); b) naming tags in terminology (in order to face the synonyms or variance issue); c) deriving hierarchies (using Formal Conceptual Analysis). The author's stance is that "linguistic perspectives should be considered while building ontologies". Furthermore, he underlines the need to develop a "lightweight ontology"[3]. Consequently, ontology editors must first construct a conceptual system, after which editors should identify hierarchical structures among concepts". In [4] the authors argue - once again - that ontologies should be based on linguistics. The goal was to enrich current formalisms such as RDFS/OWL to include linguistic information such as "part-of-speech metadata of the lexical items", morphological information and variations, expressed as RDFS/OWL properties. One of the main areas where terminology interacts with ontology engineering is that of ontology learning. As mentioned by [3], "Term extraction is a prerequisite for all aspects of ontology learning from text". However, we consider that the use of terminology within knowledge representation contexts is wider than the use given by the ontology-learning field, that is, mainly corpus tagging for information extraction.

The common denominator of the work found is learning ontological relations focusing on seeking lexical patterns within fully automatic processes. At no time, collaboration issues were addressed. The aim here is to provide a general toolkit for specialists of any field of expertise, allowing them to conceptualise in a network environment. The authors in [1] address specific issues on corpus analysis regarding the global process conceptual modelling, rather than gathering lexical patterns to fulfill ontological relations as [15], for example. The work[1] presents a set of

[3] A schema like taxonomy which comprises a conceptual system used to model knowledge.

requirements to select and organise texts and a method on how to deal with the wide range of terminological information provided by the corpus, during conceptual modelling.

3 Collaboration in Knowledge Representation

According to the CPP definition, the following elements were identified as the main components of an abstract representation of a typical CPP scenario: a) **the inputs** - information and knowledge internally or externally accessible; b) **social interactions** - they may contribute to move a resource from an internal source to an external source or, on the other hand, to bring up new resources crucial for negotiation activities [16]; c) **negotiation activities** – these could be supported by different and external perspectives (for example, argumentation-based strategies or decision support methods). However, the CCP takes into account a negotiation baseline comprising discussions activities around concepts, by exchanging contextual-information (that is, information considered relevant to understand the concepts) enriching the developed conceptual structures. Additionally, it comprises a process of calculating the degree of similarity between conceptual structures, conducted by the users, thus making it possible to merge two or more conceptual structures; d) **the outcome (conceptual structures)** - models (or artifacts) representing a perception of reality using diagrammatic views which form a network of concepts interconnected by meaningful linking phrases. Considering the previous elements and inspired by the empirical studies conducted as part of ColBlend [18], a small set containing the major building blocks or top-level requirements are described. The following building blocks (B1 to B3 in the list below) are considered fundamental for a collaborative elicitation of conceptual representations based on a corpus-based approach to be successful.

B1 - Terminological methods for corpus organisation: Each organisation represents its conceptualisation proposal fed by a set of knowledge sources (such as URIs specifying documents, webpages) shared by all parties. The use of textual resources should be preceded by a task where the aim is to organise the domain corpus. Having such a repository could enable a more efficient extraction of term candidates for the construction of the initial conceptualisation proposals [1], [21], [14].

B2 - Basic top-level conceptual structures: It is taken as an assumption that the main conceptualisation result is a less formal knowledge representation, which could be shaped into a shared conceptual model. One possible way of ensuring a common interpretation of the created conceptual models includes sharing a set of basic top-level conceptual structures and the meaning of their concepts and relations. Following a top-down approach on representing a domain knowledge, the process of creating conceptual structures could, among others, be based on patterns, which could be translated from text and from an ontological and synthesised perspective regarding its reuse potential. Gradually, these structures could be fine-tuned using term contexts and further negotiation activities.

B3 - Methods for analysing consistency in the conceptual structures: Ensuring semantic consistency for conceptual structure interchange requires something more than just gathering conceptual structure patterns and defining a set of basic templates to be (re)used. If it is acceptable that templates could help create generic domain conceptual structures, inversely the specifics of a domain field enclose a set of particular details that typical top-level templates may not consider. Some assistance should be given to the users in order to help them determine the meaning of each structure. This could be performed on one hand by attaching specific metadata to conceptual structures and, one the other hand, by implementing real-time mechanisms to validate the ontological compliance of the model under construction. Examples include suggestion mechanisms, either based on context analysis or on cross-checking the various models of the same domain.

The latest research efforts of this team have been towards shaping the building blocks mentioned above in order to support conceptualisation. For that, several research lines have been traced, one of which is based on terminology exploring how corpus analysis contributes to the CP, while other, for instance, is based on foundational ontological analysis, where recently a conceptual relation reference model was developed and implemented using conceptual relations templates [22] - on pursuing (B2). The challenge tackled in this paper is presenting a more comprehensive approach based on continuous corpus analysis, following the progressive details that constructing conceptual representations adds across time. With the typical scenario described as background, the ConceptME was designed to cope with those CPP building blocks within an extensible architecture.

4 Supporting Conceptualisation Processes: The ConceptME Approach

4.1 Conceptualisation Framework (CF)

The CF depicted below (figure 1) strengthens the advances of this research on methods and tools to support the representation of conceptual structures. This framework (figure 1) provides a structured and multidimensional view on the CP regarding its main phases, activities and artifacts, combining the terminological and knowledge representation view.

The CP is divided into four phases - concept elicitation, concept organisation, and concept sharing and concept negotiation - whose execution depends on a set of supporting activities. Two main processes are considered, the terminological process and the knowledge representation process. The terminological process encompasses methods for identifying/selecting lexical resources and its classification, as well as terminological extraction techniques. Moreover, terminological processes are also used to assist the negotiation activities during concept discussion. KR processes encompass activities to elicit, organise, share and discuss the conceptual representations. Concept elicitation can be supported by the terminological process helping overcome difficulties related to concept identification (such as naming,

meanings, contexts of use). In order to accomplish domain structuring, users could resort to available templates; however each domain has its own specificity which asks for specific relations. As the detail level of the domain description increases, it calls for real-time term contexts formatted to highlight the words or compound words, which could be used to detail existing representations, either to designate new concepts or new conceptual relations.

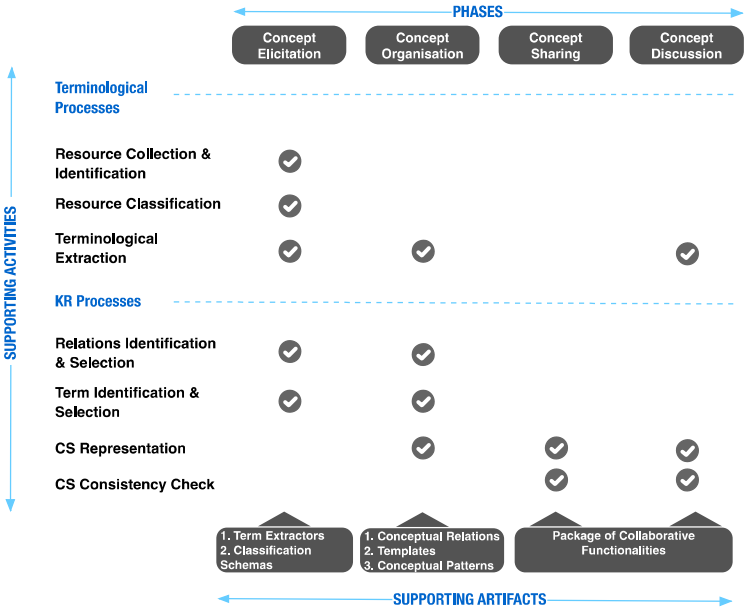


Fig. 1. The Conceptualisation framework (CF)

4.2 An Overview of the ConceptME

The ConceptMe is a technological platform that allows the user to create and share conceptual models resulting from conceptualisation processes. It holds a twofold commitment aligning terminology and knowledge representations within technological collaborative spaces in order to support and enrich the construction of domain conceptual models (Cmodels). In its core, the ConceptME is a wiki-based platform that combines semantic technologies with content and metadata management. The main objective is to support the collaborative development of Cmodels in an easy and intuitive way, giving focus to graphical knowledge representations and terminological methods accommodated into a service library. The platform enhances negotiation and discussion capabilities by means of specialised extensions. Cmodel manipulation, management and consistency analysis are also supported within the platform. However, the aim with this paper is not to fully describe the ConceptME, but rather to focus on specific approaches of how corpus-analysis enhances concept elicitation and discussion. Hence, the development of

Cmodels is supported by following a corpus-based approach (allowing users to associate relevant resources to their projects, performing extraction operations to retrieve candidate terms that can be used in their conceptualisation process). By pursuing a more pragmatic approach to knowledge representation, the ConceptME was endowed with the current approach attempting to reduce KR constraints by providing: i) a means of organising and classifying corpora; ii) tools to build conceptual structures from templates; iii) real-time term contexts to detail existing representations; iv) mechanism to discuss collected concepts and v) a model negotiation baseline enclosing a set of features (such as merging individual input structures, suggestion mechanisms, cross-checking corpus-based validation, auto-complete and categorisation, equivalents visualisation) to ensure simple negotiation mechanisms, towards a common shared model. A detailed description of the core features of the ConceptME is provided in [19],[12].

4.3 Corpus-Based Conceptual Modelling: Describing the Architecture

Over the last decade, along with the maturity level of NLP tools and a multi-disciplinary view on knowledge representation, corpus-based approaches for conceptual modelling have gained prominence in the literature, whether by new methodologies/frameworks [1] or information extraction algorithms for knowledge discovery [20]. Despite their quality, existing extraction algorithms usually require a large customisation effort as mentioned in [13]. The ConceptME approach in its turn was designed towards devising an architecture capable of providing a set of tools to domain experts that were almost exclusive to on computational linguistics experts. The base idea is on the assumption that some linguistic patterns could be found in texts denoting the existence of a concept or a relation. A verb or a verb plus a pronoun, for instance, could indicate a possible relation, while a noun could indicate a concept. Moreover, the pattern <*noun*><*verb*><*noun*> could indicate a brand new conceptual structure.

The figure bellow (figure 2) depicts the architecture designed to fit the ConceptME requirements to construct conceptual representations. A special effort was made to ensure: a) term context and metadata retrieval from unstructured data, and b) derived facts taken from achieved conceptual representations.

At this level, contexts are equivalent to the sentences extracted from the several resources made available for the CPP project, in relation to a specific term.

As mentioned earlier, the CCP requires an organised domain corpus as it could be used to validate the use of specific terms within conceptual representations. This pre-required activity includes tasks such as collecting, describing, storing and classifying provided resources. The main objective is to describe a specific resource regarding its bibliographic properties and later to classify it according to terminological characteristics. Upon creating the corpus, the resources should be indexed using (in this case) the apache Solr (http://lucene.apache.org/solr/). Afterwards, and at any time during the CCP, term contexts could be immediately retrieved in order to obtain clues on new possible concepts or relations related to the current concept and to the corresponding conceptual representation. Moreover, term contexts could additionally

support discussion around a specific concept, justifying its use by showing evidence of term occurrence in a corpus or inferring on concept semantic metadata (using an RDF triple store), or even highlighting patterns (*<noun><verb><noun>)* within the text where at least one of the terms in the pattern is already in use. This suggests that the overall conceptual structure is incomplete. This is the first iteration for the blend space construction, according to the ColBlend method.

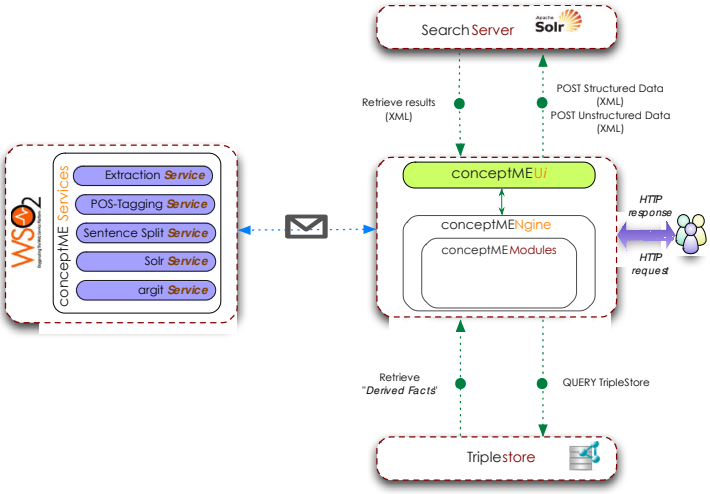


Fig. 2. The ConceptME architecture

Identifying Patterns in Contexts

To identify pattern in the retrieved term contexts, a simple method is proposed which depends on pre-processing tasks, such as sentence split, resource indexing and part-of-speech tagging. For this purpose, a set of services were developed to support the pre-processing tasks associated to index the resources of the domain corpus (POST) and to interact with the Solr engine in order to return the contexts associated to a specific term (RETRIEVE). The POST and RETRIEVE workflow are depicted below in figures 3 and 4 respectively.

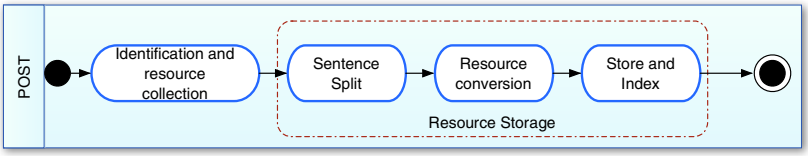


Fig. 3. POST workflow

The POST workflow encloses activities to support users in collecting relevant resources related to the intended conceptualisation goal. For each textual resource attached, the sentences are identified and the document is converted to a standard format (XML) and indexed in Solr.

The RETRIEVE workflow (figure 4) encloses activities to support the search, identification and retrieval of the context related to each term/concept in the current model.

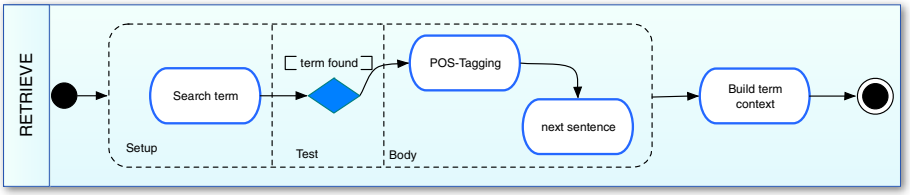


Fig. 4. The RETRIEVE workflow

The method retrieved works as follows: for each term found, the sentence containing the term is tagged. At the end, the context is retrieved and specific words/compound words are highlighted (coloured) according to the tagged value (for instance, Name, Verb, Adverb). This retrieving task could occur at several stages of the CP according to the following scenarios: a) any time a concept is included in the platform, its context could be retrieved and viewed; b) when linking two concepts, the ConceptME could suggest possible linking phrases (conceptual relations) if the terms co-occur; c) at any time during the construction of the CR, the user may call for neighbour terms of a certain concept, thus making it possible to collect possible related concepts; d) at any time during concept discussion on finding new concept variances.

The method for identifying and extracting patterns in the retrieved contexts calls for the aforementioned workflows and it is proposed as the following image illustrates (fig. 5).

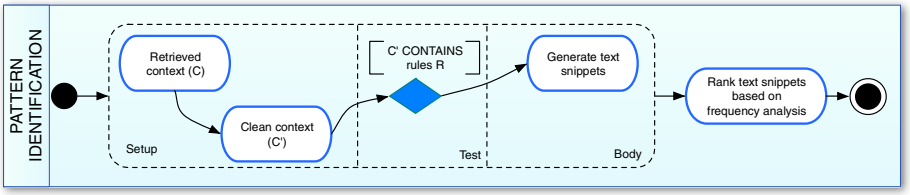


Fig. 5. Pattern identification workflow

For a context *C* retrieved for a term *t*, all numeric values *D,* punctuations *P* and other special characters and words from the stop-words list *L* will be eliminated from the context. Moreover, a stemming task will be performed. The resulting "cleaned context" *C'* will be segmented into text snippets *S* according to the following rules *R*:

a) *S* must contain exactly one occurrence of *t* and another term whose tag denotes a noun *n*; b) between *t* and *n* there should be at least a verb or verb-phrase *v*; c) the maximum length of *S* is *w* words; d) *S* could contain *t* at the beginning, middle or at the end of the text snippet. *S* will then be ranked according to the frequency of each *S* (fS_i). fS_i is calculated by counting how much times term *t* co-occurs with other term either present in the model or in *C* in each *S*. (see figure 6).

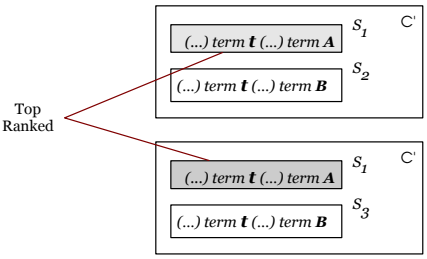


Fig. 6. Ranking patterns

5 Results

An experiment was conducted illustrating a possible scenario where the approach presented here could support the collaborative construction of conceptual representations. The objective was to: i) qualitatively assess the relevance of the description and understanding of the facts behind the approach; and ii) to legitimise the approach as a means to improve the creation of an agreed conceptualisation with a group of domain experts. Therefore, the illustration example started to be outlined as part of the H-Know (http://www.h-know.eu/) project. H-know was a European project (FP7-NMP-2007-SME-1) in the area of old building restoration and maintenance, particularly in the cultural heritage domain. It was a large project involving 15 partners from 5 countries and 7 work packages. Our research group has participated as an RTD partner, responsible for the work package 2, whose goal was to specify the methodology, ontology framework and the services for the H-Know solutions. This revealed to be a fitting backdrop to apply our approach, taking advantage of an existing domain corpus. The corpus was comprised by scientific, technical, juridical and didactic texts according to a predefined typology that took into consideration the domain's communicative and professional specificities. It had 532 000 tokens.

However, even though the current approach has been designed to "elaboration activities" of the blend space creation, the example described focuses on the private modelling task of each group. Nonetheless, the designed example, and similarly to other features already developed, can be used in the same manner after the model is shared. The scenario that would underpin the example was established, selecting two specific groups (G1 and G2) from two different partners (P1 and P2). There was a third group acting as moderators/observers (ourselves). Due to the very short time frame available, and restrictions of the several partners' agenda, only two groups were invited. The approach was explained to both groups, but while G1 had complete

access to the context-based features developed so far, the access to those features was denied to G2. Additionally, three documents selected from the urban rehabilitation corpus were posted (indexed) on Solr. While the groups were performing their modelling tasks, their actions were being monitored. Along the CP process, users added concepts and relationships between concepts, either on their own using extraction services or available templates. After releasing a term (here referred to as concept) on the canvas, several tasks may follow: a) providing a definition to the concept; b) completing the structure adding another term (concept) and a linking phrase between two terms (concepts). Context-analysis could be helpful on both tasks.

Still in the scope of urban rehabilitation, at a certain stage of CP, the conceptual structure started being defined of around the concept of *moisture control.* The task was initiated by adding the *moisture control* term. The main challenge was finding how and to which term *moisture control* should be linked to. For G1the solution was achieved after analysing the contexts by identifying terms (nouns) that co-occur with the added term (*moisture control*) and possible linking phases connecting them. The linking phases are typically verbs or expressions that match the following lexical patterns: i) a verb preceded or followed by a preposition or subordinating conjunction; ii) a verb preceded or followed by a coordinating conjunction; iii) a verb preceded or followed by a "TO"; iv) a verb preceded or followed by a determiner; v) a verb preceded or followed by another verb.

The information retrieved by the contexts related to *moisture control* provided the following information:

"*However, good* ***moisture control*** *design depends on (pattern) a variety of parameters (noun) such as climate conditions (noun) and construction type (noun) which changes (verb) from region (noun) to region (noun).*"

Within the platform, the nouns, verbs and lexical patterns described above appear in colour. Here, a tag was added in brackets for better reading. According to the context retrieved, it is possible to infer that "*moisture control depends on climate conditions and construction type*", since *climate conditions* and *construction type* co-occur with *moisture control* mediated by the lexical pattern "*depends on*".

Additionally, the following context is also provided:

Effective ***moisture control*** *has to (pattern) respond to (pattern) the exterior as well as the interior moisture loads (noun) acting on (pattern) building constructions (noun).",* from which it can be deduced that "*moisture control responds to moisture loads".*

G2 started a discussion on the possible terms and linking phrases when browsing the available documents, while G1 was indeed faster in finding the appropriate term with a high level of agreement as the achieved conceptual structure turned out to be justified by the information that came from the contexts.

After analysing the context, G1 added the following propositions: a) *Moisture control depends on climate conditions; b) Moisture control depends on construction type; c) Moisture control responds to moisture loads.*

The same approach was used for the "*moisture load*" term, which led to the following assertion: "***Construction moisture*** *is a* ***moisture load***." The process continues for "construction moisture" until the conceptual structure is achieved.

The focus of the designed example followed the perspective of evaluating the approach according to previously defined top-level requirements. The scenario was based on the simple observation of the process of creating conceptual representation, with or without a specific variable. In this case, the variable was the presence or lack of specific features related to the retrieval of term contexts and pattern identification. The results from the described example showed that the presence of such a corpus-based approach is an interesting and promising add-on to assist groups of users in creating conceptual representations.

6 Conclusions

This article discusses the problems and requirements related to supporting the collaborative conceptualisation process, more specifically, how to aid concept elicitation and concept discussion tasks. A corpus-analysis approach was presented providing methods and tools to domain specialists to help them continuously align the knowledge models under construction and the texts on a certain domain. Defining a set of terminologically based mechanisms to assist users with the creation of conceptual structures in conceptual modelling tasks introduced some precision to informal knowledge representation approaches, such as concept maps (the main formalism used in the ConceptME). Together with other services (extraction services, templates, ontological guidance and specific negotiation services) developed by this team, the perspective shown here provides an interesting and complementary approach that bypasses the difficulty of starting a conceptualisation. Additionally, it provides a baseline for implementing mechanisms such as term suggestion and/or auto-complete features, necessary for "running" the blend space of the ColBlend method. Furthermore, this modular and service oriented architecture moves towards a new generation of technological tools, compliant with the requirements to support the CCP within a perspective that promotes human interaction rather than a full automation of the CP tasks. The preliminary empirical study conducted with a small group of users was fundamental to clarify the roles and tasks within the process and the level of expertise required to perform each task. Future work will be focused on experiments aimed at obtaining further feedback from specialists, in several domains, to improve and fine-tune the methods and tools developed so far. Preliminary experiments have already been conducted and the results have shown that this approach improved the specialist performance in concept elicitation and discussion. Regarding conceptual modeling, this research surely proposes a refreshing view on this subject.

References

1. Aussenac-Gilles, N., Biébow, B., Szulman, S.: Corpus Analysis for Conceptual Modelling (2000)
2. Aussenac-Gilles, N., Biébow, B., Szulman, S.: Revisiting Ontology Design: A Method Based on Corpus Analysis. In: Dieng, R., Corby, O. (eds.) EKAW 2000. LNCS (LNAI), vol. 1937, pp. 172–188. Springer, Heidelberg (2000)

3. Buitelaar, P., Magnini, B.: Ontology Learning from Text: An Overview. In: Paul Buitelaar, P., Cimiano, P., Magnini, B. (eds.) Ontology Learning from Text: Methods, Applications and Evaluation, pp. 3–12 (2005)
4. Buitelaar, P., Cimiano, P., Haase, P., Sintek, M.: Towards Linguistically Grounded Ontologies. In: Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvönen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E. (eds.) ESWC 2009. LNCS, vol. 5554, pp. 111–125. Springer, Heidelberg (2009)
5. Camarinha-Matos, L.M., Afsarmanesh, H., Galeano, N., Molina, A.: Collaborative Networked Organizations – Concepts and Practice in Manufacturing Enterprises. Computers & Industrial Engineering 57(1), 46–60 (2009), doi:10.1016/j.cie.2008.11.024
6. Camarinha-Matos, L.M., Afsarmanesh, H.: Collaborative Networks. In: Wang, K., Kovacs, G., Wozny, M., Fang, M. (eds.) Knowledge Enterprise: Intelligent Strategies in Product Design, Manufacturing, and Management. IFIP, vol. 207, pp. 26–40. Springer, Boston (2006), http://dx.doi.org/10.1007/0-387-34403-9_4
7. Ceausu, V., Desprès, S., et al.: Learning term to concept mapping through verbs: a case study. Citeseer (2009)
8. Gillam, L., Tariq, M., Ahmad, K.: Terminology and the Construction of Ontology. Terminology 11, 55–81 (2005)
9. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Q. 28(1), 75–105 (2004)
10. Hjørland, B.: Semantics and Knowledge Organization. Annual Review of Information Science and Technology 41(1), 367–405 (2008)
11. Horrocks, I.: Semantic web: the story so far. In: Proceedings of the 2007 International Cross-Disciplinary Conference on Web Accessibility (W4A), New York, NY, USA, pp. 120–125 (2007)
12. Costa, L., Sousa, C., Soares, A.L., Pereira, C.: ConceptME: Gestão colaborativa de modelos conceptuais. 11th Conferência da Associação Portuguesa de Sistemas de Informação (2011)
13. Baroni, M., Murphy, B., Barbu, E., Poesio, M.: Strudel: A corpus-based semantic model based on properties and types. Cognitive Science 34(2), 222–254 (2010)
14. McEnery, T., Wilson, A.: Corpus Linguistics: An Introduction. Edinburgh University Press (1996)
15. Nabila, N.F., et al.: Enriching non-taxonomic relations extracted from domain texts. In: 2011 International Conference on Semantic Technology and Information Retrieval (STAIR), pp. 99–105. IEEE (2011)
16. Nonaka, I.: A dynamic theory of organisational knowledge creation. Organisation Science 5, 14–37 (1994)
17. Pereira, C., Soares, A.L.: Ontology Development in Collaborative Networks as a Process of Social Construction of Meaning. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM-WS 2008. LNCS, vol. 5333, pp. 605–614. Springer, Heidelberg (2008)
18. Pereira, C., Sousa, C., Soares, A.L.: Supporting Conceptualisation Processes in Collaborative Networks: a Case Study on an R&D Project. International Journal of Computer Integrated Manufacturing, 1–21 (July 2, 2012), doi:10.1080/0951192X.2012.684714
19. Sá, C., Pereira, C., Soares, A.L.: Supporting Collaborative Conceptualization Tasks through a Semantic Wiki Based Platform. In: Meersman, R., Dillon, T., Herrero, P. (eds.) OTM 2010. LNCS, vol. 6428, pp. 394–403. Springer, Heidelberg (2010)

20. Sarawagi, S.: Information Extraction. Found. Trends Databases 1(3), 261–377 (2008)
21. Sinclair, J.: Corpus, concordance, collocation. Oxford University Press (1991)
22. Sousa, C., Soares, A., Pereira, C., Costa, R.: Supporting the Identification of Conceptual Relations in Semi-formal Ontology Development. In: ColabTKR 2012 - Terminology and Knowledge Representation Workshop at International Conference on Language Resources and Evaluation, Istanbul, Turkey (2012)
23. Swidler, A., Arditi, J.: The New Sociology of Knowledge. Annual Review of Sociology 20(1), 305–329 (1994), doi:10.1146/annurev.so.20.080194.001513
24. Yu-Liang, C.: Elicitation synergy of extracting conceptual tags and hierarchies in textual document. Expert Systems with Applications 32(2), 349–357 (2007)

Effect of Demography on Mobile Commerce Frequency of Actual Use in Saudi Arabia

AbdulMohsin Alkhunaizan and Steve Love

School of Information Systems, Computing & Mathematics,
Brunel University, UK
{AbdulMohsin.Alkhunaizan,Steve.Love}@Brunel.ac.uk

Abstract. When developing and aiming to achieve success in the arena of mobile commerce, user behaviour is one of the main aspects for consideration. This research seeks to analysis whether individuals' (gender, age, education level) influences their mobile commerce usage within the context of Saudi Arabia. The individuals analysed are own smartphone. We further present three hypotheses that investigate whether demographic factors have a significant statistical impact on the perception of those factors for mobile commerce acceptance in the Kingdom of Saudi Arabia. Survey data were collected from 574 participants in several cities across Saudi Arabia. The results emphasise that age affect statically on the actual usage. However, gender and education level all considerably not affect on the mobile commerce actual usage.

Keywords: m-commerce, Saudi Arabia, mobile usage.

1 Introduction

In Saudi Arabia, the concept of mobile commerce (m-commerce) has incited significant interest over recent years and benefitted from considerable development. According to Kathleen [1] , in Saudi Arabia 60% of the mobile users own smart phones. By comparison, the market penetration of Smartphones in the U.S. was 44 % according to the study. Moreover, in Saudi Arabia 85% of the Smartphone users have access to the internet [2] , which represent one of the mobile commerce applications. The total number of mobile subscriptions grew to around 53.7 million by the end of 2011, up from only 2.5 million in 2001[3] .

Mobile users are able to gain access to information in real-time, and can communicate and purchase anywhere, at any time. As can be seen through such advantages, m-commerce provides a key innovation, and induces a number of opportunities, particularly for organisations. Furthermore, it is considered that user satisfaction may be significantly enhanced through m-commerce, thus providing the potential for user behaviours to be promoted and user satisfaction enhanced.

In a number of industrialised, technology-savvy countries, namely China, the USA, and those within Europe, various theoretical frameworks have been devised in an attempt to understand the factors affecting the adoption of information and communication technologies [4][6] . Accordingly, the overall aim of this paper is to examine the effect of demographic differences in specific consideration to Saudi Arabia.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 125–131.
DOI: 10.1007/978-3-642-36981-0_12

2 Demographic Differences

2.1 Gender

Gender has been considered through a number of studies with the aim of examining whether or not there are differences between men and women in regard to technology utilisation. Various academics have examined the role adopted by gender in the utilisation and implementation of technology [7][9]. Moreover, a number of researches have emphasised that gender is markedly impact when considering the use and implementation of technological in a business-related context. In this regard, it is illustrated that males utilise computers more so than women [10]. In addition, the differences between genders have been highlighted in regard to internet utilisation [9]. Notably, [8] have found that perceived usefulness in regard to behavioural intention is controlled by gender. On the other hand, it was found that women and men do not exhibit important differences in shopping behaviours in regard to mobile technology [11].

In the context of this study, the works of various scholars—including [12][14] will be followed, with gender taken into account as a social, independent variable in an attempt to explain the differences between technology users, the researcher proposes the following hypothesis.

H1: *The Actual Use of m-commerce will be greater amongst males than females.*

2.2 Age

Finch (1986, cited in [13]) stated that age may be utilised as a factor to rationalise a certain social grouping or process, or individual or collective behaviour. Various IS studies have found that age has important, direct and moderating effects on the adoption and use of behaviours, as well as behavioural intention [7][8][10][11]. Furthermore, found that the 15 -17 years age group were most likely to utilise computers in the USA, with the 26–35 age group following subsequently. Similarly, in several West European countries [15] found that older demographic groups are less inclined to use computers and the internet compared to the younger population. In the United Kingdom, 85% of 16–24 year olds have internet access, but among older age groups only 15% of 65–74 years, and 7% of those 75 years and older have access [15]. In addition, [13] found that most broadband subscribers are aged 25–45 years. Accordingly, the present study predicted that, younger and middle-aged individuals are expected to exhibit differences in m-commerce usage, with the younger groups expected to shows more indifferent use, whereas older individuals will use m-commerce less often in Saudi Arabia. Accordingly, through this research, the following hypothesis is considered:

H2: *There will be a difference between the m-commerce users of various age groups.*

2.3 Education

Burgess (1986; cited in [13]) emphasise that those possessing educational qualifications are far more likely to ascertain a greater occupational position. Accordingly, it is believed that such individuals are therefore more likely to implement new technologies. Studies carried out previously in the arena of

technology utilisation suggest a strong link between usage, and the level of education and technological ownership [7] . With this in mind, the work of [7] states that people with higher educational qualifications use computers more often than those with lesser qualifications. In this context, it is then emphasised that education is largely recognised as being one of the most fundamental drivers in South Korea [16] [17]. In specific consideration to internet access, it has been suggested that those individuals with either secondary or tertiary education are more likely to have access to the internet [18] . In addition, it has also been argued that education may be viewed as an independent variable able to justify the differences between those who utilise technology and those who do not [13] [14] . With this in mind, the hypothesis as stated as follows:

H3: *A difference between m-commerce usages will be seen in terms of education.*

3 Research Methodology

3.1 The Survey Instrument Development

A quantitative survey approach was employed in this study. And to analyse the data objectively, statistical software known as Statistical Package for the Social Sciences (SPSS) was used. Questionnaires were distributed online and manually to approximately 1700 mobile smartphone users in Saudi Arabia. Of that number, 820 were returned (48.24% response rate) and 574 were usable. The other 246 responses were eliminated because they were the response pattern noticed a careless response or incomplete.

3.2 Sample Demographic Profile of Participants

Table 1 describes the profile of the participants based on gender, age, and education level. This part will help to inform the propositions in that mobile commerce acceptance differs by personal (gender, age, education level).

Table 1. Demographic characteristics of the participants (N=574)

Demographic	***Category***	***Frequencies***	***%***
Gender	Male	249	43
	Female	325	57
Age	15-18 years	65	11
	18-25 years	260	45
	26-35 years	198	34
	36-45 years	39	7
	46+ years	12	2
Education	Less than high school	31	5
	High school	121	21
	Diploma	43	7
	Bachelor	303	53
	Postgraduate	76	13

Of the 574 respondents, a slight majority were female (57%, n=325) while 43% (n=249) were male. The majority were young in the age group of 18-25 years (45%, n=260), followed by 26-35 years (34%, n=198), 15-18 years (11%, n=65), 36-45 years (7%, n=39), while only 2% (n=12) were aged 45 years and over. As for education, the majority (53%, n=303) held undergraduate level degrees, 13% (n=76) held postgraduate degrees, 21% held high school certificates, 7% (n=43) held a diploma, while 5% (n=31) had less than high school qualifications.

Hypothesis 1: The Actual Use of m-commerce will be greater amongst males than females

A chi-square test of independence was performed to examine the relation between gender and frequency of actual use. The relation between these variables was not significant. Table 2 represents the Pearson's chi-square test that confirmed that there was no significant difference between males and females in their frequency of actual use ($X2$ (3, N=574) = 5.60, p = .133).

Table 2. Crosstab Analysis: Gender by Actual Use of Mobile commerce

			Gender		
			Male	Female	Total
Actual use	Monthly	Count	6	12	18
		% within gender	2%	4%	3%
	Weekly	Count	4	10	14
		% within gender	2%	3%	2%
	More than Once per Week	Count	12	28	40
		% within gender	5%	9%	7%
	Daily	Count	227	275	502
		% within gender	**91%**	**85%**	88%
Total		Count	249	325	574
		% within gender	100%	100%	100%

Note: Pearson Chi-Square = 5.60 $p < .05$; N=574.

Hypothesis 2: There will be a difference between the m-commerce users of various age groups

A chi-square test of independence was performed to examine the relation between age group and frequency of mobile commerce actual use. The relation between these variables was significant. Table 3 shows the Pearson's chi-square test was significant indicating a difference in age groups and frequency of actual use ($X2$ (12, N=574) = 22.44, $p < .05$).

Table 3. Crosstab Analysis: Age Group by Actual Use

			Age Group					
			15-18	18-25	26-35	36-45	45+	Total
Actual use	Monthly	Count	1	6	6	3	2	18
		% within age	2%	2%	3%	8%	17%	3%
	Weekly	Count	0	7	4	3	0	14
		% within age	0%	3%	2%	8%	0%	2%
	More than once per week	Count	3	20	11	4	2	40
		% within age	5%	8%	6%	10%	17%	7%
	Daily	Count	61	227	177	29	8	502
		% within age	**94%**	**87%**	**89%**	**74%**	**67%**	88%
Total		Count	65	260	198	39	12	574
		% within age	100%	100%	100%	100%	100%	100%

Note: Pearson Chi-Square = 22.44, $p < .05$; N=574.

Table 4. Crosstab Analysis: Education by Actual Use

			Education					
			Less than high school	High school	Diploma	Bachelor	Postgraduate	Total
ACTUAL USE	Monthly	Count	0	5	2	8	3	18
		% within education	0%	4%	5%	3%	4%	3%
	Weekly	Count	0	1	1	8	4	14
		% within education	0%	1%	2%	3%	5%	2%
	More than once per week	Count	0	6	2	26	6	40
		% within education	0%	5%	5%	**9%**	**8%**	7%
	Daily	Count	31	109	38	261	63	502
		% within Education	**100%**	**90%**	**88%**	**86%**	**83%**	88%
Total		Count	31	121	43	303	76	574
		% within Education	100%	100%	100%	100%	100%	100%

Note: Pearson Chi-Square = 11.97, $p < .05$; $p = .448$; N=574.

Hypothesis 3: A difference between m-commerce usages will be seen in terms of education

A chi-square test of independence was performed to examine the relation between education and frequency of actual use. The relation between these variables was not significant ($X2$ (12, N=574) = 11.97, p = .448). Table 4 shows that results suggest that those with lower education tend to report higher actual use, but this difference is not statistically significant.

4 Discussion

This paper presented the findings obtained from the data analysis of the survey that was conducted to examine demographic differences usage of the m-commerce services in Saudi Arabia. The findings were shown in several sections. The first step was a discussion of the profile of participants. After that, as noted earlier, various scholars have stated gender as being relevant in the use of technology. In contrast, however, this study has established no differences in terms of gender and the utilisation of m-commerce. This research has established that the greatest percentage of actual use was 'daily use', which was the case across all age groups considered in the study, although those in the 18–35 years bracket cited this most often, with those aged 36–45+ years stating this the least often. This means that older individuals are less likely to make use of m-commerce, which could be rationalised when considering a potentially lower level of awareness concerning the advantages of m-commerce, as well as a shortage of necessary skills for such utilisation. Furthermore, education has been found to have no significant influence on the usage of m-commerce, meaning that education and/or qualifications are not recognised as relevant in the adoption of m-commerce within Saudi Arabia.

5 Conclusion

In this paper, the three main hypotheses that affect the usage of mobile commerce have been reviewed with the aim of investigating such a service in Saudi Arabia. A sample population of 574 individuals were chosen in this study from a wide geographical area. Three hypotheses were presented that investigate how demographic differences affect the usage of mobile commerce. One of the hypotheses was observed to be statistically significant difference in team of the age. However, in terms of gender, and education level, no significant differences were observed.

References

1. Google study finds affluent middle east countries among most enthusiastic smartphone users [Internet].: google (2012), http://www.insidemobileapps.com/2012/05/23/google-study-finds-affluent-middle-east-countries-among-most-enthusiastic-smartphone-users/ (updated May 23, 2012)

2. Google analyzes how people use smartphones in different countries [Internet]: google (2012), http://www.webpronews.com/google-analyzes-how-people-use-smartphones-in-different-countries-2012-05 (updated November 26, 2012)
3. CITC. The ICT sector in the kingdom of saudi arabia. Communication and information technology commission, Saudi Arabia (2011)
4. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 319–340 (1989)
5. Carlsson, C., Carlsson, J., Hyvonen, K., Puhakainen, J., Walden, P.: Adoption of mobile devices/services—searching for answers with the UTAUT. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences, HICSS 2006. IEEE (2006)
6. Lu, J., Liu, C., Yu, C.S., Wang, K.: Determinants of accepting wireless mobile data services in china. Information & Management 45(1), 52–64 (2008)
7. Morris, M.G., Venkatesh, V.: Age differences in technology adoption decisions: Implications for a changing work force. Person Psychol. 53(2), 375–403 (2000)
8. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: Toward a unified view. MIS Quarterly, 425–478 (2003)
9. Jackson, L.A., Ervin, K.S., Gardner, P.D., Schmitt, N.: Gender and the internet: Women communicating and men searching. Sex Roles 44(5), 363–379 (2001)
10. Venkatesh, V., Davis, F.D.: A theoretical extension of the technology acceptance model: Four longitudinal field studies. Management Science, 186–204 (2000)
11. Bigne, E., Ruiz, C., Sanz, S.: The impact of internet user shopping patterns and demographics on consumer mobile buying behaviour. Journal of Electronic Commerce Research 6(3), 193–209 (2005)
12. Dwivedi, Y.K., Papazafeiropoulou, A., Gharavi, H., Khoumbati, K.: Examining the socio-economic determinants of adoption of the &# 39; government gateway &# 39; initiative in the UK. Electronic Government, an International Journal 3(4), 404–419 (2006)
13. Dwivedi, Y.K., Lal, B.: Socio-economic determinants of broadband adoption. Industrial Management & Data Systems 107(5), 654–671 (2007)
14. Al-Shafi, S., Weerakkody, V.: Factors affecting e-government adoption in the state of qatar (2010)
15. Carveth, R., Kretchmer, S.B.: The digital divide in western europe: Problems and prospects. Informing Science 5, 239–249 (2002)
16. Choudrie, J., Lee, H.: Broadband development in south korea: Institutional and cultural factors. European Journal of Information Systems 13(2), 103–114 (2004)
17. Choudrie, J., Papazafeiropoulou, A.: Lessons learnt from the broadband diffusion in south korea and the UK: Implications for future government intervention in technology diffusion. Electronic Government, an International Journal 3(4), 373–385 (2006)
18. Choudrie, J., Dwivedi, Y.K.: The demographics of broadband residential consumers in a british local community: The london borough of hillingdon. Journal of Computer Information Systems 45(4), 93 (2005)

Specialized Knowledge Systems – A Model for Intelligent Learning Management within Organizations

Isabel Mendes[1], Henrique Santos[1], and Celina Pinto Leão[2]

[1] Centro Algoritmi, Department of Information Systems,
School of Engineering of University of Minho,
Guimarães, Portugal
mimp1844@gmail.com,
hsantos@dsi.uminho.pt
[2] Centro Algoritmi, Department of Production and Systems,
School of Engineering of University of Minho,
Guimarães, Portugal
cpl@dps.uminho.pt

Abstract. This paper presents an insight of the perspectives and potentialities that may arise from the use of tools based on semantic technology, for an autonomous management of the learning and the knowledge produced within organizations. Thus, the conceptual basis of a model for the development of knowledge systems will be presented, oriented for an intelligent learning management. The scope of application of this model is software houses. Its main function is to support specialized training having as basis the use of contents adjusted to particular needs. The pillars of its functioning are essentially based on the use of semantic technologies and the dynamic self-regulation of its contents. Similarly, and bearing in mind the importance of learning theories for the scope of this study, we will include an approach to the connectivist theory, in order to better understand behavior and trends influencing learning and pedagogical processes in the digital era.

Keywords: Connectivism, e-Learning, Knowledge Systems, Learning Networks, Ontologies, Semantic Technologies.

1 Introduction

Globalization and strong market competitiveness have brought about major challenges to organization where the technology assumes a leading role. More and more business companies depend on the use of technology to manage their businesses and thus meet the challenges that arise from the need the knowledge in the digital era. This involves information and communication management and inevitably education and training too. We are undergoing a strong innovation period and one of fast technological renewal [1], which, in turn, makes companies and stakeholders be aware of the need of a re-organization and prompt adaptation to its learning mechanisms in order to follow the fast pace associated with these changes.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 133–141.
DOI: 10.1007/978-3-642-36981-0_13

In the organizational field, this view associated with the interest in innovating and renewing, learning continuously and valuating experience, is embedded in learning organizations. Their principles are based upon a systemic logic, where the individual and the whole (organization) work around those pillars in an interactive cycle towards business success [2]. On the other hand, modern theories of economic growth indicate the human development as a factor of impact for the progress of societies [3]. In light of this perspective, knowledge and learning are viewed as essential requirements to ensure sustainability and company differentiation in evermore demanding markets. The coordination of the aforementioned elements, in connection with experience may work as important aspects of influence in the development of new competencies, boosting the new spheres of value within organizations.

Therefore, focused on the importance of training within organizations and, considering the software houses` profile, this study is focused on the creation of a model that works as a basis for the development of training structures, able to meet the expectations and needs of such companies. These structures, supported by semantic technologies, are oriented towards an autonomous and collaborative learning, providing liability to the contents. Furthermore, considering the specificities of this type of business, where learning has to be continual and the access to knowledge immediate updated and frequently shared. The use of such a model will provide an answer to a market sector with huge requirements in this domain.

In a broader perspective, the aim of this study is to lead organizations towards a changing process in regard to the valuation of their internal know-how. Moreover, we aim at increasing organizations´ quality levels, by fostering their competitive and innovative potential. In this framework, the main objective is to develop collaborative learning mechanisms, contextualized with the digital era, in order to respond to the challenges of a knowledge-based economy.

Having the above-mentioned trends as a basis, in section 2, we identify the research problem and the reason of its importance, namely for organizations associated with software engineering; in section 3 we carry out an approach on the relevance of semantic technologies for the scope of this study; in section 4 we focus on the importance of the role of the connectivist theory in order to understand learning trends in the digital age. In this respect, what is important is to analyze the environments where it occurs, the nature of the resources and its influence on behavioral change; in section 5 the conceptual structure of the model will be presented and, finally, section 6 is dedicated to the results and the expected benefits arising from the implementation of this type of tools within organizations.

2 Research Problem

Technological and behavioral changes arising from the appearance of social semantic web, point out the need to acquire new paradigms and learning patterns with impact to intra-organizational training. Lifelong learning is a crucial mainstay of knowledge society [4]. Its importance is determining for the adaptation to the new job contours and to ensure the professional framing in environments highly dependent on technological mutations and its interaction with knowledge networks.

Expectations around the creation of flexible and user-friendly environments facilitating autonomous learning, having social semantic web fundamentals as basis, are enormous. Curiosity around this thematic has been object of interest in different research domains, by its innovative nature and, mainly, by the potentialities anticipated for the field of education and training, considering the importance of lifelong learning within the framework of development of modern societies. Among those areas of interest are Education, Psychology, Computing Sciences and Engineering [5]. The latter one has been seen as having particular interest for the possibility of using tools incorporating characteristics of social software and semantic technologies, as a support to teaching and learning of topics related to its field of knowledge.

The focus of this work is located precisely there, being centered on the importance such type of applications may bring to professional environments of technological development (e.g. software houses), where knowledge is hold as a critical factor for business sustainability and success. Its relevance is based upon the need to obtain specialized training, contextualized with the areas of work development, allowing access to shared and interconnected knowledge bases, sharing and interlinked to be used in a just-in-time logic. [6].

Taking into consideration the abovementioned, the research problem is centered on solutions that improve learning processes in environments having that specificity. Its main challenge lies on the possibility of removing some weaknesses existing at the level of reliability and regulation of contents made in e-Learning environments. In this way, it is expected to respond to the need and expectations of the market in regard to those issues. In line with this idea, the type of response that current technologies provide to autonomous learning needs should be analyzed, as well as how teaching and learning occur in this type of environments.

Bearing this in mind, some advances in this domain are presented next, as well as the value perspective that semantic technologies may bring to it, namely through the use of ontologies.

3 Social Semantic Web: Trends and Potentialities for New e-Learning Environments

The advent of Social Semantic Web brings about a perspective of change for the new online learning environments. Its role is determining for the competitiveness of companies, aiming at responding to the requirements of economies highly dependent on technology and knowledge.

Recent studies in the e-Learning domain emphasize the benefits that semantic technologies may bring to this field [7] [8] [9], namely due to the possibilities they offer when ensuring a better correlation and interpretation of learning contents based on its semantic [9].

The conceptual basis of web semantics resides in the attribution of well-defined meanings to information, allowing people and machines to improve their cooperation [10] [11]. Its structure integrates a set of technologies awarding a different logic to the

representation of knowledge. In the frame of semantic technologies, the importance of ontologies is highlighted as they are considered fundamental pillars in its functioning structure.

An ontology is the formal and explicit description of terms in a certain knowledge domain as well as its restrictions and properties [12]. They are used as ways of representing knowledge [12] [13] and, therefore referred to as the basis of Semantic Web [14]. In general, the information available on the Web is represented in a natural language in order to be understood by people. However, it should obey a formal representation in order to be understood by machines, whose main objective is to process, store and organize it in an efficient manner.

One of the fundamentals of Semantic Web consists in creating common formats, based on standards able to better integrate and combine data extracted from different sources [15]. This allows a person or a machine to access a database and to move through a set of unending databases related to the same issue [15]. The semantic enrichment of contents and the use of semantic technologies promoting a better interpretation of those contents, lead to more efficient research results [16]. This approaches Semantic Web to the concept of Intelligent Web.

On the other side, Social Web has marked the introduction of a set of services and applications emphasizing communication and online share (e.g. blogs, wikis, online social networks, folkonomies). It is distinguished by the assurance and interoperability of the connection established between these different services [4] [17]. Its appearance initiated a basis of collective knowledge sustained by collaboration and sharing of participants` individual contributions.

The Social Web has thus triggered a set of conditions for the development of structures of "collective intelligence", oriented to particular knowledge areas [9]. This new gamma of applications named "collective knowledge systems" should use the same Semantic Web reasoning and knowledge techniques [9]. On the other side, when analyzing the success of social networks, as shown by the Alexa statistics [18], one easily infers the potential that can be withdrawn from the aggregation of those concepts, transferred to other domains. Our approach will focus on the importance it may bring to the field of autonomous learning within organizations, starting from their specific interests in this regard.

Accordingly, for the scope of this study, the interest in structures integrating these technologies consists in different potentialities able to aggregate individual knowledge and promote intelligent knowledge management with relevance to a particular context. Furthermore, the development of tools materialized in those assumptions will bring a higher level of enrichment to the organizational activity. Additionally, the access to new knowledge degrees, expected to be more updated and specialized, is operationalized by search mechanisms and intelligent dissemination of learning resources. The learning processing is built upon the interception of principles based on collaboration, sharing and interaction. These principles are confined to the function and value that specialized knowledge systems may bring to organizations with the previously identified technological nature.

In this line of thought, knowledge consolidation may happen just-in-time and in loco, through the processing and application of concepts in a live context.

4 Behavioral and Learning Trends in the Digital Era

Technology has been uprising the way people learn, communicate and interact with one another. The rapid changes technology have faced, its ubiquity and interdependence with daily life sectors, make it an important variable for the integration of productive agents (e.g. people and organizations) in society. All these changes have originated different needs and requirements in terms of training. This makes organizations face new challenges and opportunities specially linked to valuation of knowledge and competencies of their human resources.

The level of popularity awarded by social software and the information boom that has arisen from such an active participation reflects the behavioral and social changes that have occurred in the way people relate with each other and also in the way they deal and interact with web knowledge. The analysis of this phenomenon is important as it helps to understand the mechanisms of learning processing and the current channels for knowledge dissemination.

A growing curiosity results thereof, for the search of solutions able to foster the use of models embedded in social networks, in turn capable of being adopted in other contexts. In view of that, analyzing that value that arises from the combination of the software in the network and consequent social phenomenon underlying its usage, some studies recognize the real benefit that can hence be extracted for Education [19] [20].

In view of that, if we observe the way users access and interact with resources available on the Web, some trends can be noticed in regard to the characteristics people expect to find there. Thus, these need to be attractive, in order to catch immediate attention; precise, with a clear and objective message; attractive, in a combination of form, color and organization; accessible and almost immediately traceable [21]. In the current development framework, the relationship between learning and technology supporting interactivity, collaborating and sharing assumes particular importance in the present and future teaching/learning process [21] [17].

These transformations of environments and behaviors have resulted in the appearance of new streams of thought defending the importance of approaching learners` educational reality to the current technological context. This stream of thought based on the connectivist theory analyses technology as an integrant part of the learning process and a fundamental piece to enable knowledge flow through the connections produced on the network [22]. According to this view, knowledge is seen as a pattern of specific relations and learning as the creation of new connections and standards where its enrichment results from the capacity of moving through networks [22].

In the different scenarios of modern life, learning is stored and manipulated by technology (e.g. inside an organization or in a database). Being on the network and interconnected, makes the changes that information is constantly exposed to, have an effect on the whole [22]. Therefore, in an age in which everything changes rapidly, the capacity to create and update knowledge by means of distributed sources of information; to develop new information standards and to recognize changes to

existing ones, such as, creating connections between interest communities, are vital characteristics for the learning in the digital era [22] [23].

These new approaches break up with traditional values of formal and institutional teaching, as we know it, pointing to the need of re-designing and projecting learning to more interactive and dynamic environments, where technology cannot be excluded.

Thus, the connectivist approach preconizes a set of principles based upon a vision of community, network, interaction and sharing, whose essence has strong implications in the education and training sectors. The importance of its values has a direct relation with the development of this study, by the herewith reflected need of trying to understand the principles in which learning mechanisms and processes are based, as well as knowledge diffusion in the contemporary technological context.

5 Conceptual Model

A knowledge system (with self-regulation of contents) oriented to a particular training, can be defined as a social learning platform that automatically organizes and provides the knowledge used in that areas of training [24].

From the point-of-view of knowledge enhancement within an organizational environment, the improvements, at the level of technology and communication, represent a competitive advantage for organizations [25]. In this regard, these features can also be a source of stimulus to foster innovation networks [26]. According to this perspective, and from the presented theoretical concept perspective, we believe that the adoption of these environments in a professional context can bring beneficial results for organizations. The organization`s interest lies primarily in the possibility of being able to implement an inside system training, in line with their areas of interest.

As discussed above, technological and educational indicators identify the importance of (re)thinking and (re)orienting teaching/learning strategies, adapting them to the contexts of the digital era. Moreover, it re-evaluates the methodological and pedagogical procedures to find new motivations and stimuli to acquire/produce knowledge. These new approaches indicate a trend towards less formal models [27], where the enhancement of learning occurs in the daily practice and in the opportunity to participate and interact in different networks of interest (social, personal, professional, etc.).

Nowadays, life cycles of knowledge, in some areas, happen faster than ever. The need for continuous training and readjustment of competencies throughout working life are critical elements for the integration of the individual in the profession and society. For this reason, the ability to promptly explore and assess knowledge relevant to the training is a feature increasingly appreciated [22].

Considering this combination of approaches, the baselines for the development of this study focus on the value of knowledge of experience and of learning as an essential condition for the development of the activity of the company. The dynamic management of these elements results in parallel, with other factors, from the enrichment and sharing of contents in a particular area of knowledge.

Figure 1 gives an overview of the conceptual framework of the model to be developed.

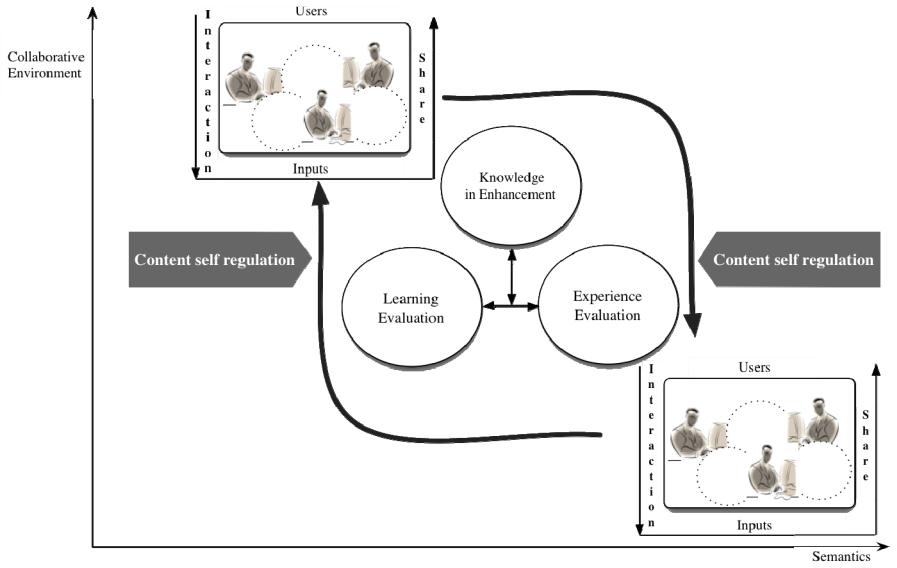


Fig. 1. Conceptual Model of Knowledge System with Content Self-Regulation (Source: Adapted from [9])

A self-regulating system of content to be fed, consulted and shared by members of the organization can support the creation of a collective knowledge base, and support the development of skills to raise the quality standards of the company. The main challenge for organizations is based on a pro-active and constructive participation of its members and on the creation of relevant contents to their activity.

Furthermore and having the abovementioned conceptual model as a basis, the use of structures supporting development in specialized collaborative environments, combining semantic and socially based technologies, play an important role for the emerging learning contexts where synergies between people and machines are created [9]. People have the role of both producers and users of knowledge, learning and communicating interactively. The system stores, interprets, organizes and manages the volume of information produced. The interoperability between these different technologies, as well as the ability to integrate and interconnect data available on the network through combination and logical inferences, produces new information flows more oriented towards the needs of the agents in those professional communities.

6 Conclusions

The approaches presented above, point to a set of new worth thinking trends at a technological and educational level. E-learning has been consolidating itself and attracting new interest in multiple knowledge areas, due to the flexibility, accessibility and autonomy it confers to the learning process and in all it dimensions. However, emerging technologies place us with new challenges and provide us with new

opportunities to explore solutions, so far unavailable. Therefore, the interest of this study is based on the opportunity to answer the needs and expectations of organizations. So, it is important to analyze the perspective of value that these systems (based on semantic technologies and social interaction) can offer to organizations. By allowing the aggregation of shared knowledge, these systems will function as personalized learning environments, providing content in specific areas of interest aligned with the organizational goals. It is expected to obtain an innovative system that provides conditions for the creation of interactive, participative and shared dynamics, ensuring reliability available materials.

Acknowledgments. This work is funded by FEDER through Programa Operacional Fatores de Competitividade – COMPETE, and by national founds through FCT – Fundação para a Ciência e Tecnologia, under project: FCOMP-01-0124-FEDER-022674. This work was supported by the FCT under the grant SFRH/BD/71792/2010.

References

[1] dos Santos, F.L.: Estratégia Passo-A-Passo, 1st edn. Centro Atlantico, Lisboa (2004)
[2] Senge, P.M.: The Fifth Discipline: The Art and Practice of the Learning Organization: Book Review (1993)
[3] UNDP, "Human Development Report 2010, The Real Wealth of Nations: Pathways to Human Development", United Nations Development Programme, Online, Global Report 2010 Report (2010)
[4] Klamma, R., Chatti, M., Duval, E., Hummel, H., Hvannberg, E., Kravcik, M., Law, E., Naeve, A., Scott, P.: Social Software for Life-long Learning (2007)
[5] IAIED | International AIED Society, http://iaied.org/about/ (accessed: April 10, 2012)
[6] Rico, D.F., Sayani, H.H., Jeffrey, V., Sone, S.: The Business Value of Agile Software Methods: Maximizing ROI with Just-in-Time Processes and Documentation. J. Ross Publishing, U.S.A. (2009)
[7] Torniai, C., Jovanovic, J., Gasevic, D., Bateman, S., Hatala, M.: E-Learning Meets the Social Semantic Web. In: Eighth IEEE International Conference on Advanced Learning Technologies, ICALT 2008, pp. 389–393 (2008)
[8] Jovanovic, J., Devedžic, V., Gasevic, D., Hatala, M., Eap, T., Richards, G., Brooks, C.: Using Semantic Web Technologies to Analyze Learning Content. IEEE Internet Computing 11(5), 45–53 (2007)
[9] Gruber, T.: Collective Knowledge Systems: Where the Social Web meets the Semantic Web. Web Semantics: Science, Services and Agents on the World Wide Web 6(1), 4–13 (2008)
[10] Berners-Lee, T., Hendler, J., Lassila, O., et al.: The Semantic Web. Scientific American 284(5), 34–43 (2001)
[11] Jovanovic, J., Gasevic, D., Devedzic, V.: E-Learning and the Social Semantic Web. In: Dicheva, D., Mizogouchi, R., Greer, J. (eds.) The Future of Learning, Semantic Web Technologies for e-Learning, Amsterdam, Netherlands, vol. 4, p. 299 (2009)
[12] W3C.org, OWL Web Ontology Language Overview (February 10, 2004), http://www.w3.org/TR/2004/REC-owl-features-20040210/#s1.3 (accessed: Maio 28, 2011)

[13] Rajiv, P., Sanjay, D.: Ontology Description using OWL to Support Semantic Web Applications. International Journal of Computer Applications 14(4), 30–33 (2011)
[14] Isotani, S., Mizoguchi, R., Bittencourt, I., Costa, E.: Estado da Arte em Web Semântica e Web 2.0: Potencialidades e Tendências da Nova Geração de Ambientes de Ensino na Internet. Revista Brasileira de Informática na Educação 17(01), 30 (2009)
[15] W3C.org, W3C Semantic Web, http://www.w3.org/2001/SW (accessed: Maio 27, 2011)
[16] Hendler, J., Berners-Lee, T.: From the Semantic Web to Social Machines: A Research Challenge for AI on the World Wide Web. Artificial Intelligence 174(2), 156–161 (2010)
[17] O'Reilly, T.: What Is Web 2.0. O'Reilly (September 30, 2005), http://oreilly.com/web2/archive/what-is-web-20.html (accessed: April 30, 2011)
[18] Alexa.com, Alexa, Top Sites by Category: Top, http://www.alexa.com/topsites/category?category=Top&q=social+networks&r=topsites_category&p=bigtop (accessed: March 10, 2012)
[19] Siemens, G., Weller, M.: Higher Education and the Promises and Perils of Social Network. Revista de Universidad y Sociedad del Conocimiento (RUSC) 8(1), 164–170 (2011)
[20] Siemens, G., Tittenberger, P.: Handbook of Emerging Technologies for Learning. University of Manitoba, Manitoba (2009)
[21] Lopp, M.: Managing Humans: Biting and Humorous Tales of a Software Engineering Manager. Springer (2007)
[22] Siemens, G.: Connectivism: A Learning Theory for the Digital Age (2005), http://www.itdl.org/journal/jan_05/article01.htm (accessed: June 23, 2012)
[23] Mota, J.: Da Web 2.0 ao e-Learning 2.0: Aprender na Rede, Tese de Mestrado. Universidade Aberta, Portugal (2009)
[24] Mendes, I., Santos, H.: Sistemas Auto-Regulados de Conteúdos Orientados para Contextos de Aprendizagem em Ambiente Profissional. In: 6th Congresso Luso-Moçambicano de Engenharia/ 3rd Congresso de Engenharia de Moçambique, Maputo, Moçambique, p. 1188 (2011)
[25] Hippel, E.: Democratizing Innovation: the Evolving Phenomenon of User Innovation. International Journal of Innovation Science 1(1), 29–40 (2009)
[26] Hippel, E.: Horizontal Innovation Networks—By and for Users. Oxford Journals 16(2), 293–315 (2007)
[27] Miller, A.: Disruptive and Transformative Education: Designing Learning in the Digital Age. eLearn Magazine, Education and Technology in Perspective 2012(11), 1 (2012)

How Small and Medium Enterprises Are Using Social Networks? Evidence from the Algarve Region

Ana Belo[1], Guilherme Castela[2], and Silvia Fernandes[2]

[1] SandoGago Inc., Portugal
anagagobelo@hotmail.com
[2] Research Centre for Spatial and Organizational Dynamics (Cieo)
Faculty of Economics, University of Algarve, 8005-139 Faro, Portugal
{gcastela,sfernan}@ualg.pt

Abstract. The evolution of internet created new opportunities for small and medium enterprises (SME), among which are social networks. This work aims at analyzing the potential of these networks for the SME in Algarve, creating a questionnaire for the purpose. The empirical study revealed that some firms have already an integrated business strategy with social networks, as well as a group in the firm responsible for it. Most of their managers consider that social networks enhance performance, but few really measure these results. A categorical principal component analysis identified two dimensions of social networks' use: social networks for product-client interaction and knowledge; and social networks with potential for marketing. A supplementary analysis (hierarchical clustering) identified three patterns of SME's involvement in social networks: cluster Social Net Level 1, cluster Social Net Level 2 and cluster Social Net Level 3. These groups validated the results described above, indicating a sustainable methodological approach.

Keywords: social networks, SME, patterns, performance, alignment.

1 Introduction

Modern drivers of change have led to new features for sustaining business competitiveness: virtual enterprises, creative design, social networks, mobile platforms, etc. These responses require IT-based tools, the increase of information content and creative teams. Any enterprise that wants to optimize its success in the information society must have a basic awareness and a strategy for dealing with this new environment. With the shortening of life cycle and time to market this has to evolve with a pace that approaches a need for its real time creation [27].

For example, the internet brings critical new functionalities to enterprises by real time business dynamics and relations. If they don't constantly improve the process of managing this, they would quickly become paralyzed. Building custom software systems to accomplish it can become so complex that the software requires an 'army' of engineers and consultants to manage it. Small to medium businesses need a way to organize their data on a smaller scale and without the expense of hiring costly

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 143–155.
DOI: 10.1007/978-3-642-36981-0_14

consultants or paying for maintenance contracts. The approach of building and managing a custom software solution is financially out of reach for small to medium sized companies. Therefore companies are moving to web-based software as it can eliminate the need to purchase and manage computer servers, firewalls and software servers just to run the application.

New communication technologies allow a global interaction like never imagined. Internet evolution, and especially the Web 2.0 [25] opened new opportunities and benefits for SME, given its ease of communication and information dissemination [5]. One of the greatest opportunities was the opening of new online applications of network environments known as social networks [41], [4], [8]. Today the internet presents itself as the platform of greater access, in which millions of individuals daily enter at any place or time [40]. In this context, new environments appeared [11] such as the social networking sites like Facebook, YouTube, LinkedIn, Twitter, Hi5, Bebo, MySpace, among others, in which users either communicate or share content [26], [4]. On these platforms people create their profiles, communicate, exchange pictures and data, group on a particular interest creating communities. These individuals, networked, can actively participate in innovation, wealth creation and social-economic development in a way never imagined before [28].

According to the study "Internet use in Portugal 2010" [39] more than 60% of social networks users in Portugal consider important that companies have a profile on these networks. The continuous entry of firms in these applications can completely change the way of doing business. Some authors have suggested that after the knowledge economy and digital economy, a new economy happens now, naming it "Socialnomics" [28] or "Economy of relations" [30], or "Economy of integrity" [3]. The key features of business and innovation, which in past decades were tangible, are now replaced by intangible assets such as knowledge, confidence and relations. Companies need to adapt to this revolution in order to be competitive.

2 The Alignment Issue

In Portugal, the enterprises (most of them small or medium sized) invest little in information system's departments due to their limited capacity (financial and organizational). A form of innovation in which they could invest consists of building an entrepreneurial culture supported by a good information structure. There are systems and architectures that can easily contribute to it, based on tools like CRM (Customer Relationship Management), CMS (Content Management Systems), ERP (Enterprise Resource Planning) and SNA (Social Networks Analysis). Integrated into well planned platforms, they can accelerate the implementation of new ideas, products/services and business processes (or improvement of existent ones). Given their enormous potential and online scale, which in some cases can lead to a reconfiguration of the whole organization or business model, entrepreneurs should not only know well this kind of platforms but also be involved at the very beginning of their strategic planning [19].

Managers take most of their decisions based on a considerable amount of data to know what products they should offer, in what quantity, from which supplier, in which way of distribution, to which location, etc. Today, as internet plays an increasing role in doing business, firms are using more analytical tools through data warehouse management and social networks. These tools can be specially used in systems designed for data and process integration. ERP systems do this work, however there is some resistance to adopt them related with time of adaptation and change of the existent systems. This can be long and firms generally avoid interrupting their normal activities to start those changes [43]. Thus many are migrating to cloud computing for benefiting from the ease of its online integration. This technology is helping firms virtually manipulate all aspects of the business chain.

Differently from departmentalized solutions, those clouds are transversal as they cross several levels, activities and agents. They make information flow in a simpler way through multiple layers to be extensively shareable. This allows managers to easily get more, diverse and accurate data without needing physical servers. Operations are well coordinated, providing a global view of processes and information flows [21]. In departmentalized systems, on the contrary, there is much data fragmentation which brings the proliferation of complex and expensive relations, as information systems are separated by areas. Today's challenges require consolidating all the data and eliminating complex relations, with a positive impact on process efficiency and business performance.

Integrating CRM function with social networks, consisting of analytical procedures for customer relationship management, makes possible to consolidate information from different sources and communication channels. Firms can use those data to know about their image, improve their products and services, acquire new clients, personalize their offers, etc. CRM techniques allow selecting key information by different points of view, helping to create unique services and critical innovations. However there is the problem, related with people alignment, that employees using these tools are not generally familiar with the criteria inherent to CRM and social networks analysis, which require specific training [43]. The enterprises should consider the adoption and diffusion of new tools from a strategic point of view and evaluate them as any other investment. For instance, many firms don't know how to consolidate several databases to extract information into management decision reports. Instead, they depend on external outsourcing or consulting services which generally increase their costs. To know all the costs and benefits, including the indirect (hidden) ones, the enterprise should define a business plan and, aligning it with the information system, evaluate regularly its results within that plan.

A comparison among small and medium firms in different sectors of the Algarve region (especially in commerce and services), revealed some issues to be considered on aligning new information technologies with business performance. Most of the enterprises interviewed on their threats and opportunities (Swot Analysis) don't align their vision of business with the potential of new information technologies and people competences [14]. In other words, analyzing how opportunities can solve weaknesses (or which threats could become opportunities) they do not fully explore how information technologies could reevaluate these issues. Main factors disturbing this

alignment are the lack of an enterprise-wide business view and the rapid rate of technological change. Other authors have also acknowledged these aspects in their studies, such as [29].

Nevertheless, some of the firms interviewed actively align their business vision with the potential of information technologies, namely those from technological sectors and commerce chains, which perceive the alignment issue enhancing their performance in terms of process integration, management efficiency, time and completeness of response to customers. Two main reasons are related with this behavior: the nature of their business processes and the right capacity for it. This mainly happens with firms in technological sectors due to their specialized know-how, and in commerce chains due to their organizational capacity [14]. The presence of technology professionals and their proximity to managers is fundamental, requiring a profile that catches the emergent trends adjusting them to the right pace [33].

Another issue assessed by this case comparison was that the sectors with more mobile activities (such as INEM medical-emergency, CTT post-offices and ALGAR eco-points), as well as the banking sector which increasingly depend on knowledge about financial markets in real time, recognize the growing potential of online and mobile platforms [13]. The ease of mobility achieved, apart from firm's location, allow accessing updated information between its intranet and the mobile gadget in use. Having this aspect in mind, besides the additional potential of online integration, the alignment issue is either a key for business performance, or a key for enterprise knowledge-base. Figure 1 illustrates this point through the main lines of information systems' development, and supporting technologies, with the resulting knowledge-base enhancement:

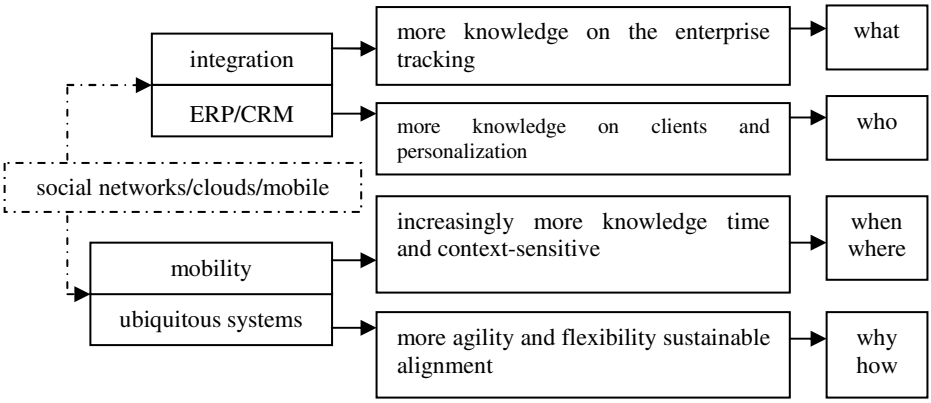


Fig. 1. Main lines of information systems' development and supporting technologies, and resulting advances in enterprise knowledge-base. Source: own elaboration.

The previous statement is not surprising as information analysts and CIO (Chief Information Officers) strongly recognize the increasing support of these architectures to process alignment and knowledge-base improvement, for more real-time decisions and activities [44], [36]. A knowledge structure can then be derived to support

matching between business needs and innovations [32]. According to [16], knowledge management is the critical layer to incorporate in enterprise management. According to [37], with the growth of global service economy leading to a huge increase in daily interactions with specialized online services, approaches should combine knowledge management with service architectures aimed at capturing the advances in emerging systems. The following empirical study opens a door for the potential of social networks for SME's performance and knowledge-base.

3 An Empirical Study

Studies on social networks have expanded, receiving increased attention from the scientific community [4]. These platforms are currently a major research focus in several areas. One example is the Facebook application, which has been studied by [10], [1], [20], [38], etc. It is essential to explore the potential of social networks in enhancing the performance of SME as these are generally the most affected by strong global competitiveness. This work aims at characterizing the representation of Portuguese SME, especially of the Algarve region, in these networks and analyzing the activities most potentiated. Some firms use these networks for marketing activities however they should not view social networks as just an online marketing tool [11]. They need to look deeper and analyze these new environments with a multiple perspective as social networks allow communication that covers millions of different features and potential customers [42], [40], [5], [8].

The main objective of this study is to analyze the potential of social networks as tools for SME's development and then indicate a methodological approach for general application. Other works, from [2], [7] and [34], also have analyzed factors and gains related to the use of social networks by SME. It then intends to verify which activities of SME in the Algarve region are already being enhanced by these platforms. For this purpose, an online questionnaire was activated and answered by 70 firms from the region. From the many resulting data, an analysis of the corresponding variables served to identify the most relevant ones and derive some patterns (clusters) according to the level of firms' involvement in social networks.

3.1 Sample and Methodology

The methodological path began with an observation of facts related with the subject, through a comparison of social networking sites and the representation/participation of SME in these, as well as how companies benefit from such involvement. For example, comparing several available videos, presentations and discussion groups on social networks. Then, we analyzed several other empirical studies already carried out, which are mentioned in the previous section. This was followed by a selection of techniques for gathering information for data analysis. First, a questionnaire was prepared to collect the data from a sample of SME in the Algarve region. The data were then analyzed through quantitative techniques which include preliminary descriptive statistics and behavioral patterns. Specifically, we used the Categorical

Principal Component Analysis (CATPCA) to detect patterns of association between variables [15], [9], [22], [24]. Then, we used the hierarchical cluster analysis based on the Ward method [31] as the criterion to detect patterns of association between firms, considering the variables highlighted by the CATPCA analysis. There are other authors that also combined these two methods of analysis such as [35].

Generally, the questionnaire directed to the SME of Algarve (QSME) aims to characterize the types of SME according to their participation in social networking sites. The questionnaire was built using a specialized online tool - SurveyMonkey (www.surveymonkey.com) - which allows design a website where the questionnaire is available. The use of this tool in research allows quick access to questionnaires, facilitating faster responses. An e-mail with the study intentions and a link to the questionnaire was sent to the SME. The data remain in a database, grouped by response [23]. As a sample, we considered 70 SME of the Algarve region having completed the questionnaire and answered it on time. Expecting for more cases would take much time as companies often don't have time to respond to free questionnaires. After closing the online questionnaire, the data were collected to be analysed. The responses directly given in the SurveyMonkey were exported to Excel, and then to SPSS v.17 for more advanced analyses.

4 Analysis of Patterns

The purpose of this analysis is to reach to a diagnosis of most relevant variables on the participation of SME in social networks. Thus, we used a Categorical Principal Component Analysis [15], [9], [22], [24] as an exploratory technique of interdependence and reduction of data dimensions in order to detect patterns of association among variables. This analysis allowed to obtain rather consistent results as described in Table 1.

Table 1. Model summary. Source: SPSS 17.0.

Dimension	Cronbach's alpha	Variance obtained
1	,899	8,353
2	,774	4,139
Total	,940	12,492

According to the Cronbach's alpha coefficient, indicator of the degree of internal consistency of observations, our sample is highly consistent in both dimensions (with a total of 94%). The related literature informs that this occurs when the coefficient exceeds 70% [15], [9], [22], [24]. We also can observe from Table 1 that the explained variance is 66.86% in the 1st dimension and 33.13% in the 2nd dimension.

The CATPCA also revealed the weights of the variables. Figure 2 illustrates the bidimensional structure of variables' distribution on the reference space, in which the vectors project the variable loadings.

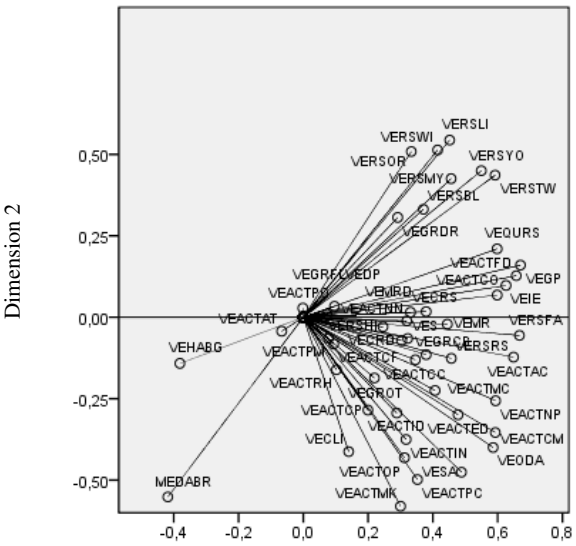


Fig. 2. Bidimensional distribution of variables. Source: SPSS 17.0.

From this analysis, all the variables had positive weights in the first dimension while the second dimension had a strong contrast of negative and positive weights. This means there are weak relations between the variables in second dimension, turning it more fragile. The analysis of the most relevant variables (12 relevant in the first dimension and 4 relevant in the second dimension) could derive the two following types or dimensions of social networks (Table 2).

Table 2. Dimensions of relevant variables (types of social networks). Source: own elaboration from the CATPCA analysis using SPSS 17.0.

Dimension/type	Characteristics
First dim.: social networks for knowledge and for product-client interaction	*Twitter, Youtube* - more oriented to (re)search, communication and knowledge (professional-oriented); and *Facebook* - more oriented to the interaction with customers (new products to customers, opinions: social-oriented)
Second dim.: social networks with potential for marketing	*Orkut, Wikis, LinkedIn* - more oriented to marketing and divulgation

Based on Table 2, the first dimension retains much more information as it captures the largest percentage of explained variance (Table 1). The second dimension retains less information and captures the residual variance [15], [31]. One aim of the analysis is to reduce the dimensionality of data with the least possible loss of information. Although extracting dimensions some information is lost, the use of all dimensions would complicate the analysis. The main goal is to extract patterns, detected from the explained variance while the remaining observations turn into residual. Thus, the 1st

dimension (social networks for knowledge and product-client interaction) is well characterized in Table 2, considering that all its relevant variables have positive weights. The 2nd dimension, due to its most relevant variable being negative (VEACTMK) and related with marketing activities, we named it as social networks with "potential" for marketing (Table 2). Another reason for this classification (mainly the term potential) is the weakness of relations between the variables in the 2nd dimension, as mentioned before.

Experts in e-business and social media emphasize that there is not sufficient empirical work in Portugal to support the dimensions and objectives described in Table 3 [12]. There are discussion groups, such as the "Social network/media strategy" and "Social media Portugal" in LinkedIn, where these aspects are discussed. However, they are limited to opinions from who usually work in these areas, but without sustainable empirical support.

Table 3. Clusters of firms (behavioral patterns). Source: own elaboration using SPSS 17.0.

Clusters obtained	Characteristics
Cluster 1 - "Social Net Level 1" - formed by 37 SME, which mostly:	- lack participation in Wikis, Orkut, Facebook, Twitter, LinkedIn, Youtube; - don't use social networks for enhancing their business performance; - don't implement an integrated strategy with social networks; - have a positive opinion on social networks' potential for business performance; - have a weak social networks' activity regarding marketing, new products, fidelization and behavioral patterns analysis; - have a significant social networks' activity regarding communication with customers; - don't have a group responsible for the representation/participation in social networks; - in average, they use social networks once a weak.
Cluster 2 - "Social Net Level 2" - formed by 21 SME, which mostly:	- have a significant presence on Facebook; - usually use social networks for enhancing their business performance; - have a strong social networks' activity regarding new products and communication with customers; - have a strong dissemination practice through social networks and a widespread positive opinion on social networks' potential for business performance; - in average, they use social networks every day.
Cluster 3 - "Social Net Level 3" - formed by 12 SME, which mostly:	- lack participation in MySpace and Wikis; - have a significant presence on Twitter, LinkedIn, and Youtube; - generally implement an integrated strategy with social networks; - have a strong social networks' activity regarding behavioral analysis; - have a group responsible for firm's representation/participation in social networks; - in average, they use social networks several times a day.

Then, a hierarchical cluster analysis was conducted, based on the Ward clustering method. It detected three different clusters [31]. The three clusters correspond to firms' behavioral patterns according to their involvement with social networks, expressed in the related variables. Considering the relevant variables, and the statistical significance given by CATPCA analysis, the clusters were classified according to the level of SME's involvement in social networks (Table 3).

Thus, the clusters obtained evolve in an ascending order of business involvement with social networks as shown by the cluster analysis in Table 3.

5 Discussion and Implications

The involvement of SME with social networks in Algarve is not uniform. This research identified three different clusters (groups) named as: 'Cluster Social Net Level 1', 'Cluster Social Net Level 2' and 'Cluster Social Net Level 3'. This classification is due to a different pattern in terms of degree of involvement in social networks. The firms in the first cluster present the lowest involvement with these platforms. Nevertheless, a company in cluster 1 is not necessarily lagging behind a company of cluster 2 as the different degree means a different pattern of involvement (not a pattern of performance). For example, both clusters 2 and 3 have companies with significant presence in social networks; however firms in cluster 3 already measure the potential gains, having an integrated strategy with these tools and using them several times a day.

We also found two dimensions of social networks' use: a first dimension oriented either to product-client interaction (Facebook - social oriented) or to knowledge/ research (Twitter, YouTube - professional oriented); and a second dimension with potential for marketing/divulgation (LinkedIn, Wikis, Orkut). So, the first dimension includes two main uses: social and professional. For example, it is enormous the influence that an opinion about a firm's service, scattered over a social network like Facebook, can have on that firm's image (contagion/imperceptible effects).

On the other hand, the second dimension showed weak associations between relevant variables. Thus, our sample is less likely to actively use social networks for marketing strategic integration, despite having propensity for it. For example, the online CRM application (Customer Relationship Management) seems to be on a starting stage for the majority of SME in Algarve[1]. This issue requires a more pro-active 'social media' strategy. The steps towards it should include: tracking the links along nodes (connection points) through social network analysis; exploring the resulting experiences/communities; and aligning with all critical partners in the necessary development areas [1].

[1] Especially SME in the technological sector, like Algardata and Visualforma, and some hotels assisted by them, are using the online cloud CRM 'Sales Force' (`www.salesforce.com`)

6 Conclusions

Most companies have just been overwhelmed with the flood of data created by networking technologies/communities. This includes click-stream data, detailed email data and also the information thrown off by the enterprise systems' transition [17]. It is interesting that by individually studying this information carefully, firms can have better knowledge of customers and products' trends, quality and defects. The field of business intelligence has been tapping into this explosion of data which is important in dealing with the alignment challenge.

Facing the increasing number of online applications with high potential for integration and mobility, the SME of Algarve already recognize it given their positioning in social networks and business clouds. However, they still lack a well defined strategy, aligned and internationalized. The relevance of aligning their information systems with people and activities continues to intensify and is normally the core of discussion on performance, especially related with the support provided by new technologies. Facing the increasing geographical scope and time-sensitiveness of services, that support is making the difference. For instance, when social networks penetrate all activities and institutions, many aspects of management and organizational structure will change radically. The speed and extent of connectivity they allow enables the creation of new business processes. Bearing in mind those challenges, a company should be able to correctly evaluate benefits and costs among which many are intangible. It is fundamental to define a plan in due course and assess, on a regular basis, the results obtained from the information systems strategically aligned with external challenges and internal competences.

The data collected for this work showed that the most performed activities through social networks are: communicating with customers, being closer to potential customers and marketing. Nevertheless, SME are also using social networks for activities such as analyzing their competitors and behavioral patterns, knowing trends, cooperation, fidelization, internationalization, new products launching, new brands/businesses, opinion searching, new ideas, and recruitment. Most of managers consider that social networks enhance the performance of their firms, but few really measure the results from this impact. Firms that most frequently accede to social networks are those whose manager has higher qualification level. And the SME that most use these platforms in Algarve are from the services and hotel/restaurant/bar sectors.

Social networks should be seen as work facilitators, and entrepreneurs need to better understand how to take advantage of their use and provide results for companies. Some advantages to be emphasized are: 'constant' (24 hours a day); 'global' (internet as support) and 'in real time' (direct/immediate). These are critical attributes for creating flexible and distinctive outcomes. However, some studies [18], [6] strongly indicate that the most important issue is not being present in large social networks (or in several of them), but to define a 'social media' strategy. Entrepreneurs must be familiar with the existing types of social networks, their characteristics and potential, in order to determine in which to be present and take full advantage. Then, they should start by taking the steps discussed in the previous section.

References

1. Acquisti, A., Gross, R.: Imagined communities: Awareness, information sharing, and privacy on the Facebook (2006), http://www.heinz.cmu.edu/~acquisti/papers/acquisti-gross-facebook-privacy-PET-final.pdf (accessed: February 03, 2011)
2. Amo, M., Constantinides, E., Romero, C.: Profiles of social networking sites users in the Netherlands. University of Twente Publications (2010) http://doc.utwente.nl/73399/ (accessed: March 10, 2011)
3. Bernasek, A.: The economics of integrity. Editora Harperstudio (2010)
4. Boyd, D., Ellison, N.: Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication 13(1), article 11 (2007), http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html (accessed: February 20, 2011)
5. Brandão, R., Marques, B.: The impact of web 2.0 in the informal learning of the Portuguese SME. AIP-CE, CEPCEP, Universidade Católica Portuguesa (2010)
6. Brandão, M.: O que são redes sociais verticais e horizontais (2011), http://www.maiswebmarketing.com/o-que-sao-redes-sociais-verticais-e-horizontais/ (accessed: February 12, 2011)
7. Carr, R., Castleman, T., Mason, C., Parker, C.: Factors affecting SME's willingness to share knowledge on-line: A path model. In: Proceedings of the 23rd Bled eConference eTrust on Implications for the Individual, Enterprises and Society. University of Maribor, Bled (2010)
8. Constantinides, E., Lorenzo, C., Gómes-Borja, M.: Social media: A new frontier for retailers? European Retail Research 22, 1–27 (2008)
9. De Leeuw, J.: Multivariate analysis with optimal scaling. In: Gupta, S., Sethuraman, J. (eds.) Progress in Multivariate Analysis. Indian Statistical Institute, Calcutta (1990)
10. Dwyer, C., Hiltz, C., Passerini, K.: Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. In: America's Conference on Information Systems (AMCIS), USA (2007)
11. Evans, D.: Marketing de media social, uma hora por dia. Editora AltaBooks (2009)
12. Falcão, H.: Social technology (2010), http://pt.scribd.com/collections/2545330/Social-Technology-Tecnologia-Social (accessed: February 15, 2011)
13. Fernandes, S.: Main differentiating issues in aligning information systems with business performance. International Journal of Information Systems in the Service Sector (forthcoming)
14. Fernandes, S.: Critical information as a success factor in organizations: objective and methodological approach. CIEO Spatial and Organizational Dynamics Discussion Papers 2, 69–82 (2010)
15. Gifi, A.: Nonlinear multivariate analysis. J. Wiley & Sons, New York (1990)
16. Gudas, S.: Enterprise knowledge management aspects and layers. Paper presented at the 12th World Multi-Conference on Systemics, Cybernetics and Informatics, Florida, USA (2008)

17. Hopkins, M.: The 4 ways IT is driving innovation. An interview with Erik Brynjolfsson (2010), http://sloanreview.mit.edu/the-magazine/articles/2010/spring/51330/it-innovation-brynjolfsson-article (accessed: April 24, 2010)
18. Inesting, Social media marketing (2011), http://www.inesting-search.com/SMM/social-media-estrategia.html (accessed: February 12, 2011)
19. Kearns, G., Sabherwal, R.: Strategic alignment between business and information technology: A knowledge-based view of behaviours, outcome and consequences. Journal of Management Information Systems 23(3), 129–162 (2006)
20. Lampe, C., Ellison, N., Steinfeld, C.: Profile elements as signals in an on-line social network. In: Conference on Human Factors in Computing Systems, San Jose, CA, USA (2007)
21. Laudon, C., Laudon, J.: Management information systems: Managing the digital firm. Prentice Hall, New Jersey (2004)
22. Meulman, J.: The integration of multidimensional scaling and multivariate analysis with optimal transformations of the variables. Psychometrika 57, 539–565 (1992)
23. Minayo, M., Suely, F., Gomes, R.: Pesquisa social, teoria, método e criatividade. Petrópolis, Brasil (2007)
24. Nishisato, S.: Elements of dual scaling: An introduction to practical data analysis. Lawrence Erlbaum Associates, Inc., New Jersey (1994)
25. O'Reilly, T.: What is web 2.0: Design patterns and business models for the next generation of software (2005), http://www.elisanet.fi/aariset/Multimedia/Web2.0/What%20Is%20Web%202.doc (accessed: April 09, 2011)
26. Pei, M., Ramayah, T., Suki, N.: Factors enhancing employed job seekers intentions to use social networking sites as a job search tool. International Journal of Technology and Human Interaction 7(2), 38–54 (2011)
27. Philipson, S.: A specification of an environment for modern product development. Paper presented at the International Conference on Advances in Management, Boston, USA (2008)
28. Qualman, E.: Socialnomics - Como os media sociais estão a mudar o mundo como vivemos e como fazemos negócios. Editorial Presença (2009)
29. Rathnam, R.: Alignment of business strategy and IT strategy: A case study of a fortune 500 financial services company. The Journal of Computer Information Systems 45(2), 1–8 (2004)
30. Robison, L., Ritchie, B.: Relationship economics: The social capital paradigm and its application to business, politics and other transactions. Editora Gower (2010)
31. Romesburg, H.: Cluster analysis for researchers. Lifetime Learning Publications, USA (1984)
32. Rosselet, U., Jolliet, Y., Wentland, M.: Knowledge management for IT project portfolio. Paper presented at the 10th European Conference on Knowledge Management, Vicenza, Italy (2009)
33. Setti, A., Fernandes, S.: The role of the information system's manager in the organization. Interscienceplace 1(3), 1–10 (2007)
34. Silver, D.: The Social network business plan: 18 strategies that will create great wealth. John Wiley & Sons, New Jersey (2009)

35. Sousa, L., Galante, H., Figueiredo, D.: Qualidade de vida e bem-estar dos idosos: um estudo exploratório na população portuguesa. Revista Saúde Pública 37(3), 364–371 (2003)
36. Spewak, S., Hill, S.: Enterprise architecture planning: developing a blueprint for data, applications and technology. John Wiley & Sons Inc., USA (1992)
37. Spohrer, J., Kwan, S.: Service science, management, engineering, and design (SSMED): an emerging discipline - outline & references. International Journal of Information Systems in the Service Sector 1(3), 1–31 (2009)
38. Stutzman, F.: Student life on the facebook (2006), http://ibiblio.org/fred/facebook/stutzman_fbook.pdf (accessed: February 20, 2011)
39. Taborda, M.: A utilização da internet em Portugal 2010. LINI - Lisbon Internet and Networks International Research Programme, Lisboa (2010)
40. Tapscott, D., Williams, A.: Wikinomics, a nova economia de multidões inteligentes. Editora Quidnovi (2007)
41. Tredinnick, L.: Web 2.0 and business: A pointer to the intranets of the future. Business Information Review 23(4), 228–234 (2006)
42. Vasconcelos, V., Campos, P.: Distributed Informal Information Systems for Innovation: An Empirical Study of the Role of Social Networks. In: Quintela Varajão, J.E., Cruz-Cunha, M.M., Putnik, G.D., Trigo, A. (eds.) CENTERIS 2010, Part II. CCIS, vol. 110, pp. 208–217. Springer, Heidelberg (2010)
43. Vasilev, J., Georgiev, G.: Tendencies in the development of ERP systems. Paper presented at the International Conference on Computer Systems and Technologies, Varna, Bulgaria (2003)
44. Zachman, J.: A framework for information systems architecture. IBM Systems Journal 26(3), 276–292 (1987)

Temporal Visualization of a Multidimensional Network of News Clips

Filipe Gomes, José Devezas, and Álvaro Figueira

CRACS/INESC TEC, Faculdade de Ciências, Universidade do Porto
Rua do Campo Alegre, 1021/1055, 4169-007 Porto, Portugal
{filipe.gomes,jld,arf}@dcc.fc.up.pt

Abstract. The exploration of large networks carries inherent challenges in the visualization of a great amount of data. We built an interactive visualization system for the purpose of exploring a large multidimensional network of news clips over time. These are clips gathered by users from web news sources and references to people or places are extracted from. In this paper, we present the system's capabilities and user interface and discuss its advantages in terms of the browsing and extraction of knowledge from the data. These capabilities include a textual search and associated event detection, and temporal navigation allowing the user to seek a certain date and timespan.

Keywords: visualization, networks, search, event detection.

1 Introduction

The exploration of large networks, using force-directed graphs [9], carries inherent challenges in the visualization of a great amount of data. One problem arises, firstly, from the large number of nodes present in the graph and, secondly, from the potentially exponential growing number of links, that clutter the screen. Associated to this problem is the smooth navigation of the graph, which is addressed with an intuitive user interface and a reduced visual disturbance in the graph's deployment.

In this paper, we describe a visualization and browsing system for a network of news clips with three dimensions for the edges — *who*, *where* and *when* — and temporal annotations for each node. This network is based on a project with the goal of connecting online news and the Social Web. The network is based on the relations induced by the coreference of entities between a compilation of news clips taken from online news sources.

2 Reference Work

Previous studies have reported that visualization has increasingly been used in data mining. While data generation has been growing exponentially, the exploration of large data sets has become a prominent but challenging problem, as

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 157–166.
DOI: 10.1007/978-3-642-36981-0_15

visual data mining is an interesting instrument handle this data deluge [11]. A visual data mining system must be syntactically simple to be useful, that is, it must have intuitive user interaction, a clear output and, thus, should be considered a powerful tool [14].

In recent years, there has been a considerable amount of literature published on the topic of large network visualization. Previous works have been undertaken, presenting visual analytics tools [13], having been identified as breakthroughs in data exploration in the context of large data sets [10]. The development of these tools analysed significant aspects of visualization tools, such as the usability or scalability [1]. Moreover, the exploration through time of a network carries its own challenges: new data should be presented in context of older data and the visualization must be dynamic for the user to realize the changes [12].

3 Data Set

Our system resorts to a collection of news clips that, for the purpose of this paper, were fetched automatically from select news outlets' websites. We developed a crawler that is scheduled to run twice daily, to gather the top news from successive days. The goal is to artificially generate, from the crawled news articles, a news clips collection similar to a user-generated one. To do that, we segmented each news article into sentences and automatically generated between two and five clips, with two to seven sentences each, by randomly selecting an initial starting sentence as well as the referred parameters. The gathered data set contains 8,981 clips, fetched between the 8^{th} and 12^{th} November 2012 for ten different news sources: Euronews, United Press International, BBC News, Daily Mail, Guardian, Reuters, The Telegraph, USA Today, New York Times, and Washington Post. News clips size varies between 1 and 376 words, with an average of 102 words. Using the methodology described in Devezas and Figueira [6], we extracted the named entities from this data set, including people, places and dates. Table 1 shows the top referenced people and places in our collection. Dates were not included in the table in order to make topic distinction simpler.

Table 1. Top 20 people and places identified by our system within the data set

Rank	Entity	Frequency	Rank	Entity	Frequency
1	China	252	11	David Cameron	83
2	England	176	12	Scotland	82
3	Syria	148	13	Jersey	78
4	Wales	137	14	Europe	71
5	United States	123	15	Africa	71
6	Afghanistan	111	16	Israel	67
7	Iraq	88	17	France	66
8	Barack Obama	85	18	Mitt Romney	61
9	India	84	19	Turkey	53
10	Russia	84	20	Mexico	50

3.1 Multidimensional Network of News Clips

We built a news clips network based on the coreference of the identified entities across text fragments. For each type of entities (people, places and dates), we established a distinct dimension, using three different edge types: *who*, *where* and *when*. This resulted in a multidimensional network with 8,981 nodes representing news clips and 323,177 edges representing a connector entity referenced in the pair of news clips it linked.

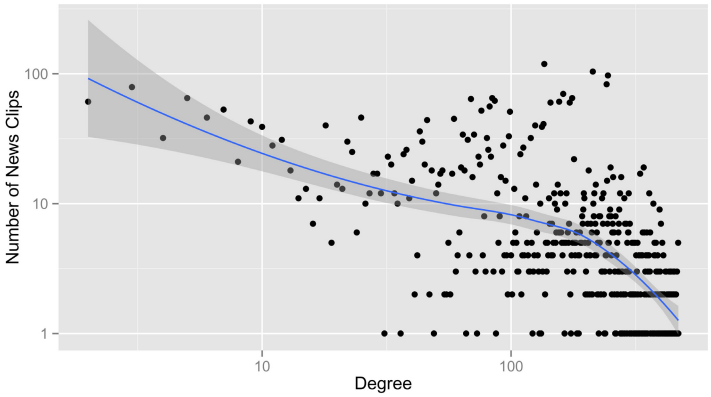


Fig. 1. Degree distribution of the news clips network

A brief analysis of the node degree distribution, depicted in Fig. 1, shows a heavy-tail behaviour, which indicates the presence of a scale-free network. As we can see from Table 2, the multidimensional network of news clips also presents the property of small-world, having a higher clustering coefficient and a lower average shortest path length than its random counterpart, as generated by the Erdős–Rényi model [8] using the same number of nodes and edges.

Table 2. Comparison of the original network with its random counterpart

Network	Clustering Coefficient	Average Path Length
Random	0.008006131	2.549247
Original	0.590759899	2.493203

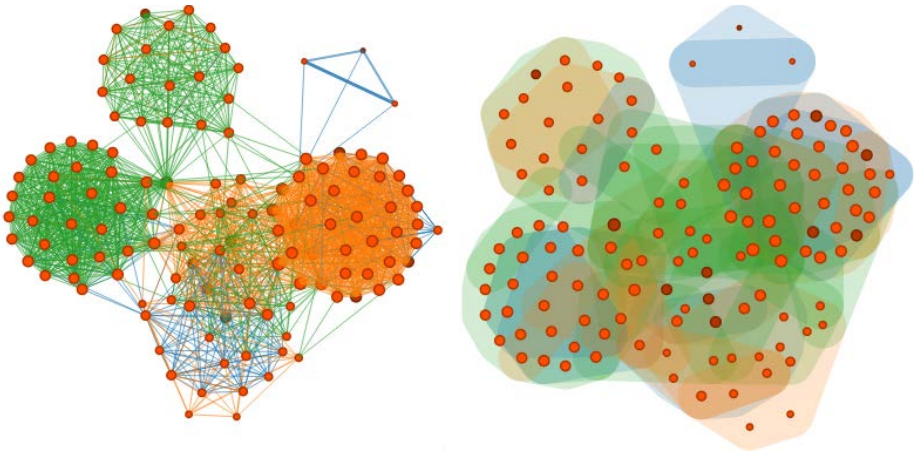
4 Visualization System

In this section, we present the visualization system, firstly introduced in [7], on which our system is built upon. We extended the system by allowing the topical

and temporal search of the news clips' database. Thus, the system indexes news clips, as described in Section 4.2. Moreover, we developed an event detection system that we present in Section 4.3. Finally, we present the system as a whole, with the navigation features that depend on the aforementioned components.

4.1 Multidimensional Network Visualization

We built our multidimensional network visualization system on top of the system presented in [7]. Ergo, it shares the same foundation, being developed using the D3.js JavaScript library [4]. The graph data is transferred using the XML-based GraphML format [5]. Data is transferred asynchronously from the requests and consists only in a relevant portion of the graph, that is shown to the user.



(a) Traditional lines for edges (b) Convex hulls replacing edges

Fig. 2. Comparing the visualization of the network using lines for edges or convex hulls to group nodes that reference a common entity

As shown in Fig. 2, the graph displayed by the visualization system is a force-directed graph. While on Fig. 2a we can see a typical illustration of a network depicted as a graph, on Fig. 2b the nodes are laid out according to their links but those links are not shown to the user.

We chose to hide the links as it allows for less graphical complexity, and to display convex hulls. This is possible for this kind of network, since every node referencing a given entity will be interconnected, thus forming a complete subgraph, that can be contained within a visual group depicted by a convex hull. The hulls' color is determined according to the entity's dimension. Due to the hulls transparency, the number of different relations and nodes involved are easily perceived. Albeit not flawless with great number of relations, the alternative is even less clear with the sheer concentration of edges.

Table 3. Comparison of a graph rendering, with visible edges or hulls

Type	Num. Nodes	Num. Edges	Iterations	Avg. Frame Time	Total Time
Edges	301	14642	298	263.1 ms	78.4 s
Hulls				11.0 ms	3.3 s

The contrast between the two cases is flagrant, particularly in large graphs, as the experimental data shown in Table 3 attests. The average time it takes to compute each frame is nearly 24 times higher, with the edges visible, taking a longer time for the graph to settle down. Moreover, when such large times, like the one obtained with a rather large graph, are experienced, the user will find the application slow and unresponsive.

4.2 Index and Search

We also address the problem of visual clutter, by providing the users with the capability of analysing a selected group of news clips that are relevant to them. To do so, we index the news clips and prepare a search engine that enables generic queries as well as queries within a specified timespan or for a given user.

We take advantage of the Apache Solr open source enterprise search platform [3], built upon the Apache Lucene project [2], to filter, index and search the Breadcrumbs news clips collection. For each news clip, we index the following fields: clip ID, text fragment, title, comment, domain name, owner username, creation date and tags. We also include a field *all*, that is used by default in a query, which results of the concatenation of the text fragment, title, comment and tags of a news clip. This way, the default behavior is to include all of the relevant information in a generic search, broadening the result space (e.g. a news clip with the tag "obama", that doesn't reference Barack Obama in the text, can be returned). In the Solr index and query schema, we define a *text_general* field type, used by all text fields, except the user and the domain names, which tokenizes the text with Lucene's *StandardTokenizer* and filters it in order to remove stop words from a custom bilingual dictionary (English and Portuguese) and to convert the text to lower case. Separate but similar filter groups are defined for the index and query phases, the difference being the addition of a synonym filter during the query phase.

An instance of Apache Solr is deployed to our application server, running alongside the Breadcrumbs application. Communication is established using a custom wrapper service implemented in the Breadcrumbs system, responsible for directly querying Solr, which is only available in the server-side, as a web service.

4.3 Event Detection

At an initial development stage of the visualization system, we identified a usability issue regarding the temporal navigation of the news clips network. Whenever there were no news clips for a selected time interval, a user would be presented with an empty canvas, having to guess a valid interval where some of the collected news clips would be displayed. In order to solve this problem, we developed an event detection subsystem, based on the detection of relevant peaks in a time series depicting the number of results of a given search for each day.

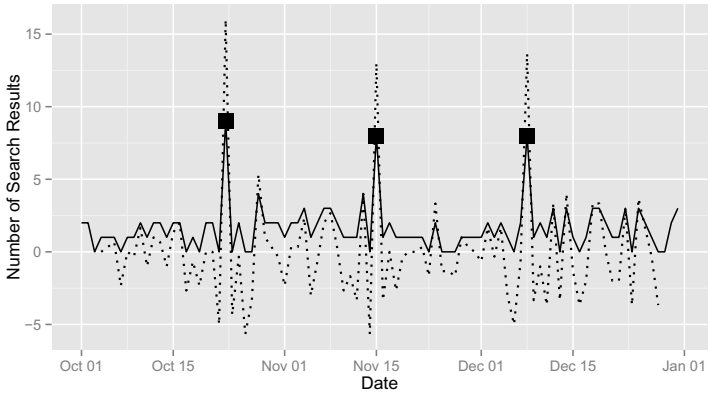


Fig. 3. Example time series depicting the number of news clips, for each day, retrieved for a hypothetical query. The solid line represents the time series itself, while the dotted line represents the event strengths, calculated by our algorithm, and the square points represent the identified events.

Our algorithm is based on the idea that a peak should only represent an event for the given query whenever its value is well above average when compared the number of news clips for each of the remaining days. Let's assume an hypothetical scenario where a user would search for a given query (e.g. "obama romney election") and a list of dated results would be returned. Fig. 3 shows a time series, illustrating the distribution over time of search results, within the range of returned dates, filling days without results with the value zero. For each known value, our algorithm calculates the event strength, a score that helps determine the probability of a peak corresponding to a relevant event. The event strength for a given point t_i depends on the value of the point y_i as well as the left and right values, y_ℓ and y_r respectively, where $\ell \in [i-n, i-1]$ and $r \in [i+1, i+n]$, for a given window of size n. In this specific example, we used $n = 3$. Event strength E_{str} is calculated as described in Equation 1.

$$E_{str} = 2 \times y_i - \left(\sum_{\ell=i-n}^{i-1} y_\ell + \sum_{r=i+1}^{i+n} y_r \right) \tag{1}$$

The event strength values for each point are also depicted in Fig. 3 using a dotted line. As we can see, irrelevant points tend to have negative or very low values, while relevant points have a much higher event strength. In order to retrieve the event points, we now need to define a threshold above which we accept a peak as an event. Notice that some of the low value peaks, also have a positive value for the event strength, however they aren't much different from their neighbourhood and shouldn't therefore be considered a relevant event. On the other hand, there are three peaks (marked with a square) that depict an out of the ordinary number of results, for the given query, on a single day. Each of these peaks has an event strength that is much higher than the peak's value. This characteristic is unique to relevant events. Thus, an event for the requested search is identified in a point t_i whenever, $E_{str}^{(i)} - y_i > \sigma_Y$, where $E_{str}^{(i)}$ is the event strength for point t_i and σ_Y is the standard deviation for the time series values.

4.4 Graph Navigation

Provided a topic and a timespan, the system fetches a given number of clips. These are selected, searching the keywords in the title, tags or body of each news clip. Furthermore, the user can specify in which field to search for a given keyword. In this way, the system is able to filter the clips quite narrowly, in accordance with the user's indications. The resultant clips are presented as the initial nodes to the user, and make up the kernel for the forthcoming graph. The system builds the graph with joining the neighbour clips: clips that have a relationship with any of the searched ones, that is, based on the people, places and dates they may both reference.

The graph itself is built iteratively, expanding with each consecutive user action. Thus, when the user enters a new topic or changes the timespan, the browser requests a list of clips for the given topic, over the specified timespan, as we described before. These are presented on the screen as unconnected nodes. The browser than requests the server a graph that contains these last clips and their neighbours.

In order to maintain an trimmed graph, some nodes are gradually withdrawn from the graph, being selected based on two factors: the clip's date and the node's order of entry in the graph. The latter is a simple threshold to prevent the overcrowding of the screen. Beyond maintaining a visually engaging graph, this factor contributes to the reducing of the computational complexity of graph visualization, as less elements need to be positioned on the screen. On the other hand, the nodes whose news clip's date is no longer over the selected timespan are the ones more commonly removed, ensuring a certain consistency to the displayed graph, as only the specified timespan is shown to the user.

When a new search is performed, the system fetches the list of events for the given topic. As depicted in Fig. 4, the user inserts the search terms in (A). The events are displayed in the timeline, as markers (E) under the slider bar. The marker's tone is proportional to the intensity of the event.

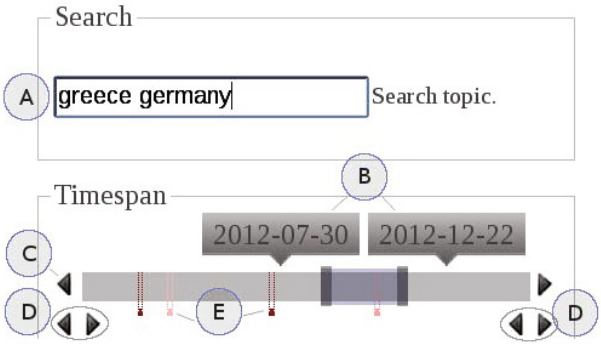


Fig. 4. System's parameter configuration: giving a search input and a timespan

In order to facilitate the navigation in a wide time range, the user can shift the slider's lower and upper bounds (D). This allows for a more or less extended view of any time range. Anticipating a substantial use of the slider, while the user is searching for a given event or exploring the evolution of different clip clusterings, this element received special attention. Accordingly, the user has various interactions methods available, the slider can be moved by clicking the arrows aside it (C), dragging its limits or date labels (B), or using the mouse wheel.

The graph is interactive, so the user can move nodes around to improve the browsing experience. Moreover, the user can pinpoint and, with a mouse click, expand a node, that is, display its neighbours within the selected timespan. This feature allows for a loosely and rather precise exploration of the graph, potentially enabling the discovery of supplementary clips.

5 Discussion

To demonstrate the applications of our multidimensional network visualization, we present some use-cases. The combination of our system's different features allows for easier information retrieval and browsing, providing insight into the knowledge underlying in the data.

The starting point, when using the system, is the search query and the timespan. The latter is quite straightforward, the user can narrow the browsing in a timespan more or less confined according to the knowledge of a relevant exact date or the larger period when a series of events might have happened. Regardless, the time span can be shorten in the case of a rather wide period, or the timespan can be readily scrolled until interesting clips emerge.

Browsing the network through time allows the user to analyse sequence of events in the news. Taking, for example, news clips related to Israel. In the wake of the US elections, clips discuss the relation between the two countries and the election's outcome on Israel's politics. In the following days, the focus

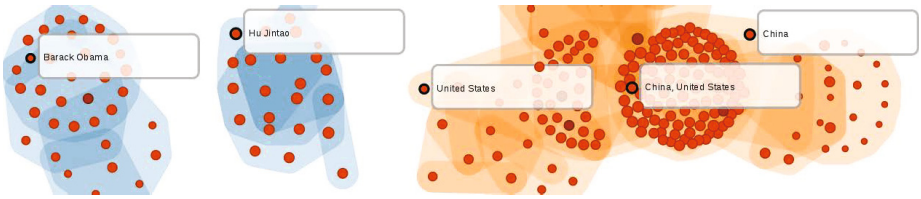


Fig. 5. Comparison between different co-reference dimensions

shifts back to the ongoing conflict in Syria and references to countries in the region emerge. Finally, news clips arise relating to the Gaza conflict, foretelling its escalation in the following days. The temporal navigation allows the user to sift through these different but related events and to perform a detailed analysis in that period context.

Turning to the visualization itself, the user is able to perform a broad analysis of the information presented. Taking an example concerning the United States and China, both undergoing a power transition at the time of the query. The resulting graph for the two countries or leaders will show them connected by the country entities, as shown in the right graph in Fig. 5. This is expected due to geopolitical factors and economical ties between the two. Nonetheless, if we filter out spacial references leaving only the entities from the class *People*, we note a much thinner, or no connection at all, between the two communities.

6 Conclusions and Future Work

In this paper, we present our temporal visualization system of a multidimensional network, with search and temporal navigation features, including an event detection system. We propose a distinct force-directed graph visualization where the edges are hidden, instead displaying convex hulls covering nodes co-referencing the same entity. This reduces the graph's rendering computational complexity and visual clutter. Our system enables the user to browse a large network of news clips spanning a wide timespan. Beyond browsing, the listed key features play an important role in the extraction of knowledge, providing an intuitive and effective user interface.

Future work in this area should focus on the scalability. Further developments would be needed in the algorithms in order to handle relatively large graphs, that may arise even for small a timespan.

Acknowledgements. This work is financed by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project PEst-C/EEI/LA0014/2011. We would like to thank Nuno Cravino for the helpful discussions, specially regarding the event detection algorithm.

References

1. Abello, J., van Ham, F., Krishnan, N.: Ask-graphview: A large scale graph visualization system. IEEE Transactions on Visualization and Computer Graphics 12(5), 669–676 (2006)
2. Apache Software Foundation: Apache Lucene (2001), http://lucene.apache.org/ (retrieved November 9, 2012)
3. Apache Software Foundation: Apache Solr (2006), http://lucene.apache.org/solr (retrieved November 9, 2012)
4. Bostock, M., Ogievetsky, V., Heer, J.: D-3: Data-Driven Documents. IEEE Transactions on Visualization and Computer Graphics 17(12), 2301–2309 (2011); IEEE Visualization Conference (Vis)/IEEE Information Visualization Conference (InfoVis), Providence, RI, October 23-28 (2011)
5. Brandes, U., Eiglsperger, M., Herman, I., Himsolt, M., Marshall, M.S.: Graphml progress report - structural layer proposal (2002)
6. Devezas, J., Figueira, A.: Finding Language-Independent Contextual Supernodes on Coreference Networks. IAENG International Journal of Computer Science 39(2), 200–207 (2012), http://www.iaeng.org/IJCS/issues_v39/issue_2/IJCS_39_2_07.pdf
7. Devezas, J., Figueira, A.: Interactive Visualization of a News Clips Network: A Journalistic Research and Knowledge Discovery Tool. In: Proceedings of the 4th International Conference on Knowledge Discovery and Information Retrieval (KDIR 2012), Barcelona, Spain (2012)
8. Erdős, P., Rényi, A.: On the evolution of random graphs. Magyar Tud. Akad. Mat. Kutató Int. Közl. 5, 17–61 (1960)
9. Fruchterman, T., Reingold, E.: Graph drawing bt force-directed placement. Software-Practice & Experience 21(11), 1129–1164 (1991)
10. Herman, I., Melancon, G., Marshall, M.: Graph visualization and navigation in information visualization: A survey. IEEE Transactions on Visualization and Computer Graphics 6(1), 24–43 (2000)
11. Keim, D.: Information visualization and visual data mining. IEEE Transactions on Visualization and Computer Graphics 8(1), 1–8 (2002)
12. Rohrdantz, C., Oelke, D., Krstajic, M., Fischer, F.: Real-Time Visualization of Streaming Text Data: Tasks and Challenges. In: Workshop on Interactive Visual Text Analytics for Decision-Making at the IEEE VisWeek 2011, Providence, RI, USA (2011)
13. Shen, Z., Ma, K.L., Eliassi-Rad, T.: Visual analysis of large heterogeneous social networks by semantic and structural abstraction. IEEE Transactions on Visualization and Computer Graphics 12(6), 1427–1439 (2006), http://dx.doi.org/10.1109/TVCG.2006.107
14. Wong, P.C.: Guest editor's introduction: Visual data mining. IEEE Computer Graphics and Applications 19, 20–21 (1999)

User Documentation: The Cinderella of Information Systems

Brigit van Loggem

Open Universiteit, Heerlen, The Netherlands
Brigit.vanLoggem@ou.nl

Abstract. User documentation is a structurally under-valued aspect of information systems and end user assistance. Popular belief has it that documentation is required only where insufficient attention has been paid to the design of usability, user satisfaction and user experience; and that documentation is not referred to by users anyhow. Both lines of reasoning are shown to be faulty, and a case is made for an interest in user documentation from the pertinent research communities.

Keywords: user documentation, documentation design.

1 The Case for User Documentation

User documentation is the Cinderella of information systems, regularly dismissed as irrelevant because only there to make up for defects in the software's user interface. If the latter were better, the argument goes, the former would not be required: so all our efforts should be geared towards improved interaction design. The dismissal is sometimes strengthened with the corollary that "nobody reads the manual anyhow"; after which the speaker turns away to attend to more pressing business.

It is my contention that both the main argument and its corollary are rooted in misconception; and also that user documentation is not something that will quietly go away if ignored long enough. Documentation is (or should be) a core deliverable of any but the most trivial of software systems [1]. In this paper I shall make a case for software documentation design as a legitimate topic of academic interest. I shall first attempt to disprove both the "documentation as unnecessary evil" and the "nobody reads the manual" lines of reasoning, and then discuss the research—or rather lack of research—on the topic.

1.1 "Well-Designed Software Needs No Documentation"

The first misconception, that documentation is no more than a patch for bad interaction design, does not stand up to close scrutiny for two reasons.

First, even if documentation could be done away with by good product design, this cannot be an excuse for refusing to pay attention to it. The imperfect world in which we live contains many badly-designed products that people have to deal with, whether

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 167–177.
DOI: 10.1007/978-3-642-36981-0_16

they like it or not. Figure 1 shows ad-hoc documentation for a sliding door; obviously created to meet a real-life requirement. Arguably, the door was designed badly in the first place and in a perfect world there would be no place for it. But it was there when the photograph was taken, at one of the Netherlands' railway stations: and people were, quite literally as well as figuratively one presumes, stuck with it.

Fig. 1. Instructions on how to use a sliding door: This is a sliding door. Therefore do not push or pull! Door does not automatically open and close. (Photo courtesy Ester Moraal, 2010)

Bad designs do happen, in the world of software as well as that of sliding doors. Moreover, the market mechanism guarantees that where the return-on-investment in user experience (or user satisfaction, or even straightforward usability) is zero, bad designs will continue to happen. Most software is employed not to satisfy the needs and requirements of those who pay to use it, and who might be willing to pay more for good user experience. Instead, the vast majority of software systems is created to do things such as controlling traffic lights, supporting the trade in stocks and bonds, or keeping track of the location of lab coats in hospitals. For most "non-consumer software", at least that which is not employed in high-risk environments such as

aerospace, any additional expenditure following from attention to usable design would never be reclaimed. And when something is not economically viable, it will not be done.

For this reason alone documentation will always be needed. But there is a more interesting reason. There exists a fundamental difference between sliding doors and software. Both are artefacts, man-made tools. In order to apply any tool to a task, its users need to learn its operations: in other words, they need to learn new ways of doing things. Yet where tools such as a sliding door, a hammer or a washing machine directly change the material world, software tools do not. Operating a sliding door causes a physical barrier to appear or disappear. Operating a hammer causes (hopefully) a nail to be driven into a piece of wood and operating a washing machine causes water, detergent and linen to be pushed around. This differs fundamentally from a computer system, where user actions frequently have a direct result only within the software environment itself.

A software system is a self-contained "world" with its own objects (think of the Clipboard in many operating systems; of templates, style sheets and fields in a word processing environment; or of layers in an image editor). These are at the basis of software's capacity for being used in many different ways to many different ends and allow not just for new ways of doing things but also for new things to do—provided we become familiar not just with the outer, visible layer of the system but also with its software-specific objects, their mutual dependencies and the rules governing their behaviour. To use a software system to its full potential, its user must look beyond the user interface and fully understand the whole of the "User Virtual Machine" or UVM; which is defined as "not only everything that a user can perceive or experience (as far as it has a meaning), but also aspects of internal structure and processes as far as the user should be aware of them" [2]. The visible part of the UVM is what these authors refer to as the "perceptual interface" and what is more commonly referred to as the user interface; but the UVM as a whole is a much larger conceptual machine that exists in the user's mind.

In 1999 Donald Norman described the "single, general" personal computer as "a great compromise, sacrificing simplicity, ease of use, and stability for the technical goals of having one device do all" and as a "fundamentally difficult machine" [3]. By trying to make one product do many things, Norman noted, complexity increases. Thus, complexity is built into any artefact that is not geared towards one specific, well-defined task—and the large-scale, sophisticated software packages that are common in the PC environment certainly meet this description. Norman concluded his 1999 book with a passionate cry for the computer to disappear into tools specific to tasks: "a world of information appliances". Some ten years later this vision had become reality when Apple's iPad™ took the world by storm. This is a flat, hand-held device consisting, as far as the user is concerned, mainly of a screen and populated with so-called "apps" that each perform one very specific task. Almost without exception, apps do not require documentation and indeed they come without. However, it is difficult to envisage all multi-functional software replaced with app-like information appliances, if only because there will always be tasks for which a complete toolbox is required rather than one single tool. And indeed, at around the

time of the spectacular rise of the iPad, Norman dedicated a whole book to the necessity of complexity, in which he wrote: “I find it interesting that we complain when a new technology requires an hour or two of study. [...] When new items are appropriately complex, it is reasonable that they require time and effort to master” [4].

It is highly unlikely that a company such as Microsoft Corporation leaves the design of the user interaction of its products to amateurs. We must therefore assume that Microsoft’s flagship word processor Word™ was carefully designed by competent professionals, striving for the highest possible usability. Yet a group of science teachers using the program extensively in the course of their everyday work gave rise to the bemused observation [5] that “almost all of them claimed that they were not fit subjects for the experiment because they were not expert users … Considering the high levels of education and experience of the subjects, the most surprising result was the simplistic level of their interaction with this sophisticated but very familiar software tool.” Word™ is complex, not unnecessarily so because it was badly designed; but necessarily so because it is a very sophisticated tool that can be put to uses that its designers could not even dream of—up to and including (though its table functionality) the design of knitting patterns.

Unnecessary complexity is the inverse of usability and can be removed through good design of the user interface. Necessary complexity on the other hand is a consequence of versatility. For those systems that are necessarily complex, insights from earlier years [6] will continue to hold: that there are types of information regarding the system that the user must know and that cannot be communicated within the user interface but for which documentation is the only medium. Even in a perfect world in which all software designs are of the highest possible standard, the idea that the need for user documentation will eventually fade away is a “persistent myth” [7].

1.2 “Nobody Reads Documentation, Anyhow”

The respondents who caused Ben-Ari and Yeshno such puzzlement over their lack of expertise [5] almost certainly worked with a version of Word™ that came equipped with online Help and possibly with an extensive printed user manual as well: although Microsoft stopped delivering printed user manuals with Word97, which came on the market in December 1996, every version of this popular word processor thus far has come with online Help. Does this then mean that users, as is often said [8], “simply won’t read the instructions”? Documentation developers the world over secretly hope for permission from management to have printed on every page the catch-phrase, “Read The F… Manual!” or RTFM for short. Yet a respondent in a study on the appropriation of the iPhone (!) was quoted to say: “I don’t have the time to explore. I miss having a manual that I can lean back and read.” [9]. A handful of research studies have been carried out to determine whether users are indeed reluctant to consult the documentation that is delivered with the product, and they are surprisingly unanimous in their findings. Invariably, it is shown that—at least for complex and unfamiliar products—the documentation *is* consulted; even if it is not read, marked, learned, and inwardly digested in its entirety [10-12, 13 p. 213, 14].

Although many of these results pertain to consumer products, we know that users of software do not necessarily avoid consulting documentation either, or have given up altogether on the idea of getting help from the written word: because we can see them vote with their feet. A visit to any bookshop will show long shelves filled with software-related books. These bear witness to a genuine demand: if there was no market for these books then they would not be there; but instead people go out of their way to pay good money for them and the market is a thriving one.

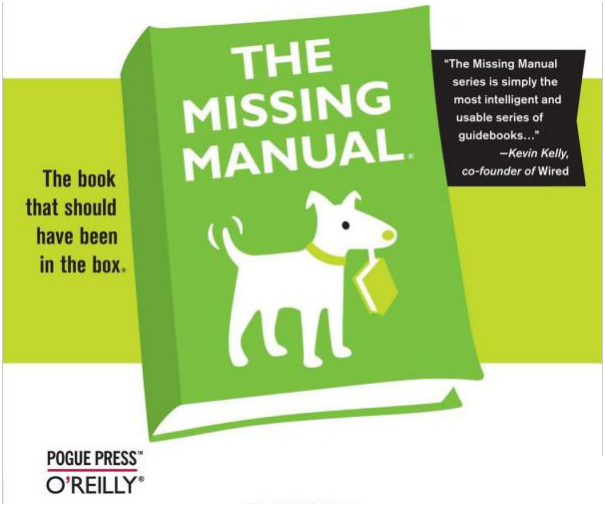


Fig. 2. Part of the cover of one book in the "Missing Manuals" series, published by O'Reilly, Sebastopol, CA, USA

2 Designing Documentation

Designers of user documentation are faced with three serious challenges.

First of all, it is often impossible for the designer of documentation to give his readers what they want and look for, namely, a direct solution to whatever problem is currently presenting itself. Users frequently want to be told "just what to do" and look in documentation for procedural information. This is not only what readers ask for, it is also what authors work hard to give them and what many researchers regard as key. Unfortunately, if the software is complex as a consequence of versatility, then it may be that the documentation designer cannot possibly know what it is that the user will want to do and thus what the user will need to do to achieve his goals. It follows then that the documentation cannot give them what they want but must give them what they need: information with which the knowledge and skills can be built to find their own solutions to their own problems [15].

Then, consider the following. Regardless of the discipline in which they work, designers find it useful to distinguish between primary and secondary tasks; the primary task being that which someone sets out to achieve, and the secondary task that which is created by any design used in order to complete the primary task. Using documentation is not a primary task, as is, for example, writing a report. It is not even a secondary task, as is using a word processor to write a report. Using documentation is in fact a tertiary task: documentation must sometimes be endured in order to use a tool (such as a word processor) in order to fulfil a motivation (such as writing a report). And it does not always end here. Some software is so complex that in order to use it to its full potential, users must learn the ins and outs of the system before they are able to do any meaningful work with it. The learning process then becomes a (quaternary!) task in its own right, three steps removed from the original motivation—and from the user's point of view, presenting yet another obstacle between his desires and their fulfillment (see Figure 3). It is not surprising then if users have little patience with the user manual.

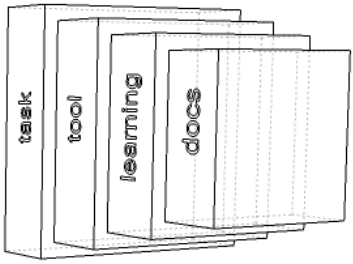


Fig. 3. Documentation as a quaternary task

Finally, a key characteristic of documentation is that the communication between the creator of the documentation—in the context of a particular documentation product more commonly referred to as the product's author—and the user in his role of reader is asynchronous. Sender and receiver of the message are separated in time and place. As a consequence, individual reader differences and preferences cannot be catered for. Although the information contained in documentation may be presented to the user interactively, its scope is then still predetermined. Because of its asynchronous nature, documentation is used at the reader's discretion. The author cannot enforce the quality or quantity of use and must depend on the readers' willingness to use it, and to do so in the way the author envisioned [16]. Referring to Figure 3, we must acknowledge that no user can be forced to work his way through the whole pyramid in order to get to the desired completion of his primary task. He may easily go round part or all of it, dispensing with the documentation and perhaps with the learning process as well. It is important to always remember that we can never force-feed the information: all we can do is attempt to drip-feed the required learning.

2.1 Stuck in a Rut

Unfortunately, documentation designers' knowledge and skills have not developed at the same pace as the software that they document [1]. User documentation in the 2010s looks very much the same as it has looked for a long time. A typical user manual or Help system is hierarchically structured with the chapter as its main unit. Sometimes the chapters are grouped into sections, or groups of chapters are split off into separate documentation products. In this manner, the "installation manual" and the "advanced user's guide" may come about. Each chapter contains an operational description of and (usually stepwise) instructions for working with a particular program concept or user task. Conceptual descriptions are added to place that which follows in context; in addition, a number of introductory chapters or appendixes provide program-wide meta-information by listing hardware requirements, providing installation and troubleshooting instructions, and informing the reader of terminology and typographical conventions used in the manual. This prototypical approach to documentation has remained unchanged for a number of decades.

Documentation designers, in the Skills Framework for the Information Age (SFIA.org.uk) referred to as "Information Content Authors" or INCA, base their designs on the prescriptions of guidebooks and standards such as the IEEE Standard for Software User Documentation (IEEE Std 1063-2001) and the ISO Standard Systems and software engineering — Requirements for designers and developers of user documentation (ISO/IEC FDIS 26514:2008). Such prescriptions change only slowly and tend not to reflect the latest developments. Yet software in the 2010s is a far cry from that in the 1980s, when the guidelines for documentation were first formulated. Whereas in the first decades of computer use by the general public people used computers mostly for reproducing their work, for automating tasks that they could at least theoretically have done without a computer, more and more software nowadays enhances people's work in such a way that it enables them to do non-routine things that were unthinkable before.

This is most salient when the end result of the work does not lie anywhere in the outside world, or even directly corresponds to a particular state in the outside world, but resides completely within the boundaries of the software world. The work is then undertaken with the sole objective of creating a particular software state, stored in a file or a set of files. People create spreadsheets to help plan their annual budgets, their vegetable gardens or their cross-stitch projects; they create websites, page-by-page or through a Content Management System; they send and receive electronic mail; they create long, camera-ready manuscripts; they write smaller or larger programs; and they set up software environments for themselves or others to work in. For this type of activity, there no longer exists a distinction between the task and the tool, as the task is defined as working with the tool. Efficiency and effectiveness, the traditional measures of performance [17], no longer suffice: suboptimal performance can be evidenced not only by inefficiency or ineffectiveness but also by under-use of the software.

When asked to document today's generation of sophisticated software, documentation designers are faced with a qualitatively different task from before. The goal posts have been moved, but the rules of the game have remained unchanged and do not yet match the new situation: the guidelines for the design of quality documentation as laid down in time-honoured standards and guidelines may no longer always be appropriate.

3 Wanted: Research

Given the requirement for quality documentation and the difficulties faced by documentation designers, it is clear that there exists a genuine need for theory-based and evidence-based guidance on how to document different types of software in an appropriate manner; a need that is felt both by the documentation design community and by the beneficiaries of its efforts: the users of software [13]. This need has not yet been answered by the academic community. A systematic search for publications uncovered no more than seven (7) scholarly papers into the design of software documentation for end users, published during the 5-year period from July 2007 to July 2012. During that same period of five years an additional nine (9) such papers were published that discussed user documentation in general, or for products other than software; bringing the total to a paltry sixteen (16)[1]. Even allowing for my overlooking some publications, these figures can be expected to fall within the correct order of magnitude. They are so extremely low as to be almost unbelievable. How can they be accounted for?

What documentation design has in common with other user-support disciplines is its multi-disciplinary nature. A manual is a form of non-fictional, instructional writing: the fields of communication science and library and information science are heavily involved, as are those of instructional design and cognitive psychology. Finally, the writing of software manuals requires technical and business knowledge of computer use and information processing. With so many godfathers, one might expect the profession to be able to draw on a wealth of research studies and theory providing guidelines for carrying out the task. Yet like a true Cinderella, it is regularly overlooked as "not sexy". Theories of learning tend to be swept up to form a starting point for the design of classroom instruction and self-study materials, while theory and experimental results from cognitive psychology have been taken up by the applied field of interaction design; which is where the interest of computer science in user support ends. Library and information science is highly system-oriented and has an established tradition of focusing on formal information systems for document retrieval from repositories rather than on information retrieval from documents. Communication science, finally, is the natural home for documentation design; but it covers a broad area of which technical communication is but a small part and the design of documentation, be it for software tools or other artefacts, an even smaller part.

[1] Details of this literature study can be obtained from the author.

Even when research into documentation is carried out, this tends to embrace new technology-enabled media of communication without questioning. Almost as soon as it was recognized that written instructions are vital for optimal performance, the research focus shifted from paper-based documentation to hypertext and soft copy [1]. Since then, it has shifted further: to the World Wide Web, then to CSCW or "computer-supported cooperative work", to social media, and at the time of my writing this to mobile appliances and ubiquitous computing channels for the delivery of information. Lured away by the beckoning of glamorous new media in cyberspace from mundane matters such as the content of the documentation or the application of printed matter, researchers "do not always make clear the boundary conditions within which their findings will apply" [1]. Research questions ask *how* a particular medium of communication can most fruitfully deployed; glossing over the question *when* (that is, under which conditions) or *why* any effects that are found may hold.

To know the conditions under which a particular design solution is useful, such a solution must be built upon a naturalistic and descriptive model on an abstract level [18]. We need to know the mechanism by which people's interaction with software changes with repeated practice before we can think about designing interventions to guide the process towards a particular outcome. And before we can think about applying a particular intervention we need to know how people interact with our proposed interventions: it has been known for a long time that the applicability of a particular solution depends heavily on context, even where domain as well as audience are the same [13, 19]. In summary, the informed design of software documentation demands that the choice for medium and format of the communication, as well as its content, be based on an understanding of the underlying processes of people interacting with software and with documentation [1, 13].

Ecologically valid working models of these processes are not currently available. Interaction with software is usually discussed with a view to immediate performance of a task lying outside the software world, rather than as an activity in and of itself that changes over time as learning takes place. Likewise, interaction with documentation is studied with a view to the immediate delivery of information that the user sets out to find, rather than the provision of an environment in which people may learn things they need to know, whether they are aware of the need or not. Among the many studies into the general case of "information behaviour" of human actors [20], none systematically target the more specific "documentation behaviour" of tool users.

When user interaction design was still a fledgling field, dreams were dreamt of finding "a set of basic design rules of the following kind: IF user(*i, j*…) AND task (*k, l* …) THEN apply design principles (*m, n* …) where *i, j* stand for variables that characterize particular (groups of) users, *k, l* for task variables, and *m, n* for design principles […]" [6]. In a similar vein dreams should at least be dreamt, even if turning them into reality will prove as elusive as it has been for interaction design, of a set of basic design rules expressed in the form IF computerized_activity (*i, j* …) AND documentation_behaviour(*k, l*…) THEN apply design_principles (*m, n* …); where *i, j* now stand for variables that characterize software-in-use (the context in which the work is done insofar as this is determined by a software tool), *k, l* for variables

characterizing the ways people may interact with and learn from documentation, and *m, n* for design principles of user documentation. What is needed is a framework within which the dreamt-of complete set could be expressed: insight is needed into the nature of the parameters *i, j* … , *k, l* … and *m, n* … . When people repeatedly use a particular software system, which are the mechanisms by which their performance changes and hopefully improves over time? Does intrinsic or extrinsic motivation perhaps play a role, or the ambition to achieve ever more satisfying results, or the degree to which presumed familiarity dampens exploration? Under which conditions do users refer to the documentation and how do they apply the information found in it?

Such a framework would allow for the expression of the design rationale underlying documentation design decisions. It would provide the scaffolding that Cinderella needs to put on her slippers and go to the ball.

References

1. Wright, P.: Printed instructions: Can research make a difference? In: Zwaga, H.J.G., Boersema, T., Hoonhout, H.C.M. (eds.) Visual Information for Everyday Use: Design and Research Perspectives, pp. 45–66. Taylor & Francis, London (1998)
2. van der Veer, G.C., van Vliet, H.: The Human-Computer Interface is the System: A Plea for a Poor Man's HCI Component in Software Engineering Curricula. In: Ramsey, D., Bourque, P., Dupuis, R. (eds.) Proceedings of the 14th Conference on Software Engineering Education and Training, pp. 276–286. IEEE Computer Society (2001)
3. Norman, D.A.: The Invisible Computer. The MIT Press, Cambridge (1999)
4. Norman, D.A.: Living with Complexity. The MIT Press, Cambridge (2010)
5. Ben-Ari, M., Yeshno, T.: Conceptual Models of Software Artifacts. Interacting with Computers 18, 1336–1350 (2006)
6. van der Veer, G.C., Tauber, M.J., Waern, Y., van Muylwijk, B.: On the interaction between system and user characteristics. Behaviour & Information Technology 4, 289–308 (1985)
7. Mehlenbacher, B.: Documentation: not yet implemented, but coming soon. In: Jacko, J.A., Sears, A. (eds.) The HCI Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, pp. 527–543. Lawrence Erlnaum, Mahwah (2003)
8. Rettig, M.: Nobody reads documentation. Commun. ACM 34, 19–24 (1991)
9. Bødker, S., Christiansen, E.T.: Poetry in motion—appropriation of the world of Apps. In: ECCE 2012, Edinburgh, Schotland, UK, pp. 78–84 (2012)
10. Jansen, C., Balijon, S.: How do people use instruction guides? Confirming and disconfirming patterns of use. Document Design 3, 195–204 (2002)
11. Mehlenbacher, B., Wogalter, M.S., Laughery, K.R.: On the reading of product owner's manuals: Perceptions and product complexity, pp. 730–734. SAGE Publications, Thousand Oaks (2002)
12. Wright, P., Creighton, P., Threlfall, S.: Some factors determining when instructions will be read. Ergonomics 25, 225–237 (1982)
13. Schriver, K.: Dynamics in document design. John Wiley & Sons, New York (1997)
14. Vromen, N., Overduin, M.: Handleiding: de Titanic onder de gebruikersondersteuning. Tekst (blad) 6, 42–46 (2000)

15. van Loggem, B.E.: Paper-Based Support for Computer-Mediated Activity: a Cognitive Task Analysis. Master's. Twente University, Enschede (2007)
16. Novick, D.G., Ward, K.: What users say they want in documentation. In: Proceedings of the 24th Annual ACM International Conference on Design of Communication, pp. 84–91. ACM, Myrtle Beach (2006)
17. Frøkjær, E., Hertzum, M., Hornbæk, K.: Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 345–352. ACM, The Hague (2000)
18. Kirlik, A.: Cognitive Engineering: Toward a Workable Concept of Mind. In: Kirlik, A. (ed.) Adaptive Perspectives on Human-Technology Interaction, pp. 3–9. Oxford University Press, Inc., Oxford (2006)
19. Wright, P., Reid, F.: Written information: Some alternatives to prose for expressing the outcomes of complex contingencies. Journal of Applied Psychology 57, 160–166 (1973)
20. Case, D.O.: Looking for Information. Emerald, Baton Rouge (2007)

Task Topic Knowledge vs. Background Domain Knowledge: Impact of Two Types of Knowledge on User Search Performance

Xiangmin Zhang[1], Jingjing Liu[2], and Michael Cole[3]

[1] School of Library and Information Science,
Wayne State University Detroit, MI 48202
ae9101@wayne.edu
[2] School of Library and Information Science,
University of South Carolina, Columbia, SC 29208
jingjing@sc.edu
[3] School of Communication and Information,
Rutgers University 4 Huntington Street,
New Brunswick, NJ 08901
m.cole@rutgers.edu

Abstract. Domain knowledge affects search success and behaviors but different aspects of domain knowledge, i.e., topic knowledge and background knowledge, has not normally been distinguished. We were interested in finding out if the two would have different effects on search performance for typical research search tasks, and we report empirical evidence from a user study. Participants' topic knowledge and background knowledge were measured separately as different aspects of domain knowledge. Our results indicate that topic knowledge and background knowledge impact search performance at different levels. While topic knowledge was significantly correlated with performance measures for individual tasks, background knowledge was significantly correlated with performance at more general levels, over multiple task sessions. This finding furthers our understanding of the role of domain knowledge in search activities.

Keywords: Personalization, information retrieval, domain knowledge, search performance, user studies.

1 Introduction

The effects of user knowledge on their search behaviors and search performance have been investigated for a long time and proved that knowledge is an important factor influencing users' search. Various studies have used the term "knowledge" or the phrase "domain knowledge", however, the investigations did not distinguish different types of knowledge involved in the search process.

In this study, we differentiate background domain knowledge from search task topic knowledge. The former represents one's knowledge in a domain, which does not

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 179–191.
DOI: 10.1007/978-3-642-36981-0_17

directly relate to the current search topic, but should be very helpful for the user to be able to understand the current search topic. The latter is one's knowledge of the search topic at hand, and thus is directly related to the current search. We attempt to find out if there are differences in the effects of the two types of knowledge on searching, and if yes, how they affect searching. Our ultimate goal is to identify properties of observable search behaviors that can serve as implicit indicators of a user's level of knowledge so that the system can personalize search results towards the user's knowledge level.

We chose genomics as the domain for this study based on several considerations: a) its knowledge is systematic and well defined; b) the knowledge is well represented by the MeSH thesaurus, a hierarchical concept structure that is used broadly for processing documents in this field; c) the availability of TREC[1] (Text REtrieval Conference) Genomics data set, which included millions of documents, search topics, and the relevance judgments by experts for the documents associated with each topics.

2 Related Work

The effects of user's knowledge on information search have received considerable attention. Nevertheless, there have rarely been explicit definitions of knowledge and/or classifications of different types of knowledge in the information science literature, except a few such as Allen [1] which noted that domain knowledge is general and topical knowledge is the specific factual knowledge of a topic. In general, knowledge, or domain knowledge, has been used to refer to "some type" of user knowledge. An examination of the literature revealed that there are two different types of content-related knowledge: 1) user's domain knowledge, which we define it in this paper as a person's background knowledge (BK) of a general subject, domain or sub-domain, for example, medicine, law, psychology, or any sub-domains under these, which is more general and broader than the topic of the specific search task that the user works on; 2) user's task topic knowledge (TK), which refers to the user's knowledge or familiarity with the topic of his/her specific task at hand, as reviewed below .

2.1 Domain Knowledge

As mentioned above, domain knowledge in previous studies was not clearly defined. Previous work about domain knowledge reviewed in this sub-section follows our above definition of domain knowledge. If some work used the term "domain knowledge" but was not actually about user's knowledge of a subject or domain, it is not included here.

One typical way to investigate the effect of people's level of domain knowledge is to recruit two groups of people, one expert group consisting of users in a discipline or

[1] http://trec.nist.gov/

familiar with the *area* of their search task topic, and one novice group consisting of users outside of a discipline or not familiar with the *area* of the search task. For example, the Hsieh-Yee [6] study found that domain knowledge affected search tactics. As compared to a familiar subject area, when users worked with a search task outside of their field, they used the thesaurus more for term suggestion, made more effort in preparing for the search, monitored the search more closely, included more synonyms, and tried more term combinations. Sihvonen & Vakkari [10] found that the number and type of terms selected from the thesaurus for expansion by experts improved search effectiveness, whereas the use of the thesaurus by novices had no impact.

Some studies have assessed the effects of differences in domain knowledge by observing users' behavioral differences at different times. Vakkari et al. [11] found that students began to use a wider and more specific vocabulary in their development of research proposals at the end of a 3-month seminar compared with in the beginning. Wildemuth [13] found that low domain knowledge, before the users took a microbiology course, was associated with less efficient selection of concepts to include in the search and with more errors in the reformulation of search tactics.

Domain knowledge has also been assessed by user ratings on some terms in a domain/area. Zhang et al. [14] had participants rate their familiarity with 200 terms from the Heat and Thermodynamics category in the Engineering Information Thesaurus, and divided them into two groups (low-level and high-level) based on their rating scores. The high-level group of users were found to have better performance (retrieved slightly more relevant documents), issued longer queries, and had more queries per task.

Another way is to test users' answers to some questions in a subject area. Duggan & Payne [3] assessed participants' knowledge by the correct rate of their answers to questions in the soccer the music domain. They found that knowledge of the music domain had little effect on search performance, but that of the football domain had much effect on search performance: it was positively correlated with search accuracy, and negatively correlated with time spent on web pages and mean query length.

In a study exploring web search behaviors using large-scale naturalistic search logs, White et al. [12] separated domain experts and novices by whether or not they visited specific web sites, named expert URL filters, which were identified through discussion with domain experts in each of the four subject areas. The study found that within their domain of expertise, experts search differently than non-experts in terms of the sites they visit, the query vocabulary they use, their patterns of search behavior, and their search success.

In sum, previous studies all found that user's domain knowledge showed significant effects on users search behaviors and search performance, although there were some inconsistencies. Most of the studies found that higher domain knowledge was associated with less use of thesaurus terms [6], but when using them, with an improved search performance [10], longer queries ([14], [12]) and so on. Others found that higher domain knowledge was associated with shorter queries in the football domain [3]. The previous studies' focus on a single general "domain

knowledge" with significant variances across studies may explain certain inconsistencies but also undermines to some degree the ability to generalize across studies.

2.2 Search Task Topic Knowledge

Some studies investigated the effect of search topic knowledge in information search. Allen [1] defined "topical knowledge" as specific factual knowledge of a topic. He examined the effect of knowledge levels of a topic on online catalog searching behaviors. Participants were given 12-question multiple-choice test to measure their level of factual knowledge of the search topic (planetary exploration), and were divided into three groups: 18 as high-knowledge who correctly answered 9 or more questions; 18 as low-knowledge who correctly answered 5 or fewer questions, and the remaining 24 participants as medium level. The study found people with high topical knowledge used more search expressions than low knowledge people.

Hembrooke et al. [4] investigated the effects of knowledge on users' search term selection and reformulation strategies for web searches in a study with 41 students searching in a simulated database. From a list of topic areas, participants were asked to choose two topics where they had some expertise (knowledge) and two in which they had none. Experts were found to issue longer and more complex queries than novices. Experts also used elaborations as a reformulation strategy more often as compared to simple stemming and backtracking modifications used by novices.

Elicitation of user's topic knowledge or familiarity can also be done by asking users to self reports on a 5- or 7-point scale with regard to their familiarity with the topic of the search task in a questionnaire. Kelly and Cool [7] found that reading time tends to decrease and efficacy increases with increasing topic familiarity. Liu and Belkin [8] found the total dwell time that users spent on a document was positively correlated with document usefulness regardless of the user's topic knowledge level, but the first dwell time (the unobtrusive duration from opening a document to the user leaving it for the first time) was not significant correlated with document usefulness, and topic knowledge affected their correlation.

While the two types of knowledge are clearly different by definition, one would ask if they play different roles in affecting users' search. It is not practical to compare their roles in a direct way just from reviewing the literature. Our study provides a means for a direct comparison based on our multiple methods employed to elicit user's knowledge of different types.

3 Methodlogy

3.1 Experiment Search System

We implemented a search system using Indri from the Lemur toolkit (http://lemurproject.org) and used it to deliver results to a web search interface presented in Internet Explorer (IE6). The experiment was conducted using the multi-dimensional logging PooDLE experiment system [2].

3.2 Data Set

The data set was taken from the TREC Genomics collection, a ten-year, 4.5 million document subset of the MEDLINE bibliographic database [5]. We used the documents from the 2000-2004 period (n=1.85 million), to allow for reasonable retrieval efficiency.

Fig. 1. A search abstract

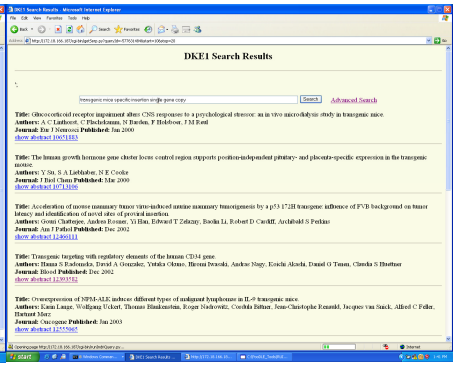


Fig. 2. A search results page

3.3 Subject Domain Based on MeSH Categories

As discussed earlier in the Introduction section, part of the reasons the TREC Genomics track data was used in this study was the knowledge in this field is systematic, well defined, and this knowledge is well represented by the MeSH thesaurus, a hierarchical concept structure that is widely used for processing documents in this field. However, genomics is a field that has a knowledge structure that is too big for this study to cover. Therefore, based on the MeSH structure for Genomics, after examining the amount of terms in each subfield and the amount of time this study could afford for each participant, three subfields were chosen as the domains in this study: Genetic Structure (G14), Genetic Process (G05), and Genetic Phenomena (G13). A pilot test was conducted to rate the terms in these three categories to make sure the amount of time needed for completing evaluating the terms for a participant would be within allowed limit.

3.4 Search Topics and Tasks

The TREC ad hoc retrieval tasks were based on 50 topics relating to five general types. Based on the selected domains from the MeSH categories, five of those tasks matching the three subfields, according to the topic description, were chosen by a hired expert for the user study. The selection criteria also included the number of relevant documents for the topic, easy or difficult to find relevant documents through searching by participants, and the balance of general and specific topics as well as easy and difficult tasks. Each topic has associated TREC expert graded relevance

judgments which provide a gold standard for the performance of the study participants. The tasks are listed in Table 1.

The experiment tasks asked participants to find and save all of the documents useful for answering the topic questions. The topics were presented unchanged from the TREC genomics track description.

Table 1. Search topics/tasks

TREC topic #	MeSH category	Topic title keywords
2	Genetic structure	Generating transgenic mice
7	Genetic processes	DNA repair and oxidative stress
42	Genetic phenomena	Genes altered by chromosome translocations
45	Genetic phenomena	Mental Health Wellness-1
49	Genetic structure	Glyphosate tolerance gene sequence

3.5 Participants

Forty students from the author's institution participated in the study. They were recruited from related schools and departments, such biology, pharmacy, animal science, biochemistry, and so on. Undergraduate students, graduate students and post-docs participated in the study. The number of graduate students and the number of undergraduate students were roughly balanced so that the different levels of knowledge could be elicited. Post-docs were assigned to the graduates group. Although there were 40 participants in the study, missing data issues and a task balance issue resulted in a data set with a different number of participants for each task. For three of the tasks there were 35 participants. One task had 20 participants and one had 15. Despite this unbalanced data set the nature of the tasks and the data analysis used minimized the impact of this less than ideal data set.

3.6 Knowledge Level Elicitation

For practical reasons, the level of participants' knowledge was assessed based on self-reporting methods. The level of BK was evaluated first. Before the experimental session started, each participant was asked to rate his/her understanding of each term in the categories (Genetic processes, genetic phenomena, and genetic structures) on a 5 point scale: 1 for “no knowledge”, 2 for “vague idea”, 3 for “some knowledge”, 4 for “high knowledge”, and 5 for “can explain to others.” These concepts may not directly relate to a particular search topic in the study, but since they are in the areas

from where the search topics were chosen, the ratings were able to represent the participants' level of this background knowledge.

For topic knowledge, after they had read the task description but before starting the task, participants were asked to rate their expertise and familiarity with the task on a 7-point scale, from 1: not at all to 7: most. The average of each participant's answers to the two questions was used as the measure of his/her level of topic knowledge. This knowledge is peculiar with this particular topic, not others.

3.7 Relevance Judgment and Search Performance Measures

Three performance measures were used in the study: precision, recall, and F measures. Since the tasks were recall-oriented tasks, our attention was focused on the recall measure and the F measure.

Participants were asked to conduct searches on the experimental system, to find as many as possible relevant documents to the topic. They could save documents and evaluate them after finishing their search activities. When evaluating, participants rated the relevance of the saved documents using a 5-point scale ranging from "not relevant" to "highly relevant" with "somewhat relevant" as the mid-point. The TREC assessors used relevance judgments of "somewhat relevant" and "highly relevant". Documents explicitly judged by TREC assessors as "not relevant" were unavailable to us, so we took every document in the TREC collection that was not explicitly judged by TREC assessors as relevant to the task to be not relevant. The participant judgments were mapped to the TREC scale with ratings of 4 or 5 as "highly relevant", and ratings of 2 or 3 as "somewhat relevant".

The performance of the participants was calculated for each task by checking for agreement with the TREC assessment. For this study we ignored the graded relevance levels in calculating precision. Recall was calculated by dividing the number of (correct) relevant documents by the number of TREC-assessed relevant documents for the task. The F score is the weighted harmonic mean of recall and precision. For our primary performance measure, we use the F2 score, which weights recall twice as much as precision, because our tasks were recall-oriented.

It should be noted that the MeSH index terms in documents were not available to the user when they viewed the search results or abstracts (Figures 1 and 2). The relevance criteria for participants were presumably based on their interaction with the available text.

3.8 Procedure

Participants were invited individually to the site of the experiment, an on-campus usability lab. They read and signed a consent form and filled out a questionnaire about their background, computer experience and previous searching experience. Next, they were asked to rate a paper version of the the selected MeSH terms that were related to the search topics. Following that was a brief demo using a training task about how to use the experiment system. Before each task the participants filled out a pre-task

questionnaire, in which topic knowledge was elicited. They were then given up to 15 minutes to conduct each of four assigned tasks. The tasks were rotated so that the order of the tasks would be different for different participants. The interaction between the participants and the system was logged by the PooDLE multi-dimensional logging system [2]. After completing each task, they completed a post-task questionnaire. After the participants finished all the tasks, they were asked to complete an exit questionnaire. Each participant was compensated $25 for their completion of the experiment. The experiment was conducted in a human-computer interaction lab, and each participant was tested individually.

4 Results

To examine the possible relationships of TK and BK with search performance, data analysis was first performed for each task, and then on the aggregated participant data over all tasks. These two levels represent a local search context and a general search context, in which different types of knowledge may have different roles.

Our results show that different knowledge levels and different types of domain knowledge had significant associations with search performance. We start by presenting the results for each of the 5 tasks and then present results for the aggregated data sets for all tasks to demonstrate the effects of the two type of knowledge.

4.1 Evidences of Differences between TK and BK

How consistent are participants in different types of self-reporting on their level of TK and BK? While a participant's BK value was the same for topics in the same MeSH category, for example, TREC topic 2 and 49, the person's TK value was unique for each individual task. The participants' average BK ratings for each category and TK ratings for each topic are presented in Table 2.

Table 2. BK and TK ratings for each topic

Topic/Task	MeSH Category	Min.	Max.	TK Mean (sd)	BK Mean (sd)
2 (n=35)	Genetic structure	1	5.5	3.23 (1.28)	1.72 (1.06)
7 (n=35)	Genetic processes	1	6	3.76 (1.38)	2.46 (0.91)
42 (n=15)	Genetic phenomena	1.5	5.5	3.67 (1.29)	2.06 (0.68)
45 (n=35)	Genetic phenomena	1	7	3.09 (1.35)	2.02 (0.83)
49 (n=20)	Genetic structure	1	4.5	2.58 (1.17)	1.89 (1.06)

A correlation analysis of the relationship between TK and BK for each task resulted in the data in Table 3 below:

Table 3. Correlations between TK and BK for each task

Topic/Task	MeSH Category	r	p
2 (n=35)	Genetic structure	0.215	0.214
7 (n=35)	Genetic processes	0.241	0.164
42 (n=15)	Genetic phenomena	0.471	0.076
45 (n=35)	Genetic phenomena	0.103	0.556
49 (n=20)	Genetic structure	-0.007	0.974

As can be seen in Table 3, for all individual tasks, there are no significant correlations between TK and BK, which demonstrates that the two are different sets of ratings.

The two sets of ratings were compared first to see if they were identical or similar. To make the two representations comparable, the TP ratings were transformed from 7-point scale to 5-point scale: ratings 2 and 3 were transformed into 2, 4 into 3, 5 and 6 into 4, and 7 into 5. 1 remained the same. The comparison of the two sets of ratings is displayed in Table 4.

Table 4. Comparison of TK and BK (N=140)

	TK	BK
Mean(standard deviation):	2.507 (0.857)	2.04 (0.960)
Pearson Correlation	r=0.209, p=0.013	
Sig. Difference (t-test)	t=-4.803, df=139, p=000.	

Table 4 shows the two sets of knowledge ratings are correlated but significantly different. It provides empirical evidence that the two are distinct types of knowledge to the participants, although closely related.

We used Pearson Correlation analysis procedure to analyze the relationships between knowledge and search performance measures. All analyses were run using 2-tailed test. The results are presented in the following sections.

4.2 Impact of the Two Types of Knowledge on Search Performance

The two different types of knowledge had different effects on search performance. The effects are analyzed at both the individual task level and the aggregated level.

4.2.1 Individual Tasks

Pearson Correlation analysis was conducted on the relation between knowledge and search performance measures for each search task (Table 5).

The significant correlations are highlighted in the table. Among the 5 tasks, 3 of the tasks were performed by 35 participants (n=35), one by 15 participants (n=15), and another one by 20 (n=20). For the 3 (n=35) tasks, topic knowledge had significant correlations with recall and the F2 score (note that all tasks were recall tasks) for 2 of them. For the other two (n <= 20) tasks, the correlations between topic knowledge and recall and F2 were close to significant at p=.05 level.

On the other hand, background knowledge had no significant correlations with any of the performance measures for any task. These results indicate that topic knowledge alone had an impact on performance in particular searches.

We then conducted MANOVA analysis on each topic data to detect the effects of the two types of knowledge, as well as the interaction of the two, on search performance. The results were in general similar to the results of correlation analysis: For topics 42 and 45, there was no significant effect on any of the performance measures by either BK or TK, or both (interaction effect of the two). For Topic 2, there was also no significant effect by BK or TK, but the interaction effect of the two was significant on precision at p=0.005. The result on Topic 7 was similar: no significant effect in general, but the interaction effect of the two was significant on precision at p=0.000.

Table 5. Correlations between knowledge and search performance measures for each individual task

Task ID	Knowledge	Precision	Recall	F2score
2	TK	r(35)=-0.204, p=0.241	r(35)=0.336**, p=0.049**	r(35)=0.328, **p=0.054**
	BK	r(35)=-0.119, p=0.497	r(35)=0.300, p=0.080	r(35)=0.304, p=0.076
7	TK	r(35)=-0.084, p=0.633	r(35)=348, **p=0.041**	r(35)=0.342**, p=0.044**
	BK	r(35)=0.156, p=0.372	r(35)=0.152, p=0.383	r(35)=0.166, p=0.342
42	TK	r(15)=0.100, p=0.722	r(15)=0.486, p=0.066	r(15)=0.486, p=0.066
	BK	r(15)=-0.085, p=0.763	r(15)=0.450, p=0.092	r(15)=0.448, p=0.094
45	TK	r(35)=-0.008, p=0.962	r(35)=--0.143, p=0.412	r(35)=--0.140, p=0.421
	BK	r(35)=0.313, p=0.0682	r(35)=0.171, p=0.325	r(35)=0.174, p=0.319
49	TK	r(20)=-0.402, p=0.079	r(20)=-0.424, **p=0.062**	r(20)=-0.425, **p=0.062**
	BK	r(20)=-0.399, p=0.081	r(20)=-0.301, p=0.197	r(20)=-0.304, p=0.192

For Topic 49, TK's effects were not significant but close: for precision: p=0.062, recall: p=0.058, and F2: p=0.057. However, the interaction effects of the two were significant on all three performance measures: precision at p=0.000, recall at p=0.003, and F2 score at p=0.002.

4.2.2 Overall – All Tasks

When the analysis was conducted on the data aggregated over all tasks, the results show that background knowledge had significant correlations with all three performance measures (Table 6).

Table 6. Correlations between knowledge and search performance over all tasks

Knowledge	Precision	Recall	F2score
TK	r(140)=-.059, p=.490	r(140)=.099, p=.246	r(140)=.091, p=.283
BK	r(140)=.223, **p=.008**	r(140)=.263, **p=.002**	r(140)=.268, **p=.001**

In contrast to the individual task case, where the topic knowledge had significant correlations with search performance measures, we found topic knowledge had no significant correlations with the performance measures for the aggregated data. Table 4 shows background knowledge is significantly correlated with search performance measures. The higher the ratings, which means more background knowledge, the better performance in terms of the precision, recall and F measures.

MANOVA results found that BK had significant effect on all performance measures: precision (p=.013); recall (p=.028), and F2score (p=.020). TK had no significant effect on any of the measures. However, the interaction of the two types of knowledge also had significant effects, and the effects were more significant than the BK itself at: precision (p=.000); recall (p=.009), and F2score (p=.006).

5 Discussion and Conclusions

There are several interesting findings from this study. The first is that the two different type of knowledge had different impact in terms of search performance: the long term searching success or the short term, current searching task. The impact of domain knowledge in general on search performance was found in previous studies [3,4,6,9,12,13,14]. Our study further identified the role of TK for the current search task, but not BK. This result makes sense because TK is the knowledge directly needed by the user for that particular topic. The second interesting finding is that despite no significant correlations between the BK and search performance measures for any individual task, BK is not useless for searching. Over multiple tasks, when the TP role seemed to fade, the BK became a significant factor in search success.

The disappearance of the significance of TK for multiple tasks is reasonable: TK is related only with a particular task. Because the topics are different from task to task, it is unreasonable to expect the same topic knowledge would work for other tasks. Our results showed that TK had significant correlations with search performance measures only for individual tasks. Whenever the analysis was conducted on multiple task data sets, there were no significant correlations with TK.

The appearance of significant correlations with BK on multiple-task data sets indicates that although BK did not show significant correlations for individual tasks, it may still have some impact. One can hypothesize that over multiple tasks the effect accumulated to the degree that its impact on search performance was detectable.

Another finding is that despite BK and TP exhibited different effects, but there were specific situations where it seemed to be difficult to differentiate the two: only the interaction of the two would have a significant effect. Further studies on such cases are needed in the future.

It should be noted that even though knowledge, either TK knowledge or BK, may not have significant direct correlations with search performance measures, it does not mean the knowledge had no impact. Knowledge can also be significantly correlated with search behaviors, which in turn may have impact on search performance. In our work so far we have analyzed knowledge effects on search behaviors and have found many significant correlations. Since task characteristics have often been found to be an important factor in search performance, we further analyzed the data by certain task characteristics (easy/hard, general/specific) to distinguish knowledge effects on different types of tasks. Limited by the space of this paper, those results will be reported in another paper.

Our primary goal in this study was to better understand the different roles of distinct types of domain knowledge, the user's TK and BK, on search performance. Ultimately, we wanted to infer automatically aspects of the user knowledge based on observations of their interactions with search systems. This would enable the search system to adapt to the user by personalizing system behavior, including search results, query reformulation, or user interface components, to a particular type of knowledge. Further research in this direction will be conducted in the future.

Acknowledgments. This research was supported by IMLS grant LG 06-07-0105-07. We thank all of the members of the Personalization of the Digital Library Experience (PoODLE) research team at Rutgers University, without whose efforts this work could not have been accomplished.

References

[1] Allen, B.L.: Topic knowledge and online catalog search formulation. Library Quarterly 61(2), 188–213 (1991)

[2] Bierig, R., Gwizdka, J., Cole, M.: A user-centered experiment and logging framework for interactive information retrieval. In: Belkin, N.J., Bierig, R., Buscher, G., van Elst, L., Gwizdka, J., Jose, J., Teevan, J. (eds.) Proceedings of the SIGIR 2009 Workshop on Understanding the User - Logging and Interpreting User Interactions in IR (2009)

[3] Duggan, G.B., Payne, S.J.: Knowledge in the head and on the web: Using topic expertise to aid search. In: Proc. SIGCHI, pp. 39–48 (2008)
[4] Hembrooke, H.A., Gay, G.K., Granka, L.A.: The effects of expertise and feedback on search term selection and subsequent learning. JASIS&T 56(8), 861–871 (2005)
[5] Hersh, W., Voorhees, E.: TREC genomics special issue overview. Information Retrieval 12(1), 1–15 (2009)
[6] Hsieh-Yee, I.: Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. Journal of the American Society for Information Science 44(3), 161–174 (1993)
[7] Kelly, D., Cool, C.: The effects of topic familiarity on information search behavior. In: JCDL 2002, pp. 74–75 (2002)
[8] Liu, J., Belkin, N.J.: Personalizing information retrieval for people with different levels of topic knowledge. In: JCDL 2010 (2010)
[9] Roberts, P.M., Cohen, A.M., Hersh, W.R.: Tasks, topics and relevance judging for the TREC genomics track: Five years of experience evaluating biomedical text information retrieval systems. Information Retrieval 12(1), 81–97 (2009)
[10] Sihvonen, A., Vakkari, P.: Subject knowledge improves interactive query expansion assisted by a thesaurus. Journal of Documentation 60(6), 673–690 (2004)
[11] Vakkari, P., Pennanen, M., Serola, S.: Changes of search terms and tactics while writing a research proposal: A longitudinal research. Information Processing & Management 39(3), 445–463 (2003)
[12] White, R., Dumais, S.T., Teevan, J.: Characterizing the influence of domain expertise on Web search behavior. In: Proceedings of WSDM 2009 (2009)
[13] Wildemuth, B.: The effects of domain knowledge on search tactic formulation. JASIS&T 55(3), 246–258 (2004)
[14] Zhang, X., Anghelescu, H.G.B., Yuan, X.: Domain knowledge, search behavior, and search effectiveness of engineering and science students. Information Research 10(2), 21 (2005)

Community Detection by Local Influence

Nuno Cravino and Álvaro Figueira

CRACS/INESC TEC, Faculdade de Ciências, Universidade do Porto
{nuno.cravino,arf}@dcc.fc.up.pt

Abstract. We present a new algorithm to discover overlapping communities in networks with a scale free structure. This algorithm is based on a node evaluation function that scores the local influence of a node based on its degree and neighbourhood, allowing for the identification of hubs within a network. Using this function we are able to identify communities, and also to attribute meaningful titles to the communities that are discovered. Our novel methodology is assessed using LFR benchmark for networks with overlapping community structure and the generalized normalized mutual information (NMI) measure. We show that the evaluation function described is able to detect influential nodes in a network, and also that it is possible to build a well performing community detection algorithm based on this function.

Keywords: graph theory, link analysis, centrality, community detection, overlapping communities.

1 Introduction

Nature and human derived complex networks follow certain patterns in their structure and development[2]. Social networks, computer networks, protein interaction networks, among others, tend to follow fat tailed distributions of node degree at least asymptotically. Many of these networks also display a community structure revealed by the existence of groups with highly interconnected nodes, with low connectivity to other groups. In most cases maintaining a low average path length between any node in the network, whether or not in the same group. Community detection algorithms are an attempt to retrieve these groups. The most currently used algorithms retrieve disjunct clusters from the networks. Recent developments have shown that many networks display overlapping clusters or an hierarchical disposition of clusters[13,10,8,11], which creates a pressing need for newer techniques that should be not only able to retrieve these communities, but that should also be capable of doing so in the large scale networks.

1.1 Previous Developments in Overlapping Community Detection

The Clique Percolation Method (CPM)[10] is a popular method for overlapping community detection which assumes that communities arise as densely connected

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 193–200.
DOI: 10.1007/978-3-642-36981-0_18

sub-graphs. The search for communities is done by the identification of all cliques of a certain size. After this step it generates a new graph containing all identified cliques as nodes that are considered adjacent if they share most of their elements (clique size minus one). The communities are then found by retrieving the connected components in this final graph. Speaker-listener Label Propagation Algorithm (SLPA)[12] is another algorithm that uses a variant of the label propagation algorithm to construct the communities. In this algorithm labels are shared between neighbour nodes, and afterwards in the distribution of these labels is processed in order to retrieve the communities.

Our approach differs from these algorithms by using a local evaluation function that discovers the structure of the network around a certain node, and uses this information to guide the search to a local maxima of the function. Clusters are then formed by nodes around the local maxima. Our algorithm is relatively stable since its results will only be affected by the order of evaluation, and if the order remains the same the results will be the same.

2 Community Detection by Local Influence

This research was developed in order to retrieve socially relevant information from a tag co-occurrence network, to be used to socially influence[4] the classification of documents in a news related social network[1]. Stability and performance were paramount to this research in order to maintain user acceptable results and delays.

2.1 The Local Influence Score

We developed a new local scoring function, the local influence, in order to retrieve hub candidates from the neighbourhood of a node. This function is built upon the properties of networks having a community structure, and of scale free networks, and is a local measure of the influence of a node in a network. Informally the local influence of a node can be defined as a score that measures the importance of a given node to the overall structure of a network. Nodes that poorly affect the structural properties of the network will have a low score, while high scoring nodes will have a significant impact. Higher influence nodes not only connect to most nodes within a certain range (i.e. in the same community), but they also provide connectivity to other sets of well connected nodes farther within the network, by forming bridges or providing increased connectivity to nodes that do so. Due to this, their removal would increase the average shortest path length between nodes within the same community, and also decrease connectivity to other communities in the network.

Scale free networks[3] have hubs which are high degree nodes that connect a large set of nodes, where the removal of just one of them can result in a significant increase in network diameter. The overall influence of these nodes over the structure of the network is high, since they ensure the low average shortest path between all nodes in the network when compared to random graphs. Based

on this structural importance of hubs in scale free networks, a node influence on the network structure needs, by our informal definition, to be proportional to its degree.

Communities[5], in the context of community detection, are loosely defined as groups of highly interconnected nodes, with significantly lower connectivity to other communities. Given this definition we add to our previous hypothesis that a node influence is higher if it connects with other high degree nodes, as would be the case of nodes within communities, and even more if they also connect to other communities.

Finally, we add that in a weighted network the neighbours importance is proportional to the weight of the edge that connects with them. Based on this set of hypothesis we construct the local influence scoring function as

$$score(n) = degree(n) \times \sum_{v \sim n} (degree(v) \times weight(n, v)) \tag{1}$$

where n is the node under evaluation, v a neighbour node of n, and $\sim$ the adjacency relation of the network. This function can be used to score nodes proportionally to the chance of being a hub, where its value will decrease for lesser connected nodes, and increase for nodes that follow our informal definition of local influence. This measure can also be viewed as an extension of the notion of degree centrality, taking into account the neighbourhood of the node being scored. The Hyperlink-Induced Topic Search (HITS)[6] algorithm also enables the identification of network hubs, using iterative improvement to compute its authority and hub scores. Unlike the local influence function here described, HITS takes a global approach, and therefore presents a higher complexity.

In figure 1 we show the nodes' local influence on the Zachary's Karate Club network[14]. It is visible that the hubs are the two nodes that originated the split, which consequentially have an high local influence score using our metric. There are other nodes that can be identified as influential given their high degree centrality and connections with other influential nodes.

3 The Community Detection Algorithm

In algorithm 1 we present the pseudo code for our algorithm. We have excluded the initialization and post-processing phases for conciseness. The initialization phase is simply the construction of a table of all node scores in order to be possible to access them in constant time.

3.1 The Community Detection Algorithm

The detection phase uses a guided search strategy to find a single path from each node to the nearest local hub, storing the set of the evaluated nodes. This path length is limited by an approximation of the average path length, since for nodes within the same community the path length will be less than the average. We use $\lfloor \ln(\ln(|N|)) \rfloor$ as the approximation, building upon the result by Cohen *et al.*

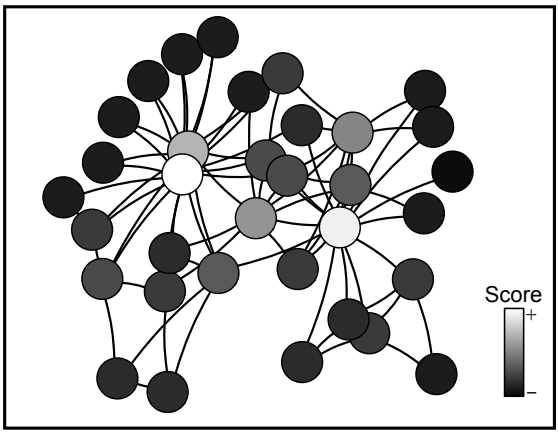


Fig. 1. Zachary Karate Club network local influence scores

Algorithm 1. Community Detection by Local Influence

```
1: function FindCommunities(VertexList)
2:     Counts ← (empty sparse matrix)
3:     maxHop ← ⌊ln(ln(|VertexList|))⌋
4:     for all candidate ∈ VertexList do
5:         maxScore ← −∞
6:         hub ← nil
7:         S ← ∅
8:         while (candidate ≠ hub) ∧ (hop < maxHop) do
9:             hub ← candidate
10:            hop ← hop + 1
11:            for all v ∈ hub.Adjacency do
12:                if (score(v) × weight(hub, v)) > maxScore then
13:                    maxScore ← score(v)
14:                    candidate ← v
15:                S ← S ∪ v
16:        for all u ∈ S do
17:            Counts[hub][u] ← Counts[hub][u] + 1
18:    return process(Counts)
```

in [3], about the order of the distance in scale free networks. This limit enables us to reduce the overall complexity of the algorithm, given the slow growth of $\ln(\ln(|N|))$.

The communities are formed by processing the node occurrence counts in the paths leading to each hub. Given $Counts[h]$ the count vector of all nodes found in the paths to hub h, we retrieve $Com[h]$, the community formed around h, as follows

$$Com[h] = \{v \in Counts[h] : Counts[h][v] \geq \frac{\sum_{u \in Counts[h]}(Counts[h][u])}{|Counts[h]|}\} \quad (2)$$

being that v and u are nodes. By joining together only those nodes with a significant occurrence count, equal to or greater than the arithmetic mean, we are able to reduce the presence of non related nodes in the same cluster.

The retrieved communities can be optionally post-processed to join similar communities and attribute titles. We cycle through all clusters and use the top n scoring nodes in each cluster to assign a title, merging all clusters sharing the exact same title. This allows us to reduce the number of clusters with very similar node sets without the need for costly comparisons. Also in word based networks we are able to obtain meaningful titles that allow to better identify the underlying context of the cluster.

3.2 Complexity

Since the scores are not updated during the detection phase, we can generate the table of all scores prior to that phase so we can have constant access time to the scores. In order to do this all the neighbours for each node must be processed, which takes a total of $\sum_{n \in N} degree(n)$ steps. This value can be simplified to a bound of $|N|d$ where d is the ceiling of the average node degree. Also, given that the node degree distribution follows a power law for scale free networks, it holds that the average node degree is significantly smaller than the number of nodes. For the detection phase we will once again have to process all the neighbours for each node, but in addition we have a cycle with at worst $\ln(\ln(|N|))$ time complexity. The complexity for processing nodes was shown before to be $|N|d$, but with the additional cycle it becomes $|N|d\ln(\ln(|N|))$. The count processing operation, after the inner loops, takes time linear on the size of the smallest set, so its complexity is bounded by the complexity of the inner loops. This is because at most the number of nodes to add to the count will be the same as the number of steps in the inner loops, therefore it only increases the complexity by a constant factor, and so it can be discarded. Therefore the overall complexity upper bound for the detection phase will be $|N|d\ln(\ln(|N|))$, being below quadratic in relation to $|N|$ for most non complete graphs, and well below that value for scale free networks.

In the count processing phase we process for each cluster i all of its nodes $|C_i|$. Since the mean calculation for each cluster only increases complexity by a constant factor, the complexity for this phase is bounded by $\sum_{i \in clusters} |C_i|$, the sum of the sizes of all clusters. This sum cannot ever exceed the bound of the detection phase, since all the cluster elements are retrieved in that phase, therefore this sum of the cardinalities of the clusters will be at most the same as the number of steps of the detection phase. Thus the complexity for the processing phase is the same as that of the detection phase.

The overall complexity of the algorithm is bounded by $|N|d+2|N|d\ln(\ln(|N|))$, that can be simplified to a time complexity bound of $|N|d\ln(\ln(|N|))$. In practice the average shortest path of most real complex networks will remain mostly

constant and it can also be assumed that the degree will always be significantly smaller than the number of nodes in the network. Therefore, this algorithm presents near linear complexity over the number of nodes, and edges, in these specific networks.

The algorithm we present is also inherently parallelizable, since the computation of the path to the hub of a node needs not to update any information relevant to any computation for any other node. Therefore, with little modification, these paths can theoretically be computed concurrently, followed by the processing of clusters.

4 Evaluation

4.1 Benchmark and Evaluation Metric

In order to evaluate the our algorithm we use the LFR model, by Lancichinetti *et al.* l[9,7], for weighted undirected graphs with overlapping structure to generate the graphs with a ground truth. LFR has various parameters, but we allow variation in only a few. The number of nodes ($|N|$) is set to $\{5000, 50000\}$ in order to test the algorithm in different scale graphs. The number of overlapping nodes (O_n) was set to $\{1\%|N|, 10\%|N|\}$ in order to evaluate the algorithm in networks with different overlapping structural tendencies. The maximum membership size for overlapping nodes (O_m) for $\{2, 6\}$ in order to assess the response of the algorithm to higher overlaps. The mixing parameter for topology (μ_t) allows, for higher values, changes on the structure of the network in order to simulate more dense networks with less defined community structures and was set to $\{0.1, 0.2, 0.3\}$. The rest of the parameters are set as Xie *et al.*[11] in the case of small communities, with $\mu_w = \mu_t$, community sizes between $[20, 50]$, average degree of 10 and maximum degree of 50. We used the generalized NMI[8] for overlapping clusters to evaluate the results obtained.

4.2 Results

Our results follow the general trend shown by others in previous comparisons[11] using the same benchmark, though showing a particular sensitivity to changes in the topology of the network and increases in the total number of overlapping nodes. This may be due to the restrictions imposed onto the path finding process, that may not be entirely compatible with the structural changes resulting from changes in μ. Overall, and given its low complexity, it performs well on networks with a well defined community structure as shown on tables 1 and 2.

Table 1. The results of our tests for $|N| = 5000$

Om	2			6		
μ_t	0.1	0.2	0.3	0.1	0.2	0.3
NMI(1%On)	0.7534	0.6148	0.4476	0.7315	0.5818	0.4260
NMI(10%On)	0.6170	0.5035	0.3335	0.4473	0.3566	0.2333

Table 2. The results of our tests for $|N| = 50000$

Om	2			6		
μ_t	0.1	0.2	0.3	0.1	0.2	0.3
NMI(1%On)	0.7734	0.6030	0.4249	0.7345	0.5863	0.4087
NMI(10%On)	0.6242	0.4821	0.3455	0.4528	0.3656	0.2728

5 Conclusion

We presented a new technique for community detection, performing as well as some other techniques previously reported while remaining mostly deterministic and presenting a good complexity profile. These characteristics are indispensable in order to provide users with a real time response and a consistent experience, necessary features of most real time non-technical systems. The scoring function here presented can be used independently from the algorithm, as a measure of influence, in order to analyse the network and identify potential hubs. It can also be used to assign meaningful titles to groups or clusters of nodes based on their scores. This algorithm can be further enhanced with optimized cluster generation, to better take into account the distribution nodes into communities.

Acknowledgements. This work is financed by National Funds through the FCT Fundao para a Cincia e a Tecnologia (Portuguese Foundation for Science and Technology) within project PEst-C/EEI/LA0014/2011.

References

1. Figueira, Á., et al.: Breadcrumbs: A social network based on the relations established by collections of fragments taken from online news, `http://breadcrumbs.up.pt` (retrieved January 19, 2012)
2. Barabási, A.-L.: Scale-free networks: a decade and beyond. Science 325(5939), 412–413 (2009)
3. Cohen, R., Havlin, S., Ben-Avraham, D.: Structural properties of scale free networks (2002)
4. Cravino, N., Devezas, J., Figueira, Á.: Using the Overlapping Community Structure of a Network of Tags to Improve Text Clustering. In: Proceedings of the 23rd ACM Conference on Hypertext and Social Media HT 2012 (2012)
5. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America 99(12), 7821–7826 (2002)
6. Kleinberg, J.M.: Hubs, authorities, and communities. ACM Computing Surveys 31(4es), 5–es (1999)
7. Lancichinetti, A., Fortunato, S.: Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. Physical Review E 80(1), 9 (2009)
8. Lancichinetti, A., Fortunato, S., Kertész, J.: Detecting the overlapping and hierarchical community structure in complex networks. New Journal of Physics 11(3), 033015 (2009)

9. Lancichinetti, A., Fortunato, S., Radicchi, F.: Benchmark graphs for testing community detection algorithms. Physical Review E - Statistical, Nonlinear and Soft Matter Physics 78(4 Pt. 2), 6 (2008)
10. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. Nature 435(7043), 814–818 (2005)
11. Xie, J., Kelley, S., Szymanski, B.K.: Overlapping Community Detection in Networks: the State of the Art and Comparative Study. Arxiv preprint arXiv11105813, V, pp. 1–30 (November 2011)
12. Xie, J., Szymanski, B.K., Liu, X.: SLPA: Uncovering Overlapping Communities in Social Networks via a Speaker-Listener Interaction Dynamic Process (2011)
13. Yang, J., Leskovec, J.: Structure and Overlaps of Communities in Networks. Arxiv preprint arXiv12056228 (2012)
14. Zachary, W.W.: An information flow model for conflict and fission in small groups. Journal of Anthropological Research 33(4), 452–473 (1977)

Predict Sepsis Level in Intensive Medicine – Data Mining Approach

João M.C. Gonçalves[1], Filipe Portela[1], Manuel Filipe Santos[1], Álvaro Silva[2], José Machado[3], and António Abelha[3]

[1] Algoritmi Centre,
University of Minho, Guimarães, Portugal
jgoncalves@di.uminho.pt,
{cfp,mfs}@dsi.uminho.pt
[2] Serviço Cuidados Intensivos, Centro Hospitalar do Porto, Portugal
moreirasilva@clix.pt,
fernandorua.sci@hgsa.min-saude.pt
[3] CCTC, University of Minho, Braga, Portugal
{jmac,abelha}@di.uminho.pt

Abstract. This paper aims to support doctor's decision-making on predicting the Sepsis level. Thus, a set of Data Mining (DM) models were developed using prevision techniques and classification models. These models enable a better doctor's decision having into account the Sepsis level of the patient. The DM models use real data collected from the Intensive Care Unit of the Santo António Hospital, in Oporto, Portugal. Classification DM models were considered to predict sepsis level in a supervised learning approach. The models were induced making use of the following algorithms: Decision Trees, Support Vector Machines and Naïve Bayes classifier. The models were assessed using the Confusion Matrix, associated metrics, and Cross-validation. The analysis of the total error rate, sensitivity, specificity and accuracy were the metrics used to identify the most relevant measures to predict sepsis level. This work demonstrates that it is possible to predict with great accuracy the sepsis level.

Keywords: Data mining, Classification, Intensive Care, Sepsis, INTCare.

1 Introduction

Everyday new patients come into Intensive Care Units (ICU) in a critical health condition. It is very difficult for the professionals to care of the patients and, simultaneously, document the operations [1, 2]. In order to overcome this limitation a project was developed called INTCare [3-5]. This project has the objective to make available anywhere and anytime [6, 7] pertinent information about the patient. During the project a lot of data (vital signs, laboratory results, fluid balance, ventilation, ICU scores and other) were converted into electronic mode, enabling the automatic acquisition in real-time. This new reality allows for obtaining fundamental knowledge to the patient treatment in the right time.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 201–211.
DOI: 10.1007/978-3-642-36981-0_19

The objective of this project is to predict the patient sepsis value in real-time, determining whether the patient is in the second or third level of the scale. This work was developed in a real environment using real data obtained from the Intensive Care Unit of Centro Hospitalar do Porto, Hospital Santo António, Portugal. This paper is divided into six chapters. After this introduction, the background and related work are presented in the second chapter. The third chapter makes an overview of the data used by the Data Mining Models. Then, the fourth chapter introduces the data mining models developed and variables used. The obtained results are presented in the chapter five. Finally, some conclusions about the work are written and the future work presented.

2 Background

2.1 SEPSIS

Sepsis is classified as a severe general infection and it is hard to define, diagnose and treat [14]. It is related to a large number of clinical conditions caused by a systemic inflammatory response of the organism to an infection, which develops in severe sepsis, also combined with simple, multiple or total organ dysfunction/failure leading to death [14]. It is one of the major causes of death in Intensive Care Units. Daily is killing around 1,400 people worldwide [15]. Although not having a clear clinical definition that can be easily adopted and communicated in its entirety, its absence turns sepsis diagnosis and treatment into a clinical challenge [14].

Tables 1, 2 and 3 show the variables associated to sepsis levels. These variables are essential to the development of classification models.

Table 1. Definition of Sepsis, adapted from [19-21]

Variables	**Sepsis is defined as a documented or suspected infection with one or more of the following**
General	Fever (core temperature >38.3°C)
	Hypothermia (core temperature <36°C)
	Heart rate >90 min–1 or >2 SD above the normal value for age
	Tachypnea
	Altered mental status
	Significant edema or positive fluid balance (>20 mL/kg over 24 hrs)
	Hyperglycemia (plasma glucose >120 mg/dL) in the absence of diabetes
Inflammatory	Leukocytosis (WBC count >12,000 μL–1)
	Leukopenia (WBC count <4000 μL–1)
	Normal WBC count with >10% immature forms
	Plasma C-reactive protein >2 SD above the normal value
	Plasma procalcitonin >2 SD above the normal value

Table 2. Definition of Severe Sepsis, adapted from [19-21]

Variables	Severe sepsis is defined as sepsis associated with organ dysfunction, hypoperfusion, or hypotension
Organ dysfunction	Arterial hypoxemia (PaO2/FIO2 <300) Acute oliguria (urine output <0.5 mL·kg-1·hr-1 or 45 mmol/L for 2 hrs) Creatinine > 2.0 mg/dL Coagulation abnormalities (INR >1.5 or aPTT >60 secs) Thrombocytopenia (platelet count <100,000 ìL–1) Hyperbilirubinemia (plasma total bilirubin > 2.0 mg/dL or 35 mmol/L)
Tissue perfusion	Hyperlactatemia (>2 mmol/L)
Hemodynamic	Arterial hypotension (SBP <90 mm Hg, MAP <65 mm Hg, or SBP decrease >40 mm Hg)

Table 3. Definition of Septic Shock, adapted from [19-21]

Septic shock is defined as acute circulatory failure unexplained by other causes
Acute circulatory failure is defined as persistent arterial hypotension (SBP <90 mmHg, MAP <60, or a reduction in SBP >40 mm Hg from baseline despite adequate volume resuscitation).

2.2 Data Mining

Due the nature of the goals, this encompasses a typical problem of Classification [8]. The main objective is to predict one target with two classes, i.e., classify if a set of data represents a patient with or without sepsis.

To overcome this problem a set of models were defined using three distinct techniques: Support Vector Machine (SVM), decision trees) (DT) and Naïve Bayes (NB). The models were induced automatically using oracle Data Mining [9, 10]. All the data used were obtained and processed automatically.

2.3 Related Work

The objective of this work only was possible achieve due to the previous achievements in the research project INTCare. The main goal of INTCare was the development of an Intelligent Decision Support System to predict organ failure and patient outcome in real-time and using online learning [11, 12]. To attain INTCare goals was necessary to perform a set of changes in the ICU environment [13] and in the way that the data were collected. INTCare introduced changes in the data sources type and format. Now, the laboratory results are collected in an open format, the patient therapeutic are accessible electronically, the vital signs are obtained automatically and a new platform to collect the nursing data in real-time called Electronic Nursing Record (ENR) was developed. ENR is used for monitoring the patient data in real-time. ENR is a web-based touch screen platform and it is available

near the patient beds. This platform is used by nurses and by physicians to insert, validate or consult the patient data. All the data are available anywhere and anytime. The new knowledge obtained by INTCare it is disseminated through ICU platforms. These advances in the ICU technology enabled the prediction of the sepsis level.

3 Data-Driven

The Data used in DM models were collected at the Intensive Care Unit (ICU) of Centro Hospitalar do Porto from 19.08.2011 to 04.07.2012, corresponding to 305 days, 394 patients and 12 beds. The data has been collected from:

- 12 monitors vital signs;
- 10 mechanical ventilators;
- Laboratory examinations.

The data sets provided by HSA for this project were generated from a set of real data collected and processed online in real time, using the following data sources:

- Electronic Health Record (EHR);
- Vital Signs Monitor (VSM);
- Laboratory (LAB).

Table 4 presents the associated attributes, their description and respective data source, referring to the initial selection.

Table 4. Attributes, description and data source

Attributes	Description	Data source
BILIRUBIN	Bilirubin	LAB
CREATININE	Creatinine	LAB
HR	Heart Rate	VSM
GLUCOSE	Glucose	LAB
TIME	Time when the clinical examination was collected	LAB/VSM
LEUKOCYTES	Leukocytes	LAB
PID	Patient Identifier	EHR/VSM /LAB
DR	Date recording/collection value	EHR/VSM /LAB
MAP	Mean Arterial Pressure	VSM
SBP	Systolic Blood Pressure	VSM
PLATELETS	Platelets	LAB
TEMP	Temperature	VSM

Table 5 presents the data obtained after the transformation, some new attributes and records were added. Excluding SEPSIS_FINAL variable, all the others have assigned the value 0 for the data that is within the range and the value 1 for those who are out of it.

Table 5. Data obtained after transformation

Attributes	Data
SEPSIS_BILIRUBIN	{0;1}
SEPSIS_CREATININE	{0;1}
SEPSIS_HR	{0;1}
SEPSIS_GLUCOSE	{0;1}
SEPSIS_LEUKOCYTES	{0;1}
SEPSIS_MAP	{0;1}
SEPSIS_SBP	{0;1}
SEPSIS_PLATELETS	{0;1}
SEPSIS_TEMP	{0;1}
SEPSIS_FINAL	{0;2;3}

Regarding to the SEPSIS_FINAL variable, the value 0 represents patients without sepsis, the value 2 patients with severe sepsis and the value 3 patients with septic shock. Using the value 1 of the remaining attributes present on table 10 and according to the definitions of sepsis on the tables 1, 2 and 3, a procedure was developed to calculate the sepsis level for each record. As such, the procedure assigns the values of 0, 2, or 3 through a simple verification as defined by sepsis in the referenced tables (1, 2, and 3). In the variable SEPSIS_FINAL was not considered the patient's condition with sepsis (referring to variables of Table 1) as this would have to be rated with a value of 1. For this reason it was not possible to assign the value 1 to the variable SEPSIS_FINAL because, according to ICU doctors, these values are easily confused with patient's condition without sepsis. During the data integration process, in order to merge these data it was created a data view for the entire data, combining the attributes of table 4 and those transformed on Table 5.

Finally, it was necessary to convert continuous numeric data into a range of classes. The values ranges were created using a 7-point scale adapted from the Clinical Global Impression - Severity scale (CGI-S)[16]. The CGI-S allows the doctors to evaluate the disease severity [17]. In this sense, Bin Quantile Range grouping technique was used considering seven classes [16]. The classes were created using a quantile distribution of the values received by each variable.

4 Data Mining Models

In order to obtain models from the data to classify sepsis level, it was essential to create DM tasks to obtain new knowledge. After the data preparation and data processing tasks, a modeling phase has been carried out. Figure 1 shows the data transformations of the classification models for sepsis and the figure 2 shows the sepsis classification models.

Fig. 1. DM process to the Sepsis target

Once variables correspond to continuous and discrete values, the classification models were used distinguishing the independent variables (lab results and vital signs) and the dependent variables (the score for sepsis). For the models modeling/evaluation 70% of the data was used for training and the remaining 30% for testing. During the modeling task some parameters adjustments were made. All continuous and discrete values with more than 10 classes, excepting the episode number and date, were classified using the Quantile Bin Range.

Each model (M1 to M4) was developed using three techniques (SVM, DT and NB) related to sepsis level condition. The input attributes (independent variables) are:

CaseMix = {PID; Data}
Sepsis = {Bilirubin; Creatinine; FCMax; FCMin; Glucose; Leukocytes; MAPMax; MAPMin; SBPMax; SBPMin; Platelets; TempMax; TempMin}
VarSepsis = {SepsisBilirubin; SepsisCreatinine; SepsisFC; SepsisTemp; SepsisGlucose; SepsisLeukocytes; SepsisPAM; SepsisPAS; SepsisPlatelets}

The target attribute (dependent variables) is:

Target = SepsisFinal

Each model can be represented as follows:

```
Model{M1}*{SVM;DT;NB}*{target} = {CaseMix; Sepsis}
Model{M2}*{SVM;DT;NB}*{target} = {CaseMix; VarSepsis}
Model{M4}*{SVM;DT;NB}*{target} = {CaseMix; Sepsis; VarSepsis}
Model{M4}*{SVM;DT;NB}*{target} = {Automatic}
```

For example, the model independent variables relating to M1 (casemix and sepsis), can be represented as follows:

```
M1 = SVM * {target} * {CaseMix; Sepsis}
M1 = DT * {target} * {CaseMix; Sepsis}
M1 = NB * {target} * {CaseMix; Sepsis}
```

It was generated and tested a total of 12 classification models (1 (target) * 4 (scenarios) * 3 (DM techniques)) to predict sepsis level.

5 Results

In order to assess the results for the models developed for sepsis a set of measures were applied. Figure 2 shows, ordered by average accuracy, the acuity values for each classification model and the technique used. All of the M4 models for each technique use variables defined automatically by the engine. The other models were characterized using a manual selection of the variables.

From the twelve scenarios considered relatively to the sepsis level, the top three non-automatic models (one of each technique) were selected for analysis: SEPSIS_SVM_M3, SEPSIS_DT_M3 SEPSIS_NB_M1 (Figure 2). Tables 6, 7 and 8 show the best prediction results for each set of variables selected, considering the technique and the corresponding scenario and the values for total error, sensitivity, specificity and accuracy.

Models

Name	Average Accuracy %	Algorithm
SEPSIS_SVM_M4	100	Support Vector Machine
SEPSIS_SVM_M3	100	Support Vector Machine
SEPSIS_SVM_M1	100	Support Vector Machine
SEPSIS_DT_M4	100	Decision Tree
SEPSIS_DT_M3	100	Decision Tree
SEPSIS_DT_M1	100	Decision Tree
SEPSIS_NB_M1	99,851	Naive Bayes
SEPSIS_SVM_M2	99,8084	Support Vector Machine
SEPSIS_NB_M3	99,8084	Naive Bayes
SEPSIS_DT_M2	99,8084	Decision Tree
SEPSIS_NB_M4	99,7233	Naive Bayes
SEPSIS_NB_M2	99,0209	Naive Bayes

Fig. 2. All the models developed for sepsis

Table 6. Confusion matrix model SVM_M3

Model SVM_M3	**Severe sepsis**	**Septic shock**	**Total**	**Corrects**
Severe sepsis	334	0	334	100%
Septic shock	0	1.415	1.415	100%
Total	334	1.415	**1.749**	
Corrects	100%	100%		

Total error	**Sensitivity**	**Specificity**	**Accuracy**
0%	100%	100%	100%

Table 7. Confusion matrix model DT_M3

Model DT_M3	**Severe sepsis**	**Septic shock**	**Total**	**Corrects**
Severe sepsis	334	0	334	100%
Septic shock	0	1.415	1.415	100%
Total	334	1.415	**1.749**	
Corrects	100%	100%		

Total error	**Sensitivity**	**Specificity**	**Accuracy**
0%	100%	100%	100%

Table 8. Confusion matrix model NB_M1

Model NB_M1	**Severe sepsis**	**Septic shock**	Total	Corrects
Severe sepsis	109	0	109	100%
Septic shock	1	436	437	99,77%
Total	110	436	**546**	
Corrects	99,09%	100%		

Total error	**Sensitivity**	**Specificity**	**Accuracy**
0,18%	100%	99,09%	99,82%

Through the confusion matrices, we conclude that the best predictions for sepsis level are provided by SVM_M3, DT_M3 and NB_M1 as presented on tables 6, 7 and 8 respectively.

From the results obtained for sepsis levels, the Figure 3 shows the ROC curve obtained from SEPSIS_DT_M3 and SEPSIS_SVM_M3 models, considered the best models in terms of acuity (100%). In figure 4 it is represented the decision tree obtained from the SEPSIS_DT_M3 model with 100% accuracy.

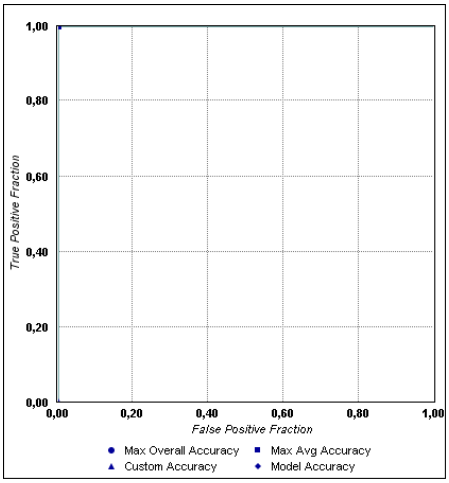


Fig. 3. ROC Curve

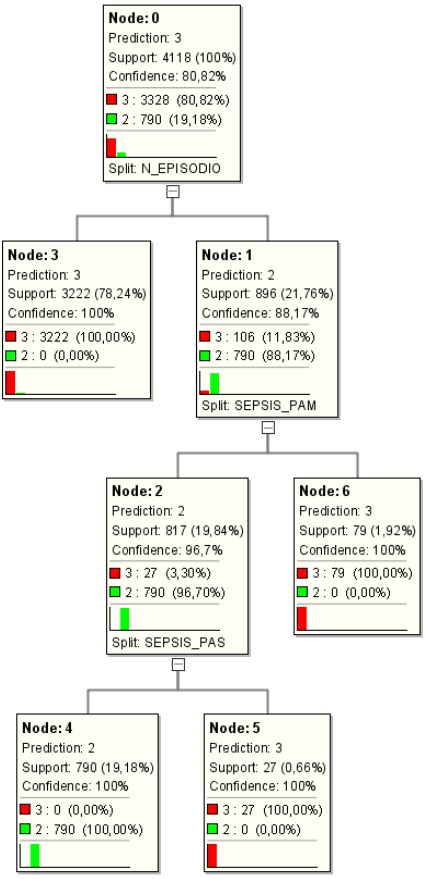


Fig. 4. Decision tree model DT_M3

In conclusion, it is possible to predict sepsis level with a great accuracy. Regarding sepsis level, the accuracy, sensitivity and specificity results were 100% for DT_M3 and SVM_M3 models. NB_M1 model presented 100% for sensitivity, 99.99% for specificity and 99.82% for accuracy.

6 Conclusions and Future Work

The results obtained in this work can be considered excellent and prove that it is possible to predict the Sepsis level in real-time using Data Mining.

This paper presents the classification models obtained using data collected in real time from the ICU of HSA, Porto, Portugal. It was considered a large initial data set which resulted, after processing in a total of 193 112 records, relating to episodes that occurred during 305 days and 394 patients.

BI systems allow for the combination of data collection with analysis tools, with the main objective of providing information for decision making. The DM techniques are needed to extract knowledge and are a part of BI systems. These techniques were very important for this work, especially by applying learning algorithms (SVM, DTs - CART and NB) in order to search patterns (SEPSIS Levels) and subsequent discovery of useful information.

This study was supported by Oracle tools. In the modeling phase three classification techniques were adopted: SVMs, DTs and NBs.

Analyzing the confusion matrix and associated metrics, we can conclude that the results obtained have high acuity.

The assessment of sepsis level is a crucial task for intensive care environments, so as soon as the risk is identified, more quickly it is applied the best and more accurate treatment. The development of classification models for sepsis is associated to some benefits, such as mortality decreasing and substantial costs reduction for institutions. Besides diagnosing the correct sepsis level, predicting the correct treatment also avoids medication testing costs.

The development of classification models for sepsis can be seen as a major contribute for developing a decision support system. Some experiences were done in order to understand some therapeutics tendencies, however the results weren't satisfactory. The results suggest that sepsis level and patient therapeutic plan aren't related.

At level of future work some work will be done in order to:

- Determine new variables that may be used to predict therapeutic;
- Build new models about therapeutic, due to the great correlation between the input variables (sepsis level) and target (therapeutic);
- Implement a decision support system based on the models of sepsis developed in this project.

Acknowledgements. This work is supported by FEDER through Operational Program for Competitiveness Factors – COMPETE and by national funds though FCT – Fundação para a Ciência e Tecnologia in the scope of the project: FCOMP-01-0124-FEDER-022674. The authors would like to thank FCT (Foundation of Science and Technology, Portugal) for the financial support through the contract PTDC/EIA/72819/ 2006. The work of Filipe Portela was supported by the grant SFRH/BD/70156/2010 from FCT.

References

1. Mador, R.L., Shaw, N.T.: The impact of a Critical Care Information System (CCIS) on time spent charting and in direct patient care by staff in the ICU: a review of the literature. International Journal of Medical Informatics 78, 435–445 (2009)
2. Häyrinen, K., Saranto, K., Nykänen, P.: Definition, structure, content, use and impacts of electronic health records: A review of the research literature. International Journal of Medical Informatics 77, 291–304 (2008)
3. Santos, M.F., Portela, F., Vilas-Boas, M., Machado, J., Abelha, A., Neves, J.: INTCARE - Multi-agent approach for real-time Intelligent Decision Support in Intensive Medicine. In: 3rd International Conference on Agents and Artificial Intelligence, Rome, Italy (2011)
4. Gago, P., Santos, M.F., Silva, Á., Cortez, P., Neves, J., Gomes, L.: INTCare: a knowledge discovery based intelligent decision support system for intensive care medicine. Journal of Decision Systems (2006)
5. Portela, F., Santos, M.F., Vilas-Boas, M.: A Pervasive Approach to a Real-Time Intelligent Decision Support System in Intensive Medicine. In: Fred, A., Dietz, J.L.G., Liu, K., Filipe, J. (eds.) IC3K 2010. CCIS, vol. 272, pp. 368–381. Springer, Heidelberg (2013)
6. Bardram, J.E., Baldus, H., Favela, J.: Pervasive computing in hospitals. Pervasive Computing in Healthcare, 49–77 (2007)
7. Varshney, U.: Pervasive healthcare and wireless health monitoring. Mobile Networks and Applications 12, 113–127 (2007)
8. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and regression. Wadsworth, Belmont (1984)
9. O. D. M. Concepts: 11g Release 1 (11.1), Oracle Corp, vol. 2007 (2005)
10. Tamayo, P., Berger, C., Campos, M., Yarmus, J., Milenova, B., Mozes, A., et al.: Oracle Data Mining. In: Data Mining and Knowledge Discovery Handbook, pp. 1315–1329 (2005)
11. Portela, F., Santos, M.F.: Data Mining Predictive Models for Pervasive Intelligent Decision Support in Intensive Care Medicine. In: Presented at the KMIS 2012 - International Conference on Knowledge Management and Information Sharing, Barcelona (2012)
12. Portela, F., Gago, P., Santos, M.F., Silva, A., Rua, F., Machado, J., et al.: Knowledge Discovery for Pervasive and Real-Time Intelligent Decision Support in Intensive Care Medicine. Presented at the KMIS 2011- International Conference on Knowledge Management and Information Sharing, Paris, France (2011)
13. Portela, F., Santos, M.F., Silva, Á., Machado, J., Abelha, A.: Enabling a Pervasive approach for Intelligent Decision Support in Critical Health Care. Presented at the HCist 2011 – International Workshop on Health and Social Care Information Systems and Technologies, Algarve, Portugal (2011)
14. SSC, Surviving sepsis campaign (2010), `http://www.survivingsepsis.org/Introduction/Pages/default.aspx` (website accessed on November 24, 2011)
15. Bone, R.C., Balk, R.A., Cerra, F.B., Dellinger, R.P., Fein, A.M., Knaus, W.A., Schein, R.M., Sibbald, W.J.: Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM consensus conference committee. American college of chest physicians/society of critical care medicine. Chest 101(6), 1644–1655 (1992)
16. Guy, W.: ECDEU Assessment Manual for Psychopharmacology: 1976. National Institute of Mental Health (1976)

17. Guy, W.: Clinical global impressions (CGI) scale. Modified from: Rush J., et al. Psychiatric measures (2000)
18. ODM, Binning (discretization) (2012), http://docs.oracle.com (website accessed on September 20, 2012)
19. Dellinger, R.P., Levy, M.M., Carlet, J.M., Bion, J., Parker, M.M., Jaeschke, R., Reinhart, K., Angus, D.C., Brun-Buisson, C., Beale, R., et al.: Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2008. Intensive Care Medicine 34(1), 17–60 (2008)
20. Levy, M., Fink, M., Marshall, J., Abraham, E., Angus, D., Cook, D., Cohen, J., Opal, S., Vincent, J., Ramsay, G.: 2001 SCCM/ESICM/ACCP/ATS/SIS international sepsis definitions conference. Critical Care Medicine 31(4), 1250–1256 (2003a) (Special Articles)
21. Levy, M., Fink, M., Marshall, J., Abraham, E., Angus, D., Cook, D., Cohen, J., Opal, S., Vincent, J., Ramsay, G.: 2001 SCCM/ESICM/ACCP/ATS/SIS international sepsis definitions conference. Intensive Care Medicine 29(4), 530–538 (2003b)

Constructing Conceptual Model for Security Culture in Health Information Systems Security Effectiveness

Ahmad Bakhtiyari Shahri[1,*], Zuraini Ismail[2], and Nor Zairah Ab. Rahim[2]

[1] Faculty of Computer Science and Information Systems,
Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia
[2] Advanced Informatics School (AIS) Universiti Teknologi Malaysia
54100 Kuala Lumpur, Malaysia
bsahmad2@live.utm.my

Abstract. The security of information systems (IS) depends on many factors, however, numerous technical advances alone cannot always create a safe and secure environment. Security incidents related to patients' data in healthcare organizations continue to increase due to human behaviors causes serious concerns. This study attempts the exploration of security culture in Health Information Systems (HIS). The scope of this paper is confined to the literature review on existing models on security culture. A conceptual model was constructed in identifying the antecedents that could influence security culture in HIS security effectiveness. We found that education and training, and communication may contribute towards a more effective implementation of security culture for HIS users. This in-progress work will then proceed to the next phase in evaluating the proposed model.

Keywords: Security culture, Security communication, Security education and training, Health information system, Security effectiveness.

1 Introduction

Advances of Information and Communications Technology (ICT) in healthcare domain is unavoidable and its presence had contributed to an increased complexity of multi-facet problems. Security of Information Systems (IS) changed from a technological issue to a behavioral one. According to UK Government report [1], as high as 95% is related to security behaviours of people while the remaing dealth with technological issues. Similarly, this holds true for Health Information Systems (HIS); healthcare organizations have sustained losses not because of insufficient or faulty technology, but rather by faulty behaviors of users [2]. More recent, a national survey conducted in United States of America [3] had confirmed a high number of data security breaches against health information resources exist, arising to the need of further attention to HIS security effectiveness central to human behavior [4].

Notablely, information security is a concept that is grounded on the process of ensuring information confidentiality, integrity and availability on technology, processes

* Corresponding author.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 213–220.
DOI: 10.1007/978-3-642-36981-0_20

and people [5]. Straub [6] described IS security effectiveness as the ability of IS security measure to protect against unauthorized or deliberate misuse of IS assets by people. Based Straub's definition, this study redefines HIS security effectiveness as the ability of HIS security measures to protect HIS assets against different threats by users' behaviors. In understanding information security effectiveness, users' culture remains a key success factor [7]. Security culture [8] is defined as a focus on security in the development of information systems and networks. It also emphasize the adoption of new ways of thinking and behaving when using and interacting within information systems and networks. In spite of the fact that the overarching objective of IS security effectiveness is through people behaviors [9], there is still lack of empirical explanations that focus on for HIS security base on security culture of users [10]; whereby most of the existing methods were based on technical viewpoint [11]. Therefore, further research on the generalizability of factors associated with user behavioral studies such as security culture is needed [12].

Hence, the objective of this paper is to construct a conceptual model in identifying the antecedents that will influence security culture in HIS security effectiveness. This paper is organized into five sections. The following section examines the related works on IS security effectiveness. The proposed conceptual model is constructed and portrayed in the third section. The fourth section describes the research methodology. Finally, conclusion and future works, ends this paper.

2 Related Works on Information Security Effectiveness

This study builds upon three streams of research. The first stream examines security culture while the second stream advocates to security education and training. Subsequetly, the third stream is related to security communication.

2.1 Security Culture

Security culture presents a common language to foster the understanding of policy and procedure in enforcing security practice [13]. Cultivating users' culture in security as the human-social factor intertwines inextricably with the overall effectiveness of security [1]. In 2003, Schlienger and Teufel [14] presented a practical method on analyze of information security culture, they believed that security culture includes all the social-cultural measures that support technical security measures. Therefore security of information is the most noticeable characteristic as the basis of daily activities of employees. In the same year they developed a management model for creating, changing and maintaining an information security culture [15].

Da Veiga and Eloff [16] emphasized that it is necessary to have a special attention to the information security culture in IS. Therefore, they presented a model for an information security culture in three levels of increasing detail [16]. Recently, Filho et al. [7] found that the understanding the human culture is still the key success factor to the information security. Therefore, they addressed ubiquitous methodology to develop security policies considering the evaluation of culture and its impacts over security policy adherence.

In 2010, a critical survey on information security and privacy in healthcare organizations was conducted by Appari and Johnson [17]. By providing a holistic view of the recent research areas of IS, they found that security culture plays a significant role in health information security.

Most recent, Brady [9] developed and empirically validated a model for predicting the effect of selected factors (eg.security management and security effectiveness) on Health Insurance Portability and Accountability Act (HIPAA) security compliance in Academic Medical Centers (AMC). With correlation analysis he found a strong relationship between security culture and security effectiveness. Security effectiveness was elucidate as the ensuring the confidentiality, integrity, and availability of patients' data in HIS.

Table 1 summaries the related works from 66 sources in consolidating security culture and its importance in IS. Thetable is categorized based chronological order based on author (and year) corresponding to the instrument used and highlighting their findings. The acceptance for security culture in HIS are reflected.

Table 1. Summary of related works on security culture and IS effectiveness (2012-2004)

Author(s) / Year	Instrument	Findings
Coen (2012)	Survey of 300 respondents	The culture encourages user's awareness and reduces information security risks.
Brady (2011)	Online survey of 76 health care professionals	Security culture is a significant predictor of security effectiveness.
Figg and Kam (2011)	Qualitative positivist approach for the exploration and hypothesis development	Power Relation between Users and Knowledge are the components of security culture.
Filho et al. (2011)	Theoretical	The article discusses the impacts of culture in security policy adherence.
Da Veiga and Eloff (2010)	Survey methodology to assess the information security culture in a South African firm with approximately 3000 employees	The study presented a framework for cultivating of information security culture in organization in three levels.
Gwen Greene and D'Arcy (2010)	Data collection had been done by two online surveys. The sample consisted of 127 employed working professionals taking MBA classes at two mid-Atlantic U.S. universities.	Security culture is a multidimensional construct consisting of top management commitment to security, security communication, and computer monitoring that has a positive effect on users' IS security.
Van Niekerk and Von Solms (2010)	Theoretical Model	Establish a security culture is the key to managing the human factors on the organization.

Table 1. (*continued*)

Alnatheer and Nelson (2009)	Suggestion a framework for information security culture in Saudi Arabia.	Training Programs influence on security culture.
D'Arcy and Greene (2009)	Survey of computer-using professionals in organizations in mid-Atlantic U.S. region.	There is a strong relationship between security culture and compliant user's behavior.
Rotvold (2008)	Survey of 144 business professionals, managers, IT administrators, and educators from various organizations.	Security awareness and training influence on security culture and information security program effectiveness.
Ma et al. (2008)	Survey of 354 certified information security professionals.	Manegement Suport influenced HIPAA compliant information security behavior.
Chang and Lin (2007)	Survey of 108 senior IT managers and professionals in various industries.	Organizational culture has a positive influence on security management effectiveness.
Gupta and Hammond (2005)	Survey 138 small business owners in Lynchburg, Virginia	Security awareness and training were positively correlated to fostering a security culture.
Von Solms and von Solms (2004)	Theoretical	Education and communication efforts are important for cultivating of information security culture.

2.2 Security Education and Training

Previous researches have investigated the role of education and training as a key to be competitive IS strategy [18]. The impact of training on IS acceptance [19] and the development of process models of end-user education and training [20] had been discussed. More importantly, the roles of computer interface designs with training methods [21]; and the effectiveness of web-based [22] for information security education. Security culture is fostered through the implementation of a comprehensive solution that includes physical, procedural, and logical forms of protection, along with the appropriate training within the organization [28].

Training and education of employees can be applied to affect the culture of an organization through improving of positive security practices and attitudes [10]. This change of mindset is of greatest importance, because a transform in attitude automatically leads to a subsequent cultural change and the employees instead of being a risk can get a security asset [23].

Whitman and Mattord [24] iterated that everyone in an organization needs to be instructed and made alert of information security since security education offers detailed information and practical instruction to staff so as to secure functions and operations. Therefore, effective information security must be culturally ingrained and

backed by continually taught [25]. In order to foster security culture, there is a neccessity for suitable training and awareness within the organization. The same is true for HIS [26]; if users are not properly educated and they are lack of security induction and culture and may compromise patients data.

Hence, it is expected that the higher the degree of security education and training, the higher will be the outcome of security culture for HIS security effectiveness. To this, we hypothesizes;

Hypothesis 1: Security education and training is positively associated to security culture for users of health information system security effectiveness.

2.3 Security Communication

Recently, Gebrasilase and Lessa [27] mentioned that communication is a dimension of information security culture that decreases the inconvenience experienced with IS security procedures.They found that security communication makes employees more professional in terms of becoming aware of their responsibilities regarding IS security. Mode of security communication include efforts through various organizational processes such as mail out reminders, notice boards [28]. Security communication had indicated strong influence in shaping security culture in an organization [29,28]. The issue of communication efforts related to security is one of the key components of security culture [30]. Likewise, a strong security culture consisting of top management support, and ongoing communication is likely to reflect a positive attitude toward information security throughout organization [29]. It was noted that a well-defined process of this dimension of security culture increases employee feelings of responsibility and ownership in decisions about security, and ultimately leads to a more positive attitude about security throughout the organization.

Besides this, an in–depth study by D'Arcy and Hovav [30] investigated the relationship between security culture and user behaviors. Using survey instrument with 105 professional computer' users as respondents, they proved a strong relationship between security culture and security communication. This suggested that security culture needs to be available at different levels in an organization, including individual, group and organizational level.

In accordance, security culture should be considered in structuring effective security program for healthcare domain [31]. D'Arcy and Hovav [30] mentioned that security culture may be shaped by security communication and communication channels play a crucial role in the adoption process. Mussa [32] and Knapp et al. [33] attest that information security culture should be improved by security communications. Hence, it would seem that security communication in a health organization would build a culture towards HIS security effectiveness. From this argument, we offer the following;

Hypothesis 2: Security communication is positively associated to security culture for users of health information system security effectiveness.

3 The Conceptual Model

Based on the previous section, from the literature review, the conceptual model was constructed and hypothesizes. The proposed conceptual model depicts in Figure 1, incorporating the essential variables. In a nutshell, security communication and security education and training may have a relationship with security culture on users for HIS security effectiveness.

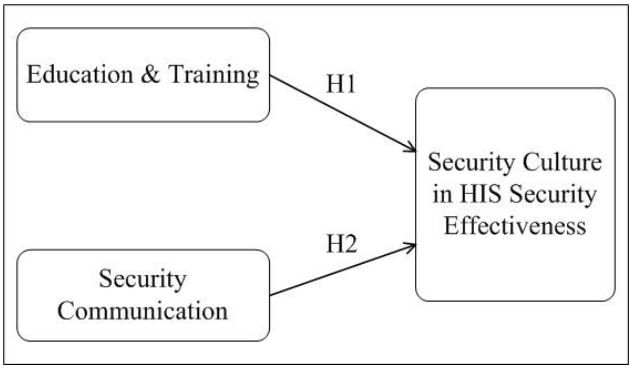


Fig. 1. The Proposed Model for Security Culture for Users in HIS Security Effectiveness

4 Research Methodology

A systematic approach and extensive search on secondary data resources was executed in deliminating construct related to HIS security effectiveness. An exploration through on-line search was carried out among the various search engines. The online databases that were given particular attention include: ACM digital library, AIS Electronic Library (AISeL), BL Direct, EBSCO host, Elsevier Science Direct, Emerald Library, IEEE Electronic Library, IGI Global, Springer Link, and Taylor & Francis Group. This databases were choosen due to its relevancy and appropriateness based on the terms chosen. Inputs for the search begins with journals related to information security, information systems, health information syatems, and security management. Specific keywords such as 'security culture', 'security effectiveness','Security communication', 'education and training', 'user behaviour' and 'security culture' were search exhausively. The outcomes of the search were consolidated according the terms as delineated in Table 1.

5 Conclusion and Future Works

In addressing the research objective, a conceptual model was successfully constructed based on the secondary data resources. The extensive search have resulted in two independent variables, namely security communication and security education and

training. Literature had exposed these variables have association with HIS security effectiveness. A proposed model hyphothesize the two relationships in establishing security culture for HIS security effectiveness. This paper may explain why security culture may play a role in estabiliahing the security effectiveness for HIS. Therefore, this work-in-progress will proceed with instrument design, subsequently continue with data collection to validate the proposed conceptual model.

Acknowledgments. This study was funded by the Research University Grant from Universiti Teknologi Malaysia (UTM) and Ministry of Higher Education (MOHE) Malaysia with the project number Q.K 130000.2138.01H98.

References

1. Colwill, C.: Human Factors in Information Security: The Insider Threat–Who Can You Trust These Days? Information Security Technical Report 14(4), 186–196 (2009)
2. Ma, Q., Johnston, A.C., Pearson, J.M.: Information Security Management Objectives and Practices: A Parsimonious Framework. Information Management & Computer Security 16(3), 251–270 (2008)
3. HIMSS Analytics: The 2010 HIMSS Analytics Report: Security of Patient Data. Technical Report (2010)
4. Winter, A., Haux, R., Ammenwerth, E., Brigl, B., Hellrung, N., Jahn, F.: Quality of Health Information Systems. Health Information Systems, 201–236 (2011)
5. Torres, J., Sarriegi, J., Santos, J., Serrano, N.: Managing Information Systems Security: Critical Success Factors and Indicators to Measure Effectiveness. Information Security 4176, 530–545 (2006)
6. Straub, D.W.: Effective IS Security. Information Systems Research 1(3), 255–276 (1990)
7. Filho, E.L., Souza, J.H.P., Chaves, A.T., Hashimoto, G.T., Rosa, P.F.: The Impact of Corporate Culture in Security Policies – A Methodology. In: The Seventh International Conference on Networking and Services (ICNS 2011), Venice/Mestre, Italy, May 22-27, pp. 98–103 (2011)
8. OECD: OECD guidelines for the security of information systems and networks: towards a culture of security. Organisation for Economic Co-operation Development (2002)
9. Brady, J.W.: Securing Health Care: Assessing Factors That Affect HIPAA Security Compliance in Academic Medical Centers. In: 44th Hawaii International Conference on System Sciences, Kauai, HI, Kauai, HI, pp. 1–10. IEEE (2011)
10. Knapp, K.J., Marshall, T.E., Rainer Jr, R.K., Ford, F.N.: Information Security Effectiveness: Conceptualization and Validation of a Theory. International Journal of Information Security and Privacy (IJISP) 1(2), 37–60 (2007)
11. Benhocine, A., Laouamer, L., Hadji, H.: Toward an Efficient Security: A New Methodology for Information Security. Journal of Economics and Administration 1(1) (2011)
12. Ball, D.M., Levy, Y., Lauderdale, F.: Emerging Educational Technology: Assessing the Factors that Influence Instructors' Acceptance in Information Systems and Other Classrooms. Journal of Information Systems Education 19(4), 431–444 (2008)
13. Figg, W.C., Kam, H.J.: Medical Information Security. International Journal of Security (IJS) 5(1), 22 (2011)

14. Schlienger, T., Teufel, S.: Analyzing Information Security Culture: Increased Trust by an Appropriate Information Security Culture. In: 14th International Workshop on Database and Expert Systems Applications 2003, pp. 405–409 (2003)
15. Schlienger, T., Teufel, S.: Information Security Culture - From Analysis to Change. South African Computer Journal 7(31), 46–52 (2003)
16. Da Veiga, A., Eloff, J.: A Framework and Assessment Instrument for Information Security Culture. Computers & Security 29(2), 196–207 (2010)
17. Appari, A., Johnson, M.E.: Information Security and Privacy in Healthcare: Current State of Research. International Journal of Internet and Enterprise Management 6(4), 279–314 (2010)
18. Ennis, M.R.: Competency Models: A Review of The Literature and The Role of The Employment and Training Administration (ETA). US Department of Labor, 1–24 (2008)
19. Sardinha, F., Costa, C.J.: Training and Interface Features in Technology Acceptance. In: Proceedings of the 2011 Workshop on Open Source and Design of Communication 2011, pp. 55–60. ACM (2011)
20. Puhakainen, P., Siponen, M.: Improving Employees' Compliance through Information Systems Security Training: An Action Research Study. MIS Quarterly 34(4), 757–778 (2010)
21. Moos, D.C., Azevedo, R.: Learning with Computer-Based Learning Environments: A Literature Review of Computer Self-Efficacy. Review of Educational Research 79(2), 576–600 (2009)
22. Van Niekerk, J., Von Solms, R.: A Web-Based Portal for Information Security Education. In: Information Security South Africa (ISSA), Johannesburg, South Africa, July 10-12, 2002, pp. 1–10. ISSA (2007)
23. Van Niekerk, J., Von Solms, R.: An Holistic Framework for the Fostering of an Information Security Sub-Culture in Organizations. Information Security South Africa (ISSA), Johannesburg (2005)
24. Whitman, M.E., Mattord, H.J.: Principles of Information Security. Course Technology Ptr., Boston (2011)
25. Lineberry, S.: The Human Element: The Weakest Link in Information Security. Journal of Accountancy 204(5), 44 (2007)
26. Sennewald, C.A.: Effective Security Management. Butterworth-Heinemann (2011)
27. Gebrasilase, T., Lessa, L.F.: Information Security Culture in Public Hospitals: The Case of Hawassa Referral Hospital. The African Journal of Information Systems 3(3), 72–86 (2011)
28. Karjalainen, M.: Imroving Employees' Information Systems (IS) Security Behavior. University of Oulu (2011)
29. Vroom, C., Von Solms, R.: Towards Information Security Behavioural Compliance. Computers & Security 23(3), 191–198 (2004)
30. D'Arcy, J., Greene, G.: The Multifaceted Nature of Security Culture and Its Influence on End User Behavior. In: International Workshop on Information Systems Security Research 2009, pp. 145–157 (2009)
31. Pattinson, M.R., Anderson, G.: How Well Are Information Risks Being Communicated to Your Computer End-Users? Information Management & Computer Security 15(5), 362–371 (2007)
32. Mussa, C.C.: A Prudent Access Control Behavioral Intention Model for the Healthcare Domain. Nova Southeastern University (2012)
33. Knapp, K.J., Marshall, T.E., Rainer, R., Ford, F.N.: Managerial Dimensions in Information Security: A Theoretical Model of Organizational Effectiveness. International Information Systems Security Certification Consortium (ISC) 2 (2005)

Using Domain-Specific Term Frequencies to Identify and Classify Health Queries

Carla Teixeira Lopes[1], Daniela Dias[1], and Cristina Ribeiro[1,2]

[1] DEI, Faculdade de Engenharia, Universidade do Porto
[2] INESC TEC
Rua Dr. Roberto Frias s/n, 4200-465, Portugal
{ctl,ei06025,mcr}@fe.up.pt

Abstract. In this paper we propose a multilingual method to identify health-related queries and classify them into health categories. Our method uses a consumer health vocabulary and the Unified Medical Language System semantic structure to compute the association degree of a query to medical concepts and categories. This method can be applied in different languages with translated versions of the health vocabulary. To evaluate its efficacy and applicability in two languages we used two manually classified sets of queries, each on a different language. Results are better for the English sample where a distance of 0.38 to the ROC optimal point (0,1) was obtained. This shows some influence of the translation in the method's performance.

Keywords: Health Information Retrieval, Health Queries, Medical Vocabularies, Web Information Retrieval.

1 Introduction

The Web is now a major source of information worldwide and the use of search engines to find health information is a common practice. In 2006, 80% of Internet users in the United States used the Web to search for health information [3]. According to Eysenbach and Köhler, over 12 million health queries are made per day in *Google* [2]. The classification of queries is frequently used to distinguish them according to the topic. This classification can be manual, may involve the comparison of a query with databases of queries or require machine learning processes. Another possibility is to use controlled vocabularies or thesaurus of terms, in areas where the quality of these structures can be trusted. Since most health queries contain terms that can be mapped to health vocabularies [5], we propose a method to detect consumer health queries that takes advantage of existing high-quality health vocabularies, can be applied in different languages and can classify queries into health categories like diseases.

2 Related Work

Two previous works report methods to identify health queries. Eysenbach and Köhler [2] proposed a method to automatically classify search strings as

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 221–226.
DOI: 10.1007/978-3-642-36981-0_21

health-related based on the proportion of pages on the Web containing the search string plus the word "health" and the number of pages containing only the search string. In another work, Lopes [4] compares the Eysenbach and Köhler's method with a method that uses health vocabularies to identify health queries. In this last method, the author considers that the presence of a health vocabulary's term in a query is sufficient to classify the query as being health-related. In this work, several variants of both methods are compared.

Like the work of Lopes [4], our work will use health vocabularies but in a different way. While the previously described method is discrete, simply indicating if a query is or not a health query, our method will produce a score, indicating the degree to which a query is related to the health domain. Moreover, our method can be used to classify health queries into categories like *Disease or Syndrom* or *Anatomical Structure*.

3 Proposed Method for Query Classification

The proposed method takes advantage of the UMLS predefined structures and the Consumer Health Vocabulary (CHV). It includes the creation of indexes to help the comparison between query terms and the health vocabulary and the calculation of the final score. Besides classifying each query as being health-related or not, we also associate it with the UMLS specialized health categories.

3.1 Health Semantic Structures

We have chosen the Consumer Health Vocabulary[1] (CHV), developed under an open source and collaborative initiative that is linked to the Unified Medical Language System (UMLS) and its many sources. The CHV has 42,977 health concepts and 158,508 concept strings in English. The UMLS is one of the most consistent and robust health semantic structures including about 1 million biomedical concepts from 100 different sources and a large semantic structure.

3.2 Vocabulary Translation

One of the main disadvantages of a method based on vocabularies is its dependence on the language and country in which it was created. Our hypothesis here is that we can apply our method by previously translating the CHV without much penalty on the results. We expect the translation process to have some influence on the classification results but also hope to minimize it using a good translation process. To evaluate the efficacy of our method in a language other than English we used the *Google Translator API*. We manually evaluated 1% of the total number of translated strings and concluded that 84.2% of the translations were good, which is very satisfactory.

[1] `http://www.consumerhealthvocab.org`

3.3 CHV Subsets

The CHV vocabulary contains concepts of several categories and some of them contain strings (e.g.: car, driving) that, when isolated from other health terms or concepts, are not useful to identify a health query. To avoid false positives we decided to obtain different subsets of the CHV vocabulary instead of using only the complete CHV. We defined four subsets: one with concept strings from UMLS categories containing concepts more likely to occur in consumer health queries (HEALTH), one with the consumer preferred string for each concept in the CHV (CHVP), one with the UMLS preferred string for each concept in the CHV (UMLSP) and the other with the MedlinePlus category concept strings (MEDP). MedlinePlus is a website for health consumers and the UMLS category with this name contains the concepts explored in this site.

3.4 Auxiliary Structures

For each subset, we created an inverted index containing the unique terms mapped to a list of unique identifiers for each concept string in the subset and their association degree with each concept string. The association degree of a term t to a concept string c, w_t^c, is computed as the ratio $tf_t^c/|c|$, where the numerator is the term frequency of t in the concept string c and the denominator is the number of terms in concept string c. If we consider the CHV strings *tooth* and *dental infection*, the terms *dental* and *infection* would be associated with the second string with a probability of 0.5 and the term *tooth* with the first string with a probability of 1.

3.5 Combining Inverted Index Entries

In the classification process, queries are tokenized and, for each term, we retrieve the corresponding posting list from the inverted index. We then combine these lists to calculate the final score for the query. As showed in Figure 1, two combination methods were tested. The first joins the lists and, when an identifier appears more than once, the w_t^c are added. The resulting list contains the weights of each CHV string in the query, w_c^q. This way we can easily identify if a query contains parts or entire health CHV strings. The second method (M2), joins the lists as M1, but also counts the occurrences of each CHV string in the query ($cf_{c,q}$). As a final step, we adjust the weights calculated in the first method as $w_c^q \times \frac{cf_{c,q}}{|q|}$, where $cf_{c,q}$ is the frequency of c in query q and $|q|$ the number of unique terms in q.

3.6 Final Score Calculation

After obtaining the query list, we calculate the final score that will be used to classify the query as health related or not. To do this, we propose some variants for the two previous methods, presented in Table 1. Here the $Query$ is the query list obtained after the previous combination in each method, $tf_{h,q}$ is the number

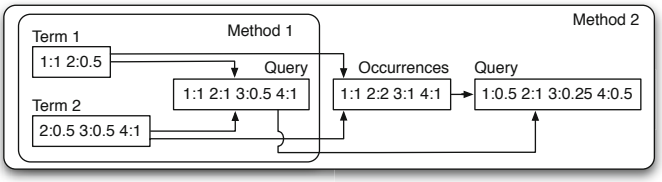


Fig. 1. Joining posting lists in Methods 1 and 2

Table 1. Variants applied to the different methods

Variant	Formula	Boost
M1Max	$\max(Query) \times (tf_{h,q} \div \lvert q \rvert)$	No
M1MaxBoost		Yes
M1Avg	$avg(top_1^5(Query)) \times (tf_{h,q} \div \lvert q \rvert)$	No
M1AvgBoost		Yes
M2Max	$\max(Query)$	No
M2MaxBoost		Yes
M2Avg	$avg(Query)$	No

of terms in query q included in the inverted index, and $|q|$ is the number of unique terms in q. M1Max and M1MaxBoost use the maximum weight of the $Query$ list under the assumption that, if a query is completely matched by a health concept, it is a health query. In M1Avg and M1AvgBoost we computed the average of the 5 largest probabilities in the query list.

The product used in the M1 variants lowers the score of the queries that have non-health terms even if the query matches an entire concept, because a concept may change when a term is added. An example of this case is the query "tooth piercing" as "tooth" is a full concept in the CHV subset as an anatomical part and the term "piercing" doesn't appear in it. Without the final product this query would have a score of 1 and it scores 0.5 with it. This is not needed in the M2 variants because the M2 already uses the occurrences of each CHV concept string in the whole query.

To promote the queries that contain terms that appear more frequently in the CHV vocabulary, we decided to test the application of a boost value b to the term weights in a CHV string ($b \times w_t^c$). This boost is similar to the document frequency *df* used in Information Retrieval and is equal to the number of strings in the CHV in which the term appears.

3.7 Classifying Health Queries

Queries that have the final score above a specific threshold will be classified as being health-related. We also used the UMLS semantic network to assign health categories to each query. In this sense we created an index similar to the one

described above where terms are replaced by CHV strings and the posting lists contain categories and not strings. After obtaining the query list as explained above we create another list with the category associated to each CHV string in the query list and the weight, w_c^q, previously associated with the string. If a category appears more than once, we select its maximum weight.

4 Findings and Discussion

To evaluate the methods using the English (EN) CHV concept strings we have used a dataset created by Beitzel and Lewis who had queries classified into 20 topical categories by a team of approximately ten human assessors. We included 1,647 queries, part classified as health queries and part classified into other categories [1]. In Portuguese (PT) we have used a collection of 1,522 queries manually classified by medical students. For each method we calculated the true positive rate (TPR), false positive rate (FPR), accuracy (ACC) and the distance (ROCD) to the optimal point in the ROC Space (0,1).

With all the CHV subsets, initial tests showed that the HEALTH subset produces the best results with respect to accuracy and distance to the ROC optimal point. However, the MEDP subset revealed a better FPR (13%-14%) due to a lower number of concept strings and its focus on consumers. In terms of TPR, M1Max using the UMLSP subset and M1Max using the CHV entire vocabulary had the best results with 68%. The UMLSP, despite having fewer strings than the CHV subset, has the same TPR probably because it contains almost all of the concept strings that led to query classification. In general, almost all methods have TPR and ACC values above 60%.

Table 2 shows the results of each method used in the classification of the sample collections in both languages with the HEALTH Subset. As shown, the best method is M2Max with a threshold of 0.17 using the English vocabulary. In Portuguese the best method is M1Max with a threshold of 0.5. We can therefore conclude that translation has impact on the results. The difference in TPR and ACC is negligible. However, differences in ROCD and FPR are more expressive. We believe our results can be improved by removing unspecialized terms that, alone, are not health-related.

Comparing our results in the English language with the results obtained by Lopes [4], we notice that our best method has a ROCD of 0.38, a little worse than Lopes's best result. Her best result was obtained applying the Eysenbach and Köhler method using the Yahoo! search engine and had a ROCD of 0.34. However, our method has a smaller ROCD than all the other variants of the Eysenbach and Köhler method and all the methods that use health vocabularies. Moreover it has the advantage of being able to associate the queries with the UMLS specialized health categories.

Table 2. Best results in the HEALTH subset. T=threshold, L=language.

M	T	L	TPR	FPR	ACC	ROCD
M1Max	0.2	EN	**0.76**	0.33	**0.73**	0.41
M1Avg	0.2	EN	0.66	**0.2**	0.7	0.39
M1MaxBoost	0.2	EN	0.71	0.29	0.71	0.41
M1AvgBoost	0.75	EN	0.72	0.33	0.71	0.43
M2Max	0.17	EN	0.68	0.21	0.71	**0.38**
M2Avg	0.1125	EN	0.67	0.32	0.68	0.46
M2MaxBoost	0.35	EN	0.71	0.29	0.71	0.41
M1Max	0.5	PT	0.65	0.31	**0.67**	**0.46**
M1Avg	0.2	PT	0.65	0.32	0.66	0.47
M1MaxBoost	0.75	PT	0.66	0.33	**0.67**	0.47
M1AvgBoost	0.2	PT	0.67	0.35	0.66	0.48
M2Max	0.5	PT	0.63	**0.30**	0.61	0.48
M2Avg	0.1	PT	**0.68**	0.40	0.65	0.51
M2MaxBoost	0.75	PT	0.66	0.33	0.66	0.47

5 Conclusions

This work proposes a new method to identify and classify health-related queries that explores the UMLS predefined structures and can be applied in different languages. The influence of the translation process in the proposed method is noticeable but does not compromise its overall effectiveness. Moreover, and not less important, our approach allows the association of queries to the UMLS semantic tree and their classification into categories like *Disease or Syndrom* or *Anatomical Structure*. The output of our method can be useful to search engines that can, for example, use it to provide contextualized query suggestions or even information about the health subject searched for. In the future, we would like to test these methods with an inverted index created with multiple data from different vocabularies and to combine them with machine learning techniques.

References

1. Beitzel, S.M., Jensen, E.C., Frieder, O., Lewis, D.D., Chowdhury, A., Kolcz, A.: Improving Automatic Query Classification via Semi-Supervised Learning. In: Fifth IEEE International Conference on Data Mining (2005)
2. Eysenbach, G., Köhler, C.: What is the prevalence of health-related searches on the World Wide Web? Qualitative and quantitative analysis of search engine queries on the internet. In: AMIA Symposium (2003)
3. Fox, S.: Online Health Search 2006. Tech. rep., Pew Internet & American Life Project (2006)
4. Lopes, C.T.: Evaluation and comparison of automatic methods to identify health queries. In: Doctoral Symposium on Informatics Engineering 2008 (2008)
5. Zeng, Q.T., Crowell, J., Plovnick, R.M., Kim, E., Ngo, L., Dibble, E.: Assisting consumer health information retrieval with query recommendations. JAMIA (2006)

Dealing with Constraint-Based Processes: Declare and Supervisory Control Theory

Sauro Schaidt, Agnelo Denis Vieira,
Eduardo de Freitas Rocha Loures, and Eduardo Alves Portela Santos,

Pontifícia Universidade Católica do Paraná, Escola Politécnica,
Imaculada Conceição 1155, Curitiba, Brazil
{sauro.schaidt,agnelo.vieira,
eduardo.loures,eduardo.portela}@pucpr.br

Abstract. The constraint-based processes are those that do not require a procedural model that explicitly specify the execution procedure. Declarative languages are more suitable for modeling and implementing this type of process. This paper aims to present and analyze two approaches that deal with the modeling and execution of constraint-based processes. The first approach makes use of Linear Temporal Logic and software Declare as a tool for modeling constraints. The second approach is based on the Supervisory Control Theory (SCT). Both approaches presented in this paper aims to monitor and restrict execution sequences of tasks such that constraints are not violated. Despite the control logic is built based on constraints, it does not limit the user by imposing rigid control-flow structures. A discussion of the results, advantages and the main drawbacks of the two approaches are presented in this paper.

Keywords: constraint-based models, declarative languages, declare, supervisory control theory.

1 Introduction

Business organizations have developed processes for the purpose of attaining their goals. A minimum set of criteria or rules should be imposed on a process to ensure that its implementation achieves the intended objectives. These criteria can be called business rules or restrictions [1] [2]. Some restrictions are established by law or governmental authorities, others by those involved in the business process. Agreements made with customers and suppliers also impose limits and rules. Moreover, the managers of the organization defined limits as well.

There are processes that require little or no rules for its implementation. These processes are said flexible processes or poorly structured. Declarative languages are better suited to modeling these processes precisely because they facilitate the declaration of little or no business rule [3] [4] [5]. These languages define the activities that must be performed to produce the expected results but not define exactly how these activities should be performed. Thus, any execution order of

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 227–236.
DOI: 10.1007/978-3-642-36981-0_22

activities is possible provided that the restrictions (imposed by the rules) are not violated.

On the other hand, there are processes that require a strong imposition of rules for their implementation. These processes are called rigid or highly structured processes. Imperative languages are better suited to modeling these processes because, unlike declarative languages, facilitating the declaration of a large amount of business rules. These languages define the activities that must be performed to produce the expected results and also exactly how these activities should be performed. Thus, we need a model that explicitly define the ordering and execution of activities.

This paper aims to present and analyze two approaches that deal with the modeling of constraint-based processes. The first approach makes use of Linear Temporal Logic [6] as a tool for modeling constraints. The second approach is based on the Supervisory Control Theory (SCT) [7], and it proposes a control system based on local supervisors [8] which force the process to not violate the constraints. In order to enrich the analysis was conducted an example in which the DECLARE [4] [9] and the SCT are employed. A discussion of the results, advantages and the main drawbacks of the two approaches are presented in this paper.

2 Declare

DECLARE is developed as a constraint-based system and uses a declarative language grounded in Linear Temporal Logic (LTL) [10] for the development and execution of process models. As opposed to traditional imperative approaches to process modeling, DECLARE uses a constraint-based declarative approach. An imperative model focuses on specifying exactly how to execute the process, i.e., all possibilities have to enter into the model by specifying its control-flow. A declarative model specifies a set of constraints, i.e., rules that should be followed during the execution. In this way, the declarative model implicitly defines the control-flow as all possibilities that do not violate any of the given constraints.

LTL can be used to define properties of sequences of tasks. Every trace represents an executed alternative where the tasks in the trace occur exactly in the order in which they appear. In the field of model checking, LTL is extensively used to check whether a system satisfies properties specified by LTL formulae. The business rules are formally described using LTL, allowing verification algorithms to identify if such rules are obeyed [9][10].

Procedural (or imperative) models (e.g. BPMN, EPC, YAWL or Petri Nets), provide constructs such as AND/OR-splits, AND/OR-Joins, etc. These constructs aim at supporting frequently needed workflow patterns. DECLARE allows for customized specification of relation types frequently needed to model constraints between tasks. Considering LTL formulas can be difficult to understand by nonexperts, DECLARE provides a graphical representation of constraints that hides the associated LTL formulas from users of declarative workflows [3] [12]. This graphical representation is called *constraints templates*. Each template has a name, an LTL formula and a graphical representation. Templates can be used to create actual constraints for a specific process [3][9].

3 Supervisory Control Theory

The Supervisory Control Theory (SCT) proposed by Ramadge and Wonham [7] is an important tool for the control of Discrete Event Systems (DES). Based upon models describing the independent behavior of each subsystem to be coordinated and a set of control specifications aiming to restrict their concurrent behavior, all of them represented by automata, it is possible to perform a formal and automatic synthesis of a supervisor. The control action of such supervisor restricts the behavior of the system so that the control specifications are satisfied in a minimally restrictive way and ensures that this behavior is non-blocking (i.e., there is always an event sequence available to complete an activity).

3.1 Monolithic Approach

The SCT introduces the concept of event controllability. The set of events is partitioned into controllable events (Σ_c) and uncontrollable events (Σ_{uc}). An event is classified as controllable if a control agent, named supervisor, can disable its occurrence. Considering a subsystem in $\{\mathbf{G_i} \mid i \in I\}$, Σ_c^{Gi} denotes its set of controllable events and Σ_{uc}^{Gi} its set of uncontrollable events. The whole set of controllable events is $\Sigma_c = \cup_{\forall i \in I} \Sigma_c^{Gi}$.

In a monolithic approach a single global supervisor, is synthesized in order to cope with all control specifications. The necessary and sufficient conditions for the existence of supervisors are established in [7]. Formally, a supervisor is a map from the closed behavior of $\mathbf{G}$ to a subset of events to be enabled ($\mathfrak{S} : L(\mathbf{G}) \rightarrow 2^{\Sigma}$). A supervisor may be represented by an automaton together with an output map ($\mathfrak{S} = (\mathbf{S}, \Phi)$), with $\mathbf{S} = (\Sigma^S, Q^S, \delta^S, q_0^S, Q_m^S)$. The automaton $\mathbf{S}$ is driven by the occurrence of events in the plant, the output map $\Phi : Q^S \rightarrow 2^{\Sigma c}$ specifies the subset of controllable events that must be disabled as a correspondence of the active state of automaton $\mathbf{S}$. The action of a supervisor includes disabling controllable events and unmarking sequence of events. The automaton $\mathfrak{S}/\mathbf{G}$ represents the optimal behavior of $\mathbf{G}$ under the supervision of $\mathfrak{S}$, where $L(\mathfrak{S}/\mathbf{G}) \subseteq L(\mathbf{G})$ and $L_m(\mathfrak{S}/\mathbf{G}) \subseteq (L(\mathfrak{S}/\mathbf{G}) \cap L_m(\mathbf{G}))$. Whenever $L_m(\mathfrak{S}/\mathbf{G})$ is a proper subset of $(L(\mathfrak{S}/\mathbf{G}) \cap L_m(\mathbf{G}))$ then $\mathfrak{S}$ is said to be a (un)marker supervisor.

3.2 Local Modular Approach

According to the LMC approach [8] the system as a whole must be decomposed into several independent subsystems and the behavior of each of them must be modeled in a high level of abstraction by a corresponding automaton. This results in the set $\{\mathbf{G_i} \mid i \in I\}$, where $i \in I$ identifies each subsystem. Each automaton can be represented by a 5-tuple $\mathbf{G_i} = (\Sigma^{Gi}, Q^{Gi}, \delta^{Gi}, q_0^{Gi}, Q_m^{Gi})$ where Σ^{Gi} is the alphabet (set) of events, Q^{Gi} is the set of states, $\delta^{Gi} : (Q^{Gi} \times \Sigma^{Gi}) \rightarrow Q^{Gi}$ is the state transition function (in general partially defined), q_0^{Gi} is the initial state and $Q_m^{Gi} \subseteq Q^{Gi}$ is the set of marker states. The model of the entire system is a Product System Representation (PSR), *i.e.*, a set of asynchronous subsystems such that all pairs of subsystems in $\{\mathbf{G_i} \mid i \in I\}$ have disjoint alphabets.

According to the LMC approach, instead of synthesizing a single global supervisor that satisfies the entire set of specifications, a local supervisor must be synthesized for each specification in $\{\mathbf{E_j} \mid j \in J\}$. This leads to a set of local supervisors $\{\mathfrak{S}_j \mid j \in J\}$. The synthesis of a local supervisor ($\mathfrak{S}_j$) is performed considering the corresponding specification ($\mathbf{E_j}$) and its corresponding local plant ($\mathbf{Gl_j}$), which is obtained by the synchronous product of only those subsystems that share some event with such specification. If at least one local supervisor in the set $\{\mathfrak{S}_j \mid j \in J\}$ disables the occurrence of an event, then the occurrence of this event is disabled in **G**. The automaton $\mathfrak{S}/\mathbf{Gl_j}$ represents the optimal behavior of the local plant $\mathbf{Gl_j}$ under the supervision of the corresponding local supervisor $\mathfrak{S}_j$.

3.3 Control Architecture

Ramadge and Wonham consider a general control paradigm for how supervisor S interacts with G. In this paradigm, S observes some—possibly all—of the events G executes. Then, S tells G which events in the current active event set of G are allowed next. More precisely, S has the capability of disabling some, but not necessarily all, feasible events of G. The set of disabled events may change whenever S observes the execution of a new event by G. Therefore, S exerts dynamic feedback control on G.

4 Running Example

Project management usually consists of various management processes, monitoring and control activities. These processes are performed in different conditions for each new project, which requires a flexible modeling. One of the most popular benchmarks for project management is the PMBOK (Project Management Body of Knowledge). PMBOK in its fourth version establishes a set of 42 macro-processes in nine knowledge areas. These areas of knowledge are the Scope Management Project which consists of five processes: *Collect Requirements*, *Define Scope*, *Create WBS*, *Verify Scope* and *Control Scope*. The *Collect Requirements* process was selected for the implementation of declarative modeling techniques and illustration of the approaches presented here. The goal of this process is to identify the set of requirements of the final product of the project.

The PMBOK provides three stages for each process: Inputs, Tools & Techniques, and Outputs. The inputs to this process are the documents *Project Charter* (PC) and *Stakeholder Register* (SR). The PC is used to provide the requirements and description of the product at a high level to develop the detailed requirements of the product. The SR is used to identify the parts that can provide detailed information about the requirements of the project and product. The tools and techniques adopted for implementing this model are: interviews, focus groups, facilitated workshops, questionnaires and surveys, prototypes and brainstorm. The outputs suggested by PMBOK are *Requirements Documentation* (RD), *Requirements Management Plan* (RMP) and *Requirements Traceability Matrix* (RTM) [13]. For this work we selected the output *Documentation Requirements*.

4.1 Implementation in DECLARE

We consider the following tasks that we provide control: Project Charter and Stakeholders Register (*t1*), Interviews (*t2*), Focus Groups (*t3*), Questionnaires and Surveys (*t4*), Facilitated Workshops (*t5*), Prototypes (*t6*), Brainstorm (*t7*) and Requirements Documentation (*t8*). There are four constraints specified for this process: constraint C1 defines that *Review Project Charter and Stakeholders Register* (*t1*) must be the first executed task in an instance; constraint C2 defines that at least one of the five tasks *Interviews* (*t2*), *Focus Groups* (*t3*), *Questionnaires and Surveys* (*t4*), *Facilitated Workshops* (*t5*) and *Prototypes* (*t6*) has to be executed, but all can be executed and each of them can be executed an arbitrary number of times; constraint C3 defines that tasks *Prototypes* (*t6*) and *Brainstorm* (*t7*) are always executed next to each other; constraint C4 defines that task *Requirements Documentation* (*t8*) is executed at least once.

Figure 1 shows the corresponding modeling process in DECLARE. It is observed that for each constraint predefined as a LTL formula there is a corresponding constraint template. The constraint C1 is graphed by the operator *init* allocated on the top of rectangle that represents the task *review PC and SR* (*t1*). The constraint C2 is represented by 1of5 connector that connects the five interrelated tasks. The constraint C3 is represented by the *succession chain* operator symbolized by an arrow linking the two interrelated activities (*prototypes* and *brainstorm*). The constraint C4 operator is represented by 1 .. * allocated on the top of rectangle which represents task *requirements documentation*.

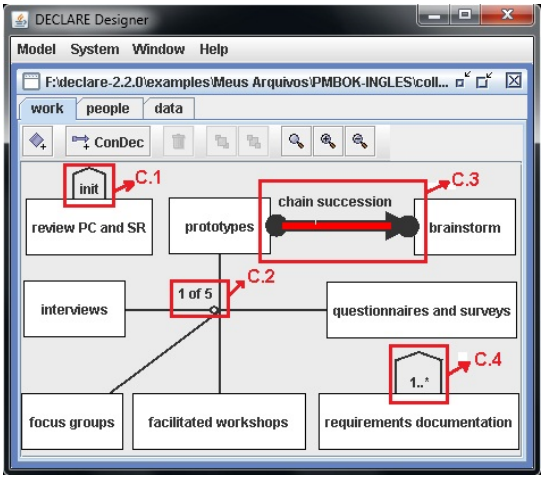


Fig. 1. Collect Requirements process in Declare

4.2 Implementation Using SCT

Activities under control are *Review Project Charter and Stakeholders Register* (*t1*), *Interviews* (*t2*), *Focus Groups* (*t3*), *Questionnaires and Surveys* (*t4*), *Facilitated Workshops* (*t5*), *Prototypes* (*t6*), *Brainstorm* (*t7*) and *Requirements Documentation*

(*t8*). We assume each task *ti* (*i*=1,…,8) is modeled as an automaton with two states: (1) an initial state means the task is not being executed (a case has not entered) and (2) another state means a case is being processed. With event *tis* (*start*), the task is initiated. When it finishes, signaled by the occurrence of event *tic* (*complete*) or *tix* (*cancel*), it returns to initial state. Notice that we consider that both events *tic* and *tix* have the same effect in the task model. This is because there is no constraint, in this example, related to cancelation of a task.

According to SCT, to calculate the control logic that does not violate the constraints described in Section 4.1, it is necessary to express such constraints in terms of automata. These automata are shown in Figure 2. According to LMC, the first step to synthesizing local supervisors is to obtain the local plant for each specification. Local plants for C1, C2, C3 and C4 are given by Gl_{C1} = t1, Gl_{C2} = t2||t3||t4||t5||t6, Gl_{C3} = t6||t7, and Gl_{C4} = t8, respectively. Using algorithms proposed by Ramadge and Wonham [7], it is possible to obtain four local supervisors, each guaranteeing the constraint expressed by corresponding automaton Cj (j=1,2,3,4). Synthesis of local supervisor Sj is performed considering corresponding specification Cj and its local plant Glj. Software TCT [14] performs synchronous composition, synthesis of supervisors, and reducing procedures of supervisors. Figure 3 shows the local supervisors Sj.

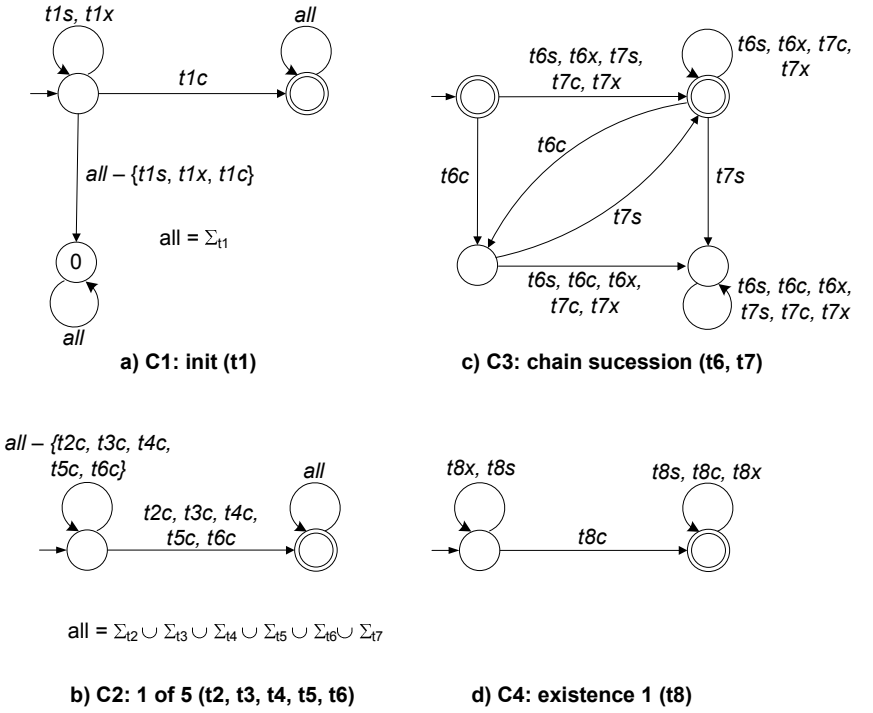


Fig. 2. Constraints

Each local supervisor disables a set of controllable events according to its states. A corresponding pair (S_j, Φ_j) represents each supervisor, where Φ_j represents the output map. Considering local supervisors S_j shown in Figure 3, their output maps are: S_1 is $\Phi_1(0)$ = {t2s, t3s, t4s, t5s, t6s, t7s, t8s}, $\Phi_1(1) = \varnothing$; S_2 is $\Phi_2(0) = \Phi_2(1) = \varnothing$; S_3 is $\Phi_3(0) = \varnothing$, $\Phi_3(1)$ = {t7s}, $\Phi_3(2)$ = {t6s}, $\Phi_3(3)$ = {t6s}, $\Phi_3(4)$ = {t7s}; and S_4 is $\Phi_4(0) = \Phi_4(1) = \Phi_4(2) = \Phi_4(3) = \varnothing$.

Figure 4 shows the interplay between tasks t6 and t7 and the local supervisor S3. In initial state of the local supervisor, there is no controllable event being disabled ($\Phi_3(0) = \varnothing$). Thus, both tasks t6 and t7 may be initiated. It means users are able to choose one of them to initiate. In the case t6 has been chosen, the state 1 of S3 is reached and a new control action is established. Figure 5 illustrated this situation. In state 3 of S3,

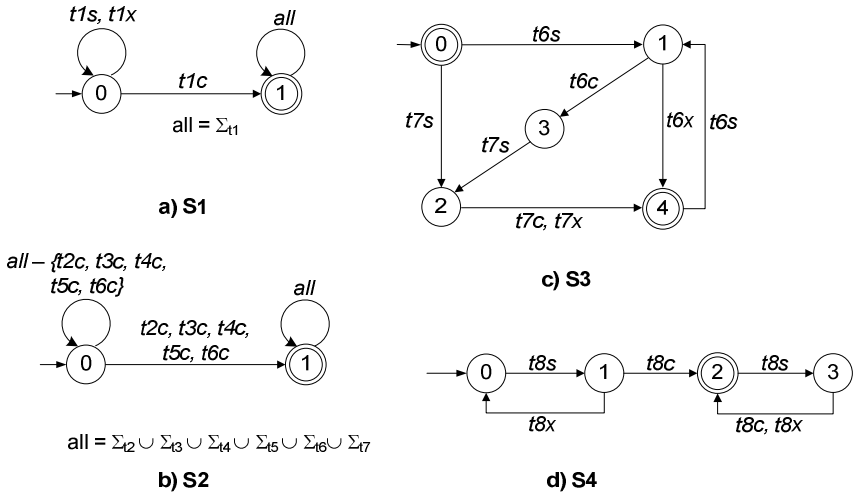


Fig. 3. Local supervisors

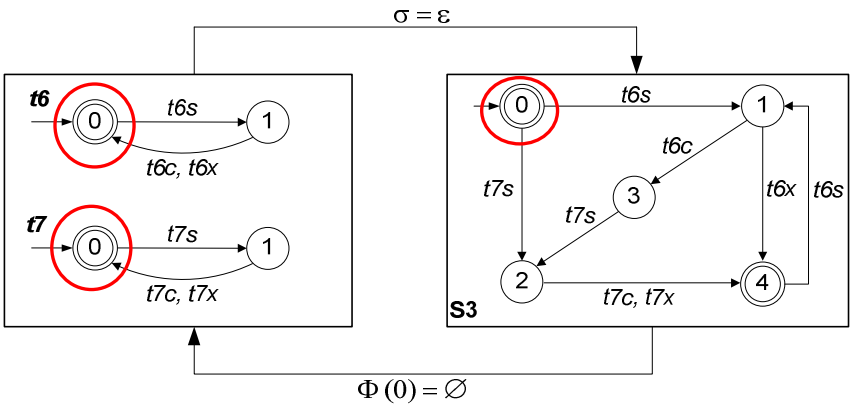


Fig. 4. Interplay between tasks and local supervisor 1

the controllable event t7s is disabled ($\Phi_3(1) = \{t7s\}$). It means users cannot initiate task t7 until task t6 has been finished (when event t6c occurs from state 1). By completing the execution sequence based on control actions from local supervisors, users are guaranteed no violation of constraints.

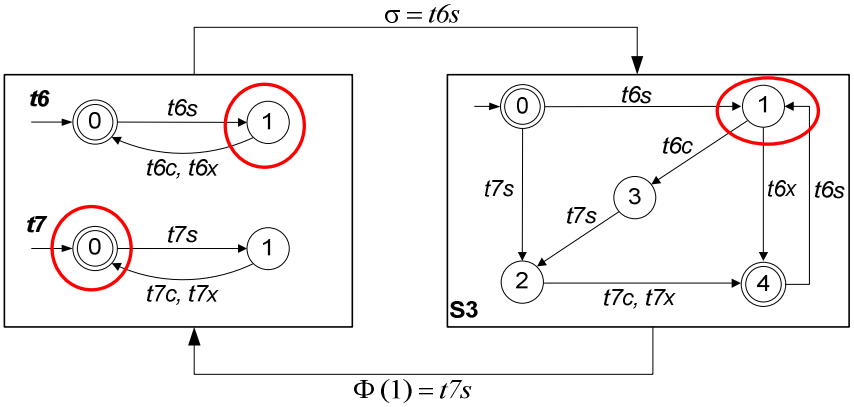


Fig. 5. Interplay between tasks and local supervisor 2

5 Discussion

Constraint-based processes require a set of rules that limit their behavior to certain boundaries. This paper presents and analyzes two approaches for treating such processes. Both models presented here consider that in a constraint-based process anything is allowed unless explicitly forbidden. The use of LTL in the construction of logical expressions establishes a systematic modeling of constraints. One can therefore construct libraries, facilitating the implementation in other projects. DECLARE also proposes the use of templates associated with each one of constraints, making it easier for users without knowledge in LTL. Also, according to [15], it is possible integration DECLARE with YAWL, adding flexibility in process execution.

According to [15], the combination of certain constraints can cause errors in constraint models that lead to undesirable effects at execution time. The authors distinguish two types of model errors: dead tasks and conflicts. The task is dead if it cannot be executed in any instance of the model without violating the specified constraints. The model contains a conflict if there are no traces that can satisfy the model, that is, instances of the model can never become satisfied. Both cases can be detected using the automaton generated for the mandatory formula of the model. [11] and [16] proposes to improve the verification in DECLARE in order to reduce the problem mainly when using a high number of constraints.

Supervisory control theory allows an automatic synthesis of supervisors that the constraints are not violated in a minimally restrictive way and ensures that this behavior is non-blocking (i.e., there is always an event sequence available to complete

an activity). Thus, new control actions may be rapidly and automatically designed when modifications, such as redefinition of constraints or tasks arrangements, are necessary. The constraint-based processes can be made to behave optimally with respect to a variety of criteria, where optimal means in minimally restrictive way. Among the criteria are safety specifications like the avoidance of prohibited regions of state space, or the observation of services priorities; and liveness specifications, as least in the weak sense that distinguished target states always remain reachable. Thus, the obtained solution using SCT is correct by construction.

Both approaches presented in this paper aims to monitor and restrict execution sequences of tasks such that constraints are not violated. Despite the control logic is built based on constraints, it does not limit the user by imposing rigid control-flow structures. In fact, the basis of these approaches is to inform users of which tasks are not allowed after an observed trace of events at run-time, and users operate with some freedom because they choose execution sequences allowed under supervision. Users can adopt this service as a guide to execute tasks with a guarantee that constraints are followed and goals are met. The approaches presented here also offers flexibility to users to choose execution sequences, and it is even possible for users to execute tasks simultaneously with no rules violations.

6 Conclusion

We argue the importance of studying the constraint-based process models. In constraint-based process is difficult to envision all possible paths and it is driven by user decisions rather than system decision. Here processes are less repetitive and the emphasis is on flexibility and user empowerment. On the other hand it is difficult to model more abstract relations between tasks when the user has many choices in each state. So, formal approaches to deal with constraint-based processes are very welcome. [15] points out that despite the many theoretical results and innovative prototypes, few of the research ideas have been adopted in commercial systems. In fact this is a limitation of the use of the SCT so far. On the other hand, DECLARE is a prototype of a workflow management system and several studies have reported their application.

References

1. Aalst, W., Hee, K., Werf, J., Kumar, A., Verdonk, M.: Conceptual model for on line auditing. Decision Support Systems 50(3), 636–647 (2011)
2. Santos, E.A.P., Francisco, R., Vieira, A.D., de F.R. Loures, E., Busetti, M.A.: Modeling Business Rules for Supervisory Control of Process-Aware Information Systems. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) BPM Workshops 2011, Part II. LNBIP, vol. 100, pp. 447–458. Springer, Heidelberg (2012)
3. Pesic, M., Aalst, W.: A Declarative Approach for Flexible Business Processes Management. In: Eder, J., Dustdar, S. (eds.) BPM Workshops 2006. LNCS, vol. 4103, pp. 169–180. Springer, Heidelberg (2006)

4. Pesic, M., Schonenberg, M.H., Sidorova, N., van der Aalst, W.M.P.: Constraint-Based Workflow Models: Change Made Easy. In: Meersman, R., Tari, Z. (eds.) OTM 2007, Part I. LNCS, vol. 4803, pp. 77–94. Springer, Heidelberg (2007)
5. Aalst, W., Pesic, M., Schonenberg, H.: Declarative workflows: Balancing between flexibility and support. Computer Science-Research and Development 23(2), 99–113 (2009)
6. van der Aalst, W.M.P., de Beer, H.T., van Dongen, B.F.: Process Mining and Verification of Properties: An Approach Based on Temporal Logic. In: Meersman, R. (ed.) OTM 2005. LNCS, vol. 3760, pp. 130–147. Springer, Heidelberg (2005)
7. Ramadge, P., Wonham, W.: The control of discrete event systems. Proceedings of the IEEE 77(1), 81–98 (1989)
8. Queiroz, M.H., Cury, J.E.R.: Modular supervisory control of large scale discrete event systems. In: Boel, R., Stremersch, G. (eds.) Discrete Event Systems: Analysis and Control, WODES 2000, pp. 103–110. Kluwer Academic Publishers (2000)
9. Pesic, M., Schonenberg, H., Aalst, W.: Declare: Full support for loosely-structured processes. In: 11th IEEE International on Enterprise Distributed Object Computing Conference, pp. 287–287. IEEE Press, New York (2007)
10. Clarke, Grumberg, O., Peled, D.A.: Model Checking. The MIT Press, Cambridge (1999)
11. Maggi, F.M., Montali, M., Westergaard, M., van der Aalst, W.M.P.: Monitoring Business Constraints with Linear Temporal Logic: An Approach Based on Colored Automata. In: Rinderle-Ma, S., Toumani, F., Wolf, K. (eds.) BPM 2011. LNCS, vol. 6896, pp. 132–147. Springer, Heidelberg (2011)
12. van der Aalst, W.M.P., Pesic, M.: DecSerFlow: Towards a Truly Declarative Service Flow Language. In: Bravetti, M., Núñez, M., Zavattaro, G. (eds.) WS-FM 2006. LNCS, vol. 4184, pp. 1–23. Springer, Heidelberg (2006)
13. Project Management Institute: A guide to the project management body of knowledge (PMBOK). An American National Standard ANSI/PMI 99-001-2008 (2008)
14. Feng, L., Wonham, W.M.: TCT: A computation tool for supervisory control synthesis. In: International Workshop on Discrete Event Systems, pp. 388–389. IEEE (2006)
15. Pesic, M., Schonenberg, H., van der Aalst, W.M.P.: Declarative workflows. In: Hofstede, A., van der Aalst, W.M.P., Adams, M., Russel, N. (eds.) Modern Business Process Automation: YAWL and Its Support Environment, pp. 175–201. Springer, Heidelberg (2010)
16. Westergaard, M.: Better Algorithms for Analyzing and Enacting Declarative Workflow Languages Using LTL. In: Rinderle-Ma, S., Toumani, F., Wolf, K. (eds.) BPM 2011. LNCS, vol. 6896, pp. 83–98. Springer, Heidelberg (2011)

Adopting Standards in Nursing Health Record – A Case Study in a Portuguese Hospital

Bruno Rocha[1] and Álvaro Rocha[2]

[1] GIMED, UFP, Pr. 9 de Abril 349, 4249-004 Porto, Portugal
bg.rocha@gmail.com
[2] LIACC, Universidade do Porto, R. Campo Alegre 1021, 4169-007 Porto, Portugal
amrrocha@gmail.com

Abstract. The investigation that is described in this paper tried to identify and acknowledge the added value created by the adoption of nursing practice standards, using a study case in a large Portuguese hospital. Following a literature review, five assumptions were proposed. Subsequently, a survey was developed and administered to a nursing staff sample of the hospital under study. Our results revealed that, except for the last, all assumptions were confirmed, leading to the conclusion that the adoption of standards is beneficial for the nursing practice.

Keywords: Nursing, Standards, Health Information Systems, Electronic Health Records.

1 Introduction

Health care is a fundamental need in society and the integration of medical information is a challenge for health information systems, to the extent that patients receive health care in different institutions which, in their turn, use multiple information systems based on different *hardware* and *software* configurations. Additionally, the health field is well known for the high number and complexity of its existing concepts [1].

The health care system is composed by organizations, which monitor health status, and individuals, the ones who receive and the ones who provide the health care [9]. Patients seek treatments that answer their needs, while permanently demanding the confidentiality of their personal information. In current medical health care, the majority of medical information pertaining to patients derives from complementary diagnostic tests that are administered as medical support tools to assure a correct pathology diagnosis. Nowadays, both analogical and digital formats support these tests, and until very recently the paper based analogical format was the preferred method. The digital format is being increasingly used [5].

The development of electronic health records has been significantly active in recent years, as well as their implementation. With the evolution of health information systems, electronic health records need to aggregate the complete clinical information

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 237–247.
DOI: 10.1007/978-3-642-36981-0_23

pertaining to a patient, allowing health care professionals to access their medical history as a whole, enhancing the communication amongst them and promoting the quality of health care [19]. This requires the existence of rules that systematize and normalize procedures and concepts, with the adoption of a common terminology for computer systems.

This common terminology is known as *standard*. These are previously agreed rules that are implemented to ensure the interoperability between computer systems. In Portugal, the nursing standard proposed by the Portuguese Ministry of Health for health care units is the ICNP (International Classification for Nursing Practice), which describes the elements of nursing practice – diagnoses, actions and results.

The main goal of this investigation was to identify and recognize the benefits connected with the adoption of standards in nursing care. Considering our goal, and a set of formulated assumptions, we chose a quantitative approach and we carried out a descriptive research – a case study, since our intention was to characterize this phenomenon. The research was developed in a Portuguese hospital and the selected data collection tool was a survey, administered to 125 nursing professionals.

2 Standards

The need to apply certain standards derives from a number of factors: the significant increase in the complexity of the health system, the growing number of health professionals and the wish to ensure an equal access to health plans, as stated by the WHO [22]. These are the factors that stimulate the adoption of electronic medical records by health care institutions [10].

The integration of electronic health records in multiple health information systems requires an aggregation of data that must be carried out with consistency and quality, and this is where the existence of standards plays a major role. According to Palhares [18], the interoperability between computer systems requires a previously defined and unambiguous structure and, most of all, the existence of data with semantics that can be interpreted by multiple systems. When two different systems follow a common standard, the communication between them and their interoperability is possible, and this enables the circulation of easily accessed information between them.

2.1 What Are They? What Is Their Purpose?

According to the definition provided by the California Healthcare Foundation [3], a standard is a set of rules that are previously agreed by organizations who intend to adopt it, rules that enable the sharing and processing of information in a uniform and consistent manner.

Standards are constantly used in our everyday lives without being noticed, as for instance in writing: a full stop in the end of a sentence; a capital letter in personal names. Another analogy can be found in traffic regulations, which state that a driver must stop when faced with a red traffic light and, in most countries, drive on the right ride of the road. Not only do these standards support our actions and communications, they also prevent errors and accidents.

This concept is applied in the health field to define rules that are capable of answering specific medical questions and providing the organizations with the means to exchange medical information, increasing the efficiency of communication and preventing errors.

For the standards to be used in practice, information systems and software development must enclose these rules and provide an efficient environment that reinforces their use, in order to ensure their reliance after their development and adjustment. This guarantees the automation of repetitive processes and the consistency across applications and, according to Sujansky & Chang [20], it can support clinical decisions in a timely manner and improve the quality of medical services.

The need to use standards in the health field derives from a number of factors [8]:

- The great diversity of sources and medical concepts (there are over 150.000 medical concepts).
- The existence of information systems that use different platforms and softwares.
- T the retrieval and exchange of information.

2.2 Nursing Standards

The development of nursing standards was stimulated by the advance of technology in the health field during the sixties. Standards "(…) promote research in nursing care, articulating available data in nursing information systems with available data in other health information systems (…)" [7] and their applicability is mandatory to sort, codify and organize every phenomenon observed in nursing care, because "(…) without classifications, nurses employ free language to express clinical decisions." [7].

2.2.1 ICNP

The ICNP (International Classification for Nursing Practice) is a classification developed by the ICN (International Council of Nurses) with the purpose of defining a scientific and unified language for nursing practice, to represent concepts and describe globally used health techniques [12]. In 1989 the ICN was called upon the need to develop an international classification, which triggered an answer in 1991 with the emergence of the ICNP project. After a long bibliographic survey and several researches on a global scale, the original ICNP project recommended the development of a terminology that was capable of describing the nursing practice while bearing in consideration a structure of terms and the definition of vocabulary [4].

According to the ICN [12], the following goals were defined for the ICNP: establishing a common language for describing the nursing practice; the ability to describe and document the nursing practice; the possibility of being used for clinical decision making; providing the vocabulary and a classification for the nursing practice, to be used when entering nursing data in computer systems; providing nursing practice data that is capable of influencing health policies.

The Alfa-version of the ICNP was published in 1996. Composed by a combination of nursing phenomena and interventions, albeit incomplete, its dissemination held the

purpose of encouraging the participation of the nursing community. Based on this version, several changes were introduced in the classification structure. New rules were defined for the selection of terms, new terms were included and a classification of nursing outcomes was also created, giving way to the Beta-version in 1999 [13].

In 2000, the ICNP was considered an official program for the professional nursing practice [4] and in 2002 the Beta-2-version was launched, offering more terms and a Portuguese version, published by the APE (the Portuguese Nursing Association). A review process followed, as well as the development of the classification system, and in 2006 its first version was published, bearing significant changes. CIPE is presently in its third version, published in 2011."

2.2.2 NANDA

A NANDA-I (North American Nursing Diagnosis Association – International) is a professional nursing care organization whose intention is to standardize the nursing terminology and to develop the vocabulary, criteria and taxonomy of nursing diagnosis. The term is defined by NANDA-I [17] as being "(…) a medical assessment of the answer given by an individual, a family or a community to real or potential health problems (…)".

NANDA-I emerged in 1973, the same year of its publication concerning nursing diagnosis. Founded in 1982 only [17], when an interested group decided to focus on the development and validation of nursing care diagnosis, it was only in the 90's that their work started to be internationally acclaimed. In 2002, NANDA (North American Nursing Diagnosis Association) decided to change its name to NANDA-I [21].

Nurses use the diagnosis developed by NANDA-I to recognize human reactions to risk, illness and lesion or health promotion, along with their ability to apply a critical judgement and reasoning and in order to give suitable answers to these reactions [15].

Lundberg et al. [14] states that the nursing diagnosis included in NANDA-I serve as a care documentation base for nurses, being used in the evaluation and planning of medical care and to categorize the condition of the patient.

2.2.3 NIC

The NIC (Nursing Interventions Classification) describes interventions performed by nursing professionals. According to Bulechek et al. [2], the NIC is a nursing practice oriented research kit oriented, which enables the analysis of the intervention impact in patient outcomes. Nursing interventions include direct and indirect care to individuals, family and communities.

The first edition of this classification system was published in 1992 and it is currently in its sixth edition. The NIC proposes a unique label and numeric code for each intervention, favouring the communication between multiple computer systems. It can be used in any clinical situation [2].

The adoption of this classification system in electronic health records enables an effective selection of nursing interventions and their subsequent communication to other health professionals. Moreover, it enables the assessment of the effectiveness and of the expenses pertaining to the allocation of resources for nursing care. This classification improves the learning of clinical decisions by nurses at the point of care [14].

2.2.4 NOC

Nurses document the outcomes of their interventions, but the absence of a common terminology made it impossible to aggregate data and to analyse the effects of nursing practice. The development of this common terminology began in 1991 but it was in 1997 that the first NOC (Nursing Outcomes Classification) edition was published, having already reached its fourth version. The development team tries to focus on the review of the connections between outcomes (NOC), interventions (NIC) and diagnosis (NANDA-I) [16].

Moorhead et al. [16] add that with this classification nurses can follow any variations in the obtained outcomes in the course of time, allowing them to understand the problem of the patient and its diagnosis. It is organized according to taxonomies, divided in three levels: domain, category and specialty. Each outcome has a label, a five-point scale of states, a list of associated indicators and support references. All this information is encoded in order to be available in the electronic health record.

This classification can be implemented both as a support tool to nursing care planning and in the measurement of outcomes, when patients are admitted, discharged or transferred.

3 Research Design and Methodology

We will now present an exploratory study involving the adoption of standards in day to day professional nursing activities of a Portuguese hospital, as well as the resulting benefits.

3.1 Object of Study

Fortin [6] identifies two scientific data collection methods: the quantitative and the qualitative method. Choosing the method is a matter of personal taste and philosophy on the part of the researcher, but it also depends on the investigation question.

We selected a quantitative data collection technique – a survey administered as a questionnaire. We applied the descriptive research method – a case study, because the purpose of our investigation strategy was to analyse a real situation in all its complexity, and after the information was collected we were able to perform a correlational study to analyse the connections between the variables involved [11].

3.2 Investigation Objectives

The main purpose of this investigation was to identify and acknowledge the added value created by the implementation of nursing practice standards in a Portuguese hospital.

More concrete goals concerned the analysis of the challenges imposed by these standards for the nursing practice:

- Understanding the opinion of professional towards current standards.
- Confirming the existence of health care improvements.
- Understanding if the standards favour the interoperability between computer systems.
- Confirming if the communication between health care professionals is enhanced.
- Identifying the implementation of mobile and portable health records and understanding whether they offer the nursing professional the basic features, as well as their benefits and/or disadvantages.

3.3 Investigation Methodology

Every investigation project that adopts the quantitative methods presupposes the identification of variables. In our study, we determined that our independent variables would bear a characterizing nature, because "(…) the pre-existing characteristics of the participants in a study (…) generally involve demographic data." [6], including: Sex; Age; Qualifications; Specialty; Years of service.

Subsequently, Pearson's correlation coefficient (or Pearson's ρ) was estimated between each question (dependent variables) and each one of our independent variables, with the purpose of measuring the degree and the direction of the observed correlations (positive or negative).

Being a linear correlation coefficient it assumes values between -1 and 1only:

- $\rho = 1$ represents a perfect and positive linear correlation, that is, when one of the variable increases the other one also increases.
- $\rho = 0$ indicates the absence of a linear correlation between the two variables.
- $\rho = -1$ represents a perfect but negative linear correlation, that is, when one of the variable increases the other one decreases.

We adopted the following scale for Pearson's linear correlation:

- $0 \leq \rho \leq 0{,}19$ for a very weak correlation
- $0{,}20 \leq \rho \leq 0{,}39$ for a weak correlation
- $0{,}40 \leq \rho \leq 0{,}59$ for a moderate correlation
- $0{,}60 \leq \rho \leq 0{,}79$ for a strong correlation
- $0{,}80 \leq \rho \leq 1$ for a very strong correlation

First, we performed a theoretical survey for the history of nursing practice standards, in order to understand their present state. Next, we developed, validated and administered a survey to a group of nursing professionals from a Portuguese hospital, in order to understand their opinion towards the influence and importance of standards in their professional occupation.

We performed a preliminary test to our survey because, in the opinion of Fortin [6], data collection tools must be tested for their effectiveness and value, to quantify the time needed to administrate the survey to a small sample from the target population and to ascertain if all the questions lead to relevant data. Therefore, a preliminary test was administered to 7 nursing professionals whose characteristics were similar to our respondents. No clarifications were needed and no alterations

were suggested by the participants. Accordingly, our survey remained unchanged after the preliminary test.

After the data collection, information needs to be treated in order to improve its interpretation since"(…) the results involve numeric information resulting from a statistical analysis of the data collected from the participants (…)" [6]. All the questions were treated taking into account their central tendency, i.e., we used the arithmetic mean, the mode and the median, as well as their dispersion, including standard deviation and variance. This treatment was performed with Microsoft® Excel.

3.4 Investigation Question and Assumptions

In our study, we tried to reach an answer to our investigation question:

How does the adoption of standards impact the nursing practice?

This investigation question led us to formulate the following assumptions, based on our bibliographic review and personal knowledge:

A_1: Health care institutions implement a nursing standard.
A_2: The nursing standards employed by the institution favour the interoperability between computer systems.
A_3: A standardized electronic health record system improves nursing care.
A_4: Computer systems that follow nursing standards allow for a complete collection of medical data.
A_5: Health units adopt mobile and portable electronic health records.

3.5 Definition of the Medium and the Sample

Our intention was to study at least one Portuguese hospital, which led us to contact two possible hospitals. These hospitals were chosen for being conveniently close, because we had privileged contacts in both institutions and because they were large organizations. Being a highly controlled environment, we presented our study to the ethics board of each hospital and asked for the necessary permissions to access and request the cooperation of nursing professionals during their work hours. We provided three different communication channels (email, mobile number and address) where we could be reached in the event of a response. One of the selected hospitals did not reach a final answer and our attention turned to the hospital that had given us a positive answer.

Beck & Polit [1] define population as "(…) the aggregate of cases that conform to designated criteria." They also define the target population as being "(…) the entire aggregate of cases in which the researcher is interested." and the accessible population as "(…) the target population cases that are accessible as subjects for a study.".

The hospital in our study serves approximately 200 thousand users has close to 500 beds available in all departments. The surgery wing has 12 operating rooms, with conventional and laparoscopic equipment. There are approximately 70 rooms for

outpatient medical and nursing appointments and 1 general and gynaecology/ obstetrics emergency service. The hospital employs almost 1850 professionals, and nearly 800 belong to the nursing staff.

Our sample included 109 nursing professionals from the different medical specialities offered by the hospital under study, from a total of 125 nursing professionals who participated. Sixteen (16) participants were excluded because they did not answer the questions or answered only a few questions; the unreliability of illegible calligraphy was also an excluding factor.

3.6 Data Collection

According to Fortin [6], there are several ways to collect data from participants and it is up to the researcher to determine the best method, bearing in mind the purpose of the study, the investigation questions and formulated hypothesis or assumptions. With this in mind we chose the survey as our data collection tool, which, in the words of the same author, is "the method most often used by researchers (…). The purpose of the survey is to draw up factual information concerning known events or situations, attitudes, beliefs, knowledge, feelings and opinions.".

Our strategy included the Boolean measure (Table 1) and the five point Likert scale measure (Table 2).

Table 1. Boolean measure

Yes	No	I don't know
1	2	3

Table 2. Measured in a five point Likert Scale

Strongly disagree	Disagree	No opinion	Agree	Strongly agree
1	2	3	4	5

The survey included twenty eight questions, organized in six parts:

Part I involved five questions, which allowed us to characterize our sample.

In part II we included three open questions and six closed questions, which tried to capture the perspective of the respondents regarding the current state of the standards adoption.

Part III tried to assess the interoperability of computer systems and involved two closed questions.

Part IV evaluated whether the application of standardized electronic health records was beneficial, both for the patient and the nursing professional and included four open questions.

Part V concerned the use of computer systems and the extent to which they met the needs of nursing professionals, and involved four open questions.

Part VI intended to evaluate whether the computer system was capable of supporting the mobile and portable nature of nursing electronic health records, and if this is beneficial or disadvantageous. This part included four open questions.

One hundred and twenty five (125) surveys were handed out during the month of February 2010, with a response rate of 97,6% and a validation rate of 87,2% (109 surveys were considered valid).

3.7 Ethical Considerations

This investigation was developed in compliance with the ethical principles underlying the investigation practice. Fortin [6] contends that the investigative practice must be carried out in compliance with the rights of the individual, regardless of the subject under study. Therefore, this survey was answered anonymously, so none of the respondents could be recognized in the results, by the researcher or the readers. The hospital under study also exercised his right to anonymity.

4 Discussion and Conclusion

In the health field, change is imperative to ensure answers to patients needs. One of the primary challenges to the modernization of computer systems pertains to strategic investment, the rationalization of resources to improve efficiency and quality. These information systems form an essential source of knowledge and constitute decision support tools, capable of increasing the security and the efficiency of health care related processes.

Existing obstacles prevent a large scale application of standards that could improve efficiency and refine the nursing practice, namely the fact that each organization develops and implements its own rules, which renders the sharing of patient data records with other institutions impracticable.

In the course of this investigation, we tried to answer our research question (How does the adoption of standards impact the nursing practice?) and to confirm the assumptions defined for the hospital under study:

A1: Health care institutions implement a nursing standard.

We can conclude that they do. According to our results, 92,7% of the respondents confirmed that they worked with a standard. When we estimated Pearson's correlation we observed that all variables correlated very weakly with the obtained answers.

A2: The nursing standards employed by the institution favour the interoperability between computer systems.

This interoperability was observed within the hospital under study (60,6% agreed), but the exchange of information between institutions did not occur (33,9% disagreed). In the first case, Pearson's linear correlation revealed higher values for "Age" and "Years of service", but these were not significant. In the second case, the variable "Specialty" displayed the highest value, but it did not reach a significant value.

A3: A standardized electronic health record system improves nursing care.

A standardized electronic health record system did indeed improve the health care provided by nursing professionals (81,7% agreed). These nursing care improvements include a better continuity of care, a lower probability of error, a greater access to information, the enhanced communication amongst professionals and the uniformization of nursing care plans. Pearson's linear correlation leaned towards zero in every variable.

A4: Computer systems that follow nursing standards allow for a complete collection of medical data.

We observed that computer systems were ready to perform a complete collection of patient medical data (58,7% agreed). Pearson's linear correlation was very weak only for "Age", every other variable under study displayed a weak correlation.

A5: Health units adopt mobile and portable electronic health records.

We can conclude that this is not the case, having obtained 37,6% negative answers and a higher percentage of "I don't know" answers (42,2%). Every variable under study revealed a significantly weak linear correlation.

With the exception of the latter, all assumptions were confirmed for the hospital under study. At the same time, the main goal of this investigation was achieved: to identify and acknowledge the added value behind the adoption of nursing care standards, in the Portuguese hospital under study.

Given the nature of this investigation (a case study), in the future it will be advisable to apply it in other health care units, to obtain more representative results.

References

1. Beck, C., Polit, D.: Fundamentos de Pesquisa em Enfermagem, 7th edn. Artmed (2011) ISBN 9788536325453
2. Bulechek, G., Butcher, H., Dochterman, J.: Nursing Interventions Classification (NIC), 5th edn. MOSBY Elsevier (2008) ISBN 9780323056045
3. CHF, Clinical Data Standards Explained. Fact Sheet, California Healthcare Foundation (2004)
4. Cho, I., Park, Y.: Evaluation of the Expressiveness of an ICNP-based Nursing Data Dictionary in a Computerized Nursing Record System. Journal of the American Medical Informatics Association 13(4), 456–464 (2006), doi:10.1197/jamia.M1982
5. Dolin, R., et al.: HL7 Clinical Document Architecture, Release 2. Journal of the American Informatics Association 13(1), 30–39 (2006), doi:10.1197/jamia.M1888
6. Fortin, M.F.: Fundamentos e Etapas no Processo de Investigação. Lusodidacta (2009) ISBN 9789898075185
7. Gaidzinski, R., et al.: Diagnóstico de Enfermagem na Prática Clínica. Artmed (2008) ISBN 9788536311807
8. Goehle, J.: Technology of Tomorrow: The Implementation of Electronic Health Records, University of Buffalo (2010)
9. Gonçalves, J., Rocha, A.: A decision support system for quality of life in head and neck oncology patients. Head & Neck Oncology 4(3), 1–9 (2012), doi:10.1186/1758-3284-4-3

10. Health Affairs, Health Policy Brief. Health Foundation Inc. (2010)
11. Hill, M., Hill, A.: Investigação por Questionário, 2nd edn. Sílabo, Lisboa (2005) ISBN 9726182735
12. ICN, International Council of Nurses, ICN (2010), http://www.icn.ch (accessed: November 25, 2011)
13. Leal, M.: A CIPE e a Visibilidade da Enfermagem. Lusociência, Lisboa (2008) ISBN 9789728930202
14. Lundberg, C., Warren, J., et al.: Selecting a standardized terminology for the electronic health record that reveals the impact of nursing on patient care. Online Journal of Nursing Informatics 12(2) (2008)
15. Lunney, M.: Critical need to address accuracy of nurses' diagnoses. The Online Journal of Issues in Nursing 13(1) (2008), doi:10.3912/OJIN
16. Moorhead, S., Johnson, M., Maas, M., Swanson, E.: Nursing Outcomes Classification (NOC), 4th edn. Elsevier (2008) ISBN 9780323054089
17. NANDA-I Nursing Diagnoses: definitions & classification. Willey-Blackwell, Oxford (2008)
18. Palhares, P.: My PEPWeb: Sistema de Prontuário Electrónico Pessoal através da World Wide Web. Undergraduate Monograph in Computer Science, Federal University of Ouro Preto, Brazil (2010)
19. Rocha, A.: Evolution of Information Systems and Technologies Maturity in Healthcare. International Journal of Healthcare Information Systems and Informatics 6(2), 28–37 (2011), doi:10.4018/jhisi.2011040103
20. Sujansky, W., Chang, S.: The California Clinical Data Project: A Case Study in the Adoption of Clinical Data Standards for Quality Improvement. Journal of Healthcare Information Management 20(3), 71–78 (2004)
21. Von-Krogh, G.: An examination of the NANDA International taxonomy for domain completeness, ontological, homogeneity, and clinical functionality. International Journal of Nursing Terminology and Classification 19(2), 65–75 (2008), doi:10.1111/j.1744-618X.2008.00083.x
22. WHO, International Classification of Diseases (ICD). World Health Organization (2011), http://www.who.int/classifications/icd/en/ (accessed: September 14, 2011)

The Relationship between Portal Quality and Citizens' Acceptance: The Case of the Kuwaiti e-Government

Adel M. Aladwani*

Department of QM & IS, College of Busines Administration, Kuwait University, Kuwait
adel.aladwani@ku.edu.kw

Abstract. This research-in-progress article tries to examine Kuwaiti citizens' perceptions of e-Government quality and its influence on their decision to use (or not to use) the official Kuwaiti portal. Specifically, it tries to examine an attitudinal model linking three groups of e-Government quality attributes and e-Government use as measured by number of visits and transactions. The statistical analyses of responses collected from a sample of Kuwaiti citizens highlight the importance of e-Government quality for inducing citizens' use.

Keywords: electronic government, quality, acceptance, use, Kuwait.

1 Introduction

In the past few years, electronic government (e-Government), or the use of advanced wire-based and wireless telecommunication technologies to strengthen the relationship with the public, has attracted much attention from many researchers across the globe. In particular, several researchers have called for putting more attention on examining the relationship between e-Government quality and acceptance [1-6]. This increased interest stems from the fact that the issue of quality has become a strategic imperative for many of today's technology initiatives [7] including e-Government websites. Unfortunately, only a handful studies focused on understanding e-Government quality issues in the Middle-east. This study tries to fill part of this gap in past research.

2 e-Governments in Kuwait

A decade ago, decision makers in the State of Kuwait noted that public workers as well as citizens should learn how to work together within a new milieu which is different than anything they have accustomed to interact within before. As a result, the Council of Ministers in Kuwait decided in the year 2000 to form a new technology committee to kick off the Kuwaiti e-Government efforts to assume responsibility of the e-Government development efforts.

Kuwaiti administrators selected an easy to remember URL address (www.e.gov.kw) to identify to the official e-Government portal. A year later, a

* This study was supported by Kuwait University (Grant No. IQ01-11).

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 249–254.
DOI: 10.1007/978-3-642-36981-0_24

decree was issued to establish the information technology agency, which has become responsible for all efforts related to the Kuwaiti e-Government project. Table 1 summarizes some of the basic content facts about the Kuwaiti e-Government portal.

An important issue facing the newly established Kuwaiti e-Government is the challenge of how to convince Kuwaiti citizens to accept this national initiative [8]. Previous research [9] has shown that the role of attitudes in an information technology context is rather complex issue that deserves close attention from technology researchers. As the success or failure of a given e-Government project is determined by the extent to which citizens' appreciate the quality of the portal [10], it becomes a reasonable choice for us to investigate Kuwaitis' evaluations of e-Government quality. Aladwani [11] found that website quality is an important driver of one's intention to accept an information technology application in Kuwait. However, there isn't much research studying Kuwaiti citizens' quality perceptions of e-Government services. Next, the author describes the results of an initial study focusing on this issue.

Table 1. e-Government in Kuwait (basic facts)

	e-Gov. Content	**Overall Gov. Content**
No. of Gov. bodies	36	64
No. of pages	~115,000	~622,000
No. of Arabic pages	~65,000	~420,000
No. of online services	49	419

Source: [8]

3 e-Government Quality and Acceptance

In the information systems field, the study of website quality has attracted some attention from researchers. For example, Aladwani and Palvia [12] developed an instrument to measure technical, content, and aesthetic website quality. Parasuraman et al. [13] developed an e-service quality instrument. Loiacono et al. [14] proposed the WebQual™ instrument based on Loiacono's dissertation work. Also, Barnes and Vidgen [15] proposed a multi-item website quality scale, and so did Palmer [16] and Aladwani [17], among others.

So far, most of the attention to analyzing the influence of website quality on certain organizational outcomes has been conducted in business settings. For example, Aladwani [18] studied the relationship between commercial websites' quality and acceptance. Therefore, several researchers have called for putting more attention on examining the relationship between e-Government quality and acceptance [1-6]. Still, the study of e-Government quality has thus far focused on content analyzing some of the portals, which are located mostly in developed countries [19]. Little, if any, studies have investigated the relationship between e-Government quality and acceptance in developing countries; rare exceptions are the studies by Aladwani [20, 21]. Pragmatically, users' perceptions of e-Government quality can help facilitating the achievement of many goals among which is e-Government acceptance.

Based on the above discussion, this investigation proposes the following question for possible answer: is there a relationship between Kuwaitis perception of e-Government quality and their acceptance of the same?

4 Methodology

A research instrument was borrowed from previous research for the purpose of this study. It consisted of two groups of items: demographics and likert-like perceptual items. The outcome variable, e-Government usage, which was borrowed from relevant research [22], was assessed using two items: visit frequency (anchored on a four-point scale: (1) at least once every few months, and (4) at least once a day) and number of transactions (anchored on a five-point range: (1) conducted no transactions, and (5) completed 7 or more transactions).

The predictor variable in this study, e-Government quality, was measured using a modified version of Aladwani & Palvia's [12] instrument, which consists of 20 items covering technical (security, ease of navigation, ease of accessing the site, search facilities, availability, valid links, speed of page loading, personalization, customization, and interactivity), content (usefulness, completeness, clarity, currency, conciseness, and accuracy), and aesthetic quality (attractiveness, organization, proper use of fonts, proper use of colors, and proper use of multimedia) attributes. The questions in the scales ranged from (1) not important to (7) very important.

The instrument was distributed to fifty-five Kuwaitis who have previous knowledge of the studied portal; and responses were received from all the participants. The mean age of the respondents in the sample is approximately thirty-two years (Standard Deviation = 5.05). Around 71 percent of the respondents are males. The mean Web experience score (on a 3 point scale where 1 = low, 2 = moderate, and 3 = high) is 2.05 (Standard Deviation = 0.68).

5 Data Analysis and Results

Table 2 summarizes respondents' importance perceptions of the different e-Government quality attributes. Overall, the table shows that the scores differ from one group of features to another. For example, technical quality features exhibit higher mean scores than aesthetic quality features. It is true that we cannot infer solid conclusions from this information but it can offer hint to the preferences of the participants in the study.

To test the proposed structural model and research hypotheses, the author used Partial Least Squares (PLS), an advanced multivariate technique for simultaneously assessing a series of regression equations in causal models. Figure 1 reports path estimates and squared multiple correlations of our structural model. Significance was calculated using the bootstrapping technique. A significance level of less than or equal to 0.05 was selected for examining the p-value, which is the probability that the resulting statistic in the investigation could have happened by pure coincidence.

Table 2. Perceived importance of e-Government quality

		Perceived Importance
Technical	Perceived security	5.55
	Ease of navigation	5.13
	Search facilities	5.20
	Availability	5.44
	Valid links	4.78
	Personalization	4.75
	Speed of page loading	5.07
	Interactivity	4.58
	Ease of accessing the site	4.71
Content	Usefulness	4.96
	Completeness	4.69
	Clarity	4.69
	Currency	4.40
	Conciseness	5.05
	Accuracy	4.51
	Attractiveness	4.38
Aesthetic	Organization	4.13
	Fonts	4.33
	Colors	4.09
	Multimedia	4.09

Source: [20]

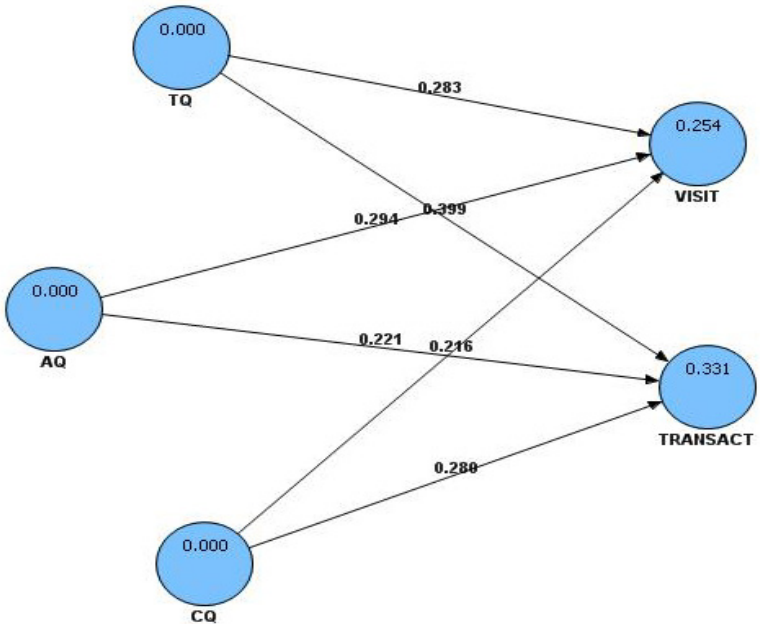


Fig. 1. Results of the PLS test

Overall, the results show that e-Government quality variables explained 25% and 33% of the visitation and transaction outcomes respectively. The findings also indicate that technical quality (TQ) has the strongest effect on number of transactions and aesthetic quality (AQ) has the strongest influence on visitation frequency.

6 Conclusions

The goal of this preliminary investigation was to explore Kuwaitis' perceptions of e-Government quality and its effect on e-Government acceptance. More specifically, it examined the relationship between technical, content, and aesthetic quality on e-Government visitation frequency and number of transactions. The paper was motivated by the fact that most of the attention to examining the influence of website quality on individuals' acceptance in Kuwait has been carried out in commercial settings.

Overall, the empirical study reveals that e-Government quality is strongly related to e-Government acceptance. More specifically, the results show that technical quality (security, ease of navigation, ease of accessing the site, search facilities, availability, valid links, speed of page loading, personalization, customization, and interactivity) has the highest impact on citizens' decision to transact with the e-Government portal. The results further show that aesthetic quality (attractiveness, organization, proper use of fonts, proper use of colors, and proper use of multimedia) is the main determinant of visits frequency. Finally, the findings indicate that content quality (content usefulness, completeness, clarity, currency, conciseness, and accuracy) affects transactional use more than visits frequency.

Future research may need to draw a larger sample size and retest the proposed relationships to get better understanding of e-Government acceptance in Kuwait.

References

1. Wangpipatwong, S., Chutimaskul, W., Papasratorn, B.: Quality Enhancing the Continued Use of E-Government Web Sites: Evidence from E-Citizens of Thailand. International Journal of Electronic Government Research 5, 19–35 (2009)
2. Wang, Y.-S.: A validation of the User-Perceived Web Quality instrument for e-Government context. In: 10th WSEAS International Conference on Systems, Athens, Greece, vol. 5, pp. 703–709 (2006)
3. Tan, C.W., Benbasat, I., Cenfetelli, R.T.: Building citizen trust towards e-Government services: Do high quality websites matter? In: HICSS 2008, Hawaii International Conference on System Sciences (2008)
4. Papadomichelaki, X., Mentzas, G.: A Multiple-Item Scale for Assessing E-Government Service Quality. Springer, Berlin (2009)
5. Magoutas, B., Mentzas, G.: Refinement, Validation and Benchmarking of a Model for E-Government Service Quality. Springer, Berlin (2009)
6. Garcia, A.C.B., Maciel, C., Pinto, F.B.: A quality inspection method to evaluate e-government sites. Electronic Government, 198–209 (2005)

7. Aladwani, A.M.: Implications of some of the recent improvement philosophies for the management of the information systems organization. Industrial Management & Data Systems 99, 33–39 (1999)
8. Aladwani, A.M.: A preliminary study of citizens' attitudes toward the Kuwaiti e-Government portal. In: ISDSI 2010, Gurgaon, India (2010)
9. Aladwani, A.M.: A deeper look at the attitude-behavior consistency assumption in information systems satisfaction research. Journal of Computer Information Systems 44, 57–63 (2003)
10. Aladwani, A.M.: Determinants of e-Government success in Kuwait. In: GITMA 2011, Las Vegas, Nevada (2011)
11. Aladwani, A.M.: Online banking: A field study of drivers, development challenges, and expectations. International Journal of Information Management 21, 213–225 (2001)
12. Aladwani, A.M., Palvia, P.C.: Developing and validating an instrument for measuring user-perceived web quality. Information & Management 39, 467–476 (2002)
13. Parasuraman, A., Zeithaml, V.A., Malhotra, A.: E-S-QUAL: A multiple-Item scale for assessing electronic service quality. Journal of Service Research 7, 213–233 (2005)
14. Loiacono, E.T., Watson, R.T., Goodhue, D.L.: WebQual: An instrument for consumer evaluation of Web sites. International Journal of Electronic Commerce 11, 51–87 (2007)
15. Barnes, S.J., Vidgen, R.T.: An integrative approach to the assessment of e-commerce quality. Journal of Electronic Commerce Research 3, 114–127 (2002)
16. Palmer, J.W.: Web site usability, design, and performance metrics. Information Systems Research 13, 151–167 (2002)
17. Aladwani, A.M.: The development of two tools for measuring the easiness and usefulness of transactional Web sites. European Journal of Information Systems 11, 223–234 (2002)
18. Aladwani, A.M.: An empirical test of the link between web site quality and forward enterprise integration with web consumers. Business Process Management Journal 12, 178–190 (2006)
19. Kuzma, J.M.: Accessibility design issues with UK e-government sites. Government Information Quarterly 27, 141–146 (2010)
20. Aladwani, A.M.: A cross-cultural comparison of Kuwaiti and British citizens' views of e-Government interface quality. Government Information Quarterly 30, 74–86 (2013)
21. Aladwani, A.M.: Web Site Usability: A cultural perspective. In: AMCIS 2003. Proceedings of AMCIS 2003, Tampa, Florida, USA, paper 266 (2003)
22. Aladwani, A.M.: A contingency model of citizens' attitudes toward e-Government use. Electronic Government, an International Journal 10, 68–85 (2013)

Knowledge Management Framework for Six Sigma Performance Level Assessment

Jevgeni Sahno, Eduard Sevtsenko, and Tatjana Karaulova

Tallinn University of Technology, Department of Mechanical Engineering,
Ehitajate tee. 5, 19086 Tallinn, Estonia
Jevgeni.Sahno@gmail.com,
{Eduard.Sevtsenko,Tatjana.Karaulova}@ttu.ee

Abstract. With the rapid growth of competition in the market, the companies have to guarantee customers a reliable, sustainable and quality proofing production system. In this paper we consider a KM framework that extracts "Six Sigma" knowledge on the basis of the data gathered from production facilities. The KM framework enables to assess the performance of a company's production and quality system by sigma value. The result will help the company to select a new development strategy in order to increase the profitability and customer satisfaction. KM framework includes well known tools like PDM, ERP system, PDM-ERP middleware and DM. The core of our framework is the DM that combines production route card data, Faults Classification standard DOE-NE-STD-1004-92 and the data from FMEA table. The combination and application of different tools and methods in the general KM framework allows the data flow between different systems, analysis of production operation and the failures occurring in the production process.

Keywords: Knowledge Management (KM), Product Data Management (PDM), Enterprise Resource Management (ERP), PDM-ERP middleware, Data Mart (DM), Production Route (PR), Faults Classification, Failure Mode and Effect Analysis (FMEA), Six Sigma, Sigma performance level.

1 Introduction

The internal issues that many manufacturing companies face today are often surround the idea that companies know the problems that lie in front of them, but they do not understand the root causes spending much time on routine work. The Pareto principle states that roughly 80% of the problems are called by 20% of the causes [1]. Hence, all problems in manufacturing start from small causes, e.g., low labor qualification and non-competent management, unbearable working conditions and old technologies applied which lead to the bad quality of product. As the result, it takes to the loss of customers' expectations and consequently position on the market. Today, the release of quality product means consistency. A customer who is satisfied by the first buying experience needs to be satisfied again. Going below is unacceptable and may ruin the business. Most customers care about quality willing to pay more than average market price to obtain product and see "the extras" as worth the additional expense [2].

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 255–267.
DOI: 10.1007/978-3-642-36981-0_25

To survive in the competitive market companies should be expedient in technological resources and they should be able to demonstrate to a customer the quality of products and a functional quality proofing production system [3, 4]. Here we propose a Knowledge Management (KM) framework that consists of well-known Information System (IS) tools. We also present a novel Data Mart and the advantages that can be achieved after its implementation in manufacturing. Our framework enables an effective and less time-consuming way to store Production Route (PR) data and reuse it for new products. Also it allows to find the root of the problems that leads to the poor quality of a product during production and to eliminate them by using the Fault Classification standard DOE-NE-STD-1004-92 and Failure Mode and Effect Analysis (FMEA) methodology. New FMEA approach of Severity, Occurrence and Detection rating assessment through ERP system will show RPN value for each operation in online mode. On the basis of the collected data and specific algorithm, our DM will demonstrate us a company's "dashboard" that indicate a sigma performance level for every product type and common production system.

2 Literature Review

Organizational knowledge is now recognized as a key resource and a variety of perspectives suggest that the ability to marshal and deploy knowledge dispersed across the organization is an important source of organizational advantage [5]. KM is performance of the activities involved in discovering, capturing, sharing, and applying knowledge in terms of resources, documents, and people skills to enhance, in a cost-effective fashion, the impact of knowledge on the company´s goal achievement [6]. KM initiatives in organizations are becoming important and firms are making significant Information Technology (IT) investments [7]. We apply in KM framework well known IS tools that accept data resources as input and process them into information products as output [8]. With the help of IS technology, a company can become competitive in all phases of its customer relationships. [9].

2.1 Concepts and Tools Applied in Knowledge Management Framework

PDM System. The software that manages product data of design files generated by Computer Aided Design (CAD) systems, it allows standardizing items, storing into repository and controlling document files, maintaining Bills of Materials (BOM) and document revision, and displaying relationships between parts and assemblies [10].

ERP System. A cross-functional enterprise system driven by an integrated suite of software modules that support the basic internal business processes of a company and the systems that serves as a platform for trans-organizational data management [11].

PDM-ERP Middleware. A standardized communication interface between PDM and ERP systems that allows designing and managing of easily adaptable workflows to exchange data and integrate processes between these two systems [12].

Data Mart. A repository of data gathered from Operational Data Stores (ODS) and other sources. The goal of the DM is to meet the specific demands of a particular group of knowledge in terms of analysis, content, presentation, and ease-of-use. DM present the data in terms that are familiar or in a format user want to see [13].

Implementations of DM in several domains such as commerce, telecommunication and medicine have been thoroughly researched. It was proven to be useful and effective in the particular application domain of clinical research in heart surgery in Germany [14]. A similar research has not been done in manufacturing, despite of the potential benefits. Different reasons can be outlined. The majority of researchers in manufacturing are not familiar with DM methods and tools; many IT researchers are not familiar with manufacturing; the few researchers that have skills in both DM and manufacturing area may not have an access to manufacturing enterprise data [15].

2.2 Concepts and Tools Applied in Data Mart

Our DM that will play the role of a company's "dashboard", that is similar to the one that a pilot has in the cockpit that describe the performance of an airplane. A business or production system may have similar indicators that define the process excellence. DM will collect the raw data from ERP system, processing them and presenting sigma performance level in real time - on the basis of data collected for the previous day.

Production Route Card. A card that gives the detail of an operation to be performed in a production line, it is used to instruct the production people to take up the production work. The content and formats of the PR card vary from a company to company. In general it contains: an item and quantities to be produced; production time; dimensions; any additional information that may be required by the production worker. PR card traces the route to be taken by a job during a production process [16].

Failure Mode and Effect Analysis Methodology. For the past few years, companies have been trying to enhance the reliability of their products to grasp such opportunity for business development. Therefore FMEA was born [17]. It is a systematic method of identifying and preventing product and process problems before they occur. It is focused on preventing defects, enhancing safety and increasing a customer's satisfaction. The risk of a failure and its effects are determined by three factors:

- **Severity (S)** – the consequence of the failure that should occur during process.
- **Occurrence (O)** – the probability or frequency of the failure occurring.
- **Detection (D)** – failure being detected before the impact of the effect realized.

Every potential failure mode and effect is rated in three factors on a scale ranging from 1 to 10. By multiplying the rating for the three factors (S×O×D), a Risk Priority Number (RPN) will be determined for each potential failure mode and effect. The RPN will range from 1 to 1000 for each failure mode or operation. It is used to rank the need for corrective actions to eliminate or reduce the potential failures [18].

All FMEAs are team based and the purpose of an FMEA team to bring a variety of perspectives and experience to the project. The team should be made of five to nine

members. All team members must have some knowledge of group behavior, must be cross-functional and multidiscipline, to handle the problem to be discussed. [19].

Faults Classification for a Machinery Enterprise. Reliability engineering is dealing with an analysis of the causes of the faults in factories. In our case study we have developed a faults classification based on DOE-NE-STD-1004-92 standard shown in Figure 1. The basic reason for using the standard is to investigate and report the causes of occurrences to enable the identification of corrective actions adequate to prevent a recurrence and thereby protect public health and safety, the workers and the environment [20]. We have adapted and modified the standard qualifier for machinery enterprises. There are seven major cause categories that have its subcategories.

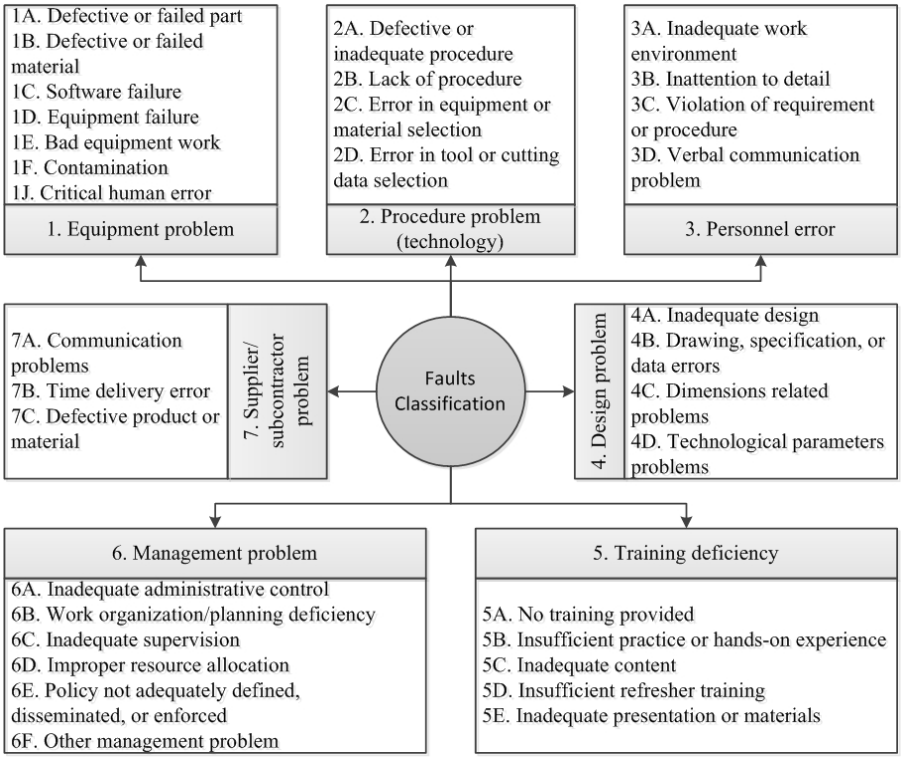


Fig. 1. Faults classification for a machinery enterprise

Six Sigma Overview. The Six Sigma method is a project-driven management approach to improve the organization's products, services, and processes by reducing defects in the organization. It is a business strategy that focuses on improving customer requirements understanding, business systems, productivity, and financial performance. Dating back to the mid of 1980s, applications of the Six Sigma methods allowed many organizations to sustain their competitive advantage by integrating their knowledge of the process with statistics, engineering, and project management

[21]. Motorola was the first company who launched a Six Sigma project in the mid-1980s [22]. Today Six Sigma has made a huge impact on industry and yet the academic community lags behind in its understanding of this powerful strategy [23].

Literature Review Summary. By the author point of view, the mentioned above concepts and tools were never used in one general framework and DM. Usually those tools were used separately for fulfillment of assigned tasks. In our paper we will combine all of them and demonstrate the benefits which company will get after its implementation.

3 Case Study

In this paper our case study will consist of two steps. The first step represents a general framework. The second step describes the functionality of our DM, its benefits and the process of data collection from ERP system to DM, its real-time data processing (data from the previous day) and sigma performance level assessment.

3.1 Step 1. Knowledge Management Framework Description

The presented KM framework shown in Figure 2 starts from a CAD system where a design engineer creates a new Item and/or a BOM structure - product. Along with the CAD models and drawings, the engineer defines an item data in the PDM system. The item data contains different attributes that are "packed" under a general Designation Code, for example - ("XYZ"). The Designation Code ("XYZ") - is an item key that is used to logically connect the data between the PDM and ERP systems. Each specific item and/or a group of items have its own unique Designation Code.

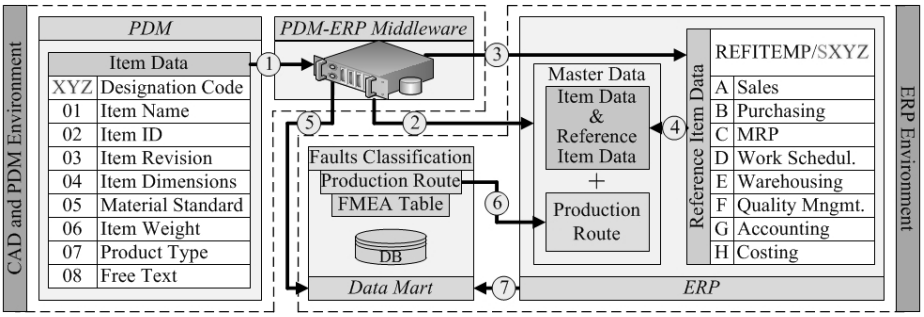


Fig. 2. Knowledge Management framework

The PDM-ERP Middleware transfers Designation Code and the accompanying the Item Data from the PDM (arrows labeled by 1 and 2), finds a Reference Item Code by a matching Designation Code (along with the set of Reference Item Data from ERP) and copies them into ERP Master Data (arrows 3 and 4). The final step (arrows 5 and 6) describes a similar process where the middleware finds PR in the DM (matched by

a Reference Item Code) for the given item and copy it into ERP Master Data [24]. Finally, the PR is released into a production floor in the form of a PR card.

Usually the Severity, Occurrence and Detection ratings in FMEA are assessed in a team, but from the author point of view this is some kind of "tricky" process because the team members can be differently educated, they can have different experience and they cannot know everything that is going on in product life cycle. Also during this process could be different opinions and argues that makes it difficult to make right decision. Therefore, in our case study we are going to monitor the changes of those ratings in online mode (based on data from the last day) to assess the sigma performance level of product and common production system. Arrow 7 in Figure 3 represents the data transfer from ERP system that was collected from production floor to DM. This data is collected in order to assess the rating of Severity (how can be expensive damaged component), Occurrence (how often scrap occurs) and Detection (what is the rating of tool that checks component´s quality) to calculate RPN value.

3.2 Step 2. DM Structure Development for KM Framework

The purpose of our DM is to store a product's PR data that are used in manufacturing for a specified order, analyze them and assess sigma performance level for product type as well as for common production system. The functionality of the DM is provided by the PR card combined with the Faults Classification standard DOE-NE-STD-1004-92 and FMEA methodology that is served by a common database.

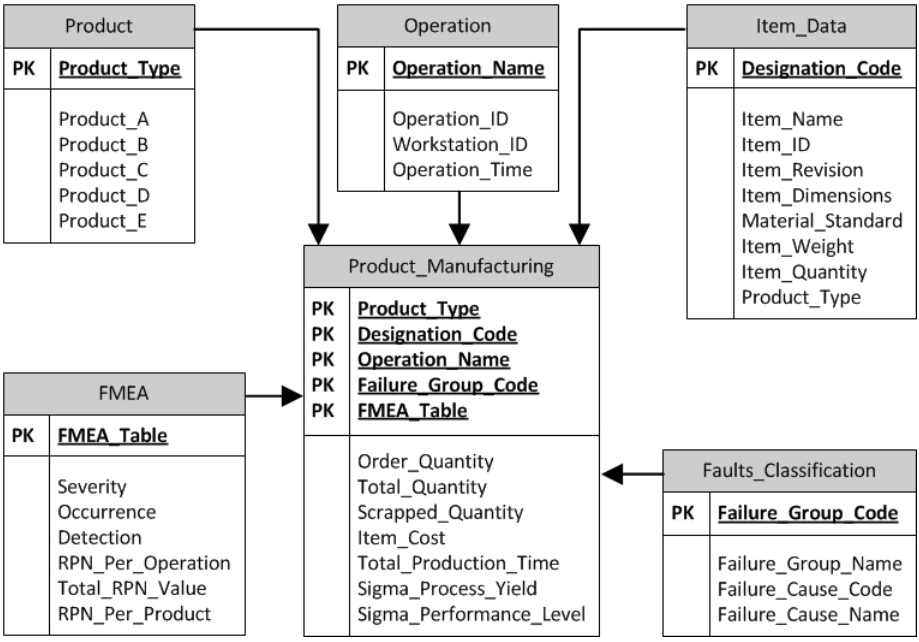


Fig. 3. Fact table of the Data Mart structure

Figure 3 illustrates the fact table of the DM structure, based on the Kimball bottom-up data warehouse design methodology [25]. This dimensional model contains data about concepts that are part of the manufacturing process. The dimensional model provides an easily communicable medium between people who understand the manufacturing process and IT workers who develop the software. It also provides the actual database structure of the DM. The developed DM enables management to access the data required for business reports from single location.

Data Mart Structure Description. In our case study we will calculate the RPN value for each operation and then for every product type. We also assess each unique operation that exists in a company's production floor by calculating the average RPN value for different product groups and then on the basis of this calculation, the average RPN for operation. The same assessment will be done for each Failure Group and Failure Cause mentioned in the Faults Classification. In other words, it will be a bi-directional analysis of a unique production operation and the failures that occurred. Table 2 (Appendix 1) shows an example of our DM where RP card, Faults Classification and a FMEA methodology/table were combined. Here we can see the Item Name and its Designation Code (used in the first step of the case study), Work Centre, Operation Sequence, Operation Name and Operation Time. In front of each operation according to the Faults Classification, a Failure Group and its sub-group are defined. Further we can see the part of the standard columns of the FMEA table where Severity, Occurrence and Detection, being multiplied and giving us the RPN value per operation. Actually one operation can have some failures. As a result, the total sum of the RPN value for the entire product production process may increase. Thanks to the fact table of DM structure we can start a bi-directional analysis of a product and production system by:

- **Operation.** We shall group equal operations that happen in a specified product or we can also sort/group all similar operations of a common production system to see the variability of the RPN value. After that we shall calculate an average RPN value for a product and for a common production system. The result will show us an average variability and occurrence of different failures by a specific operation that further allows making the right decision concerning a current status.
- **Failure Group and/or Failure Cause.** It is similar to the operation case. We shall group similar Failure Causes (sub-groups) according to their main group category – Failure Group. It will show us how many or how often similar failures occurs in some specific operation. From here we can see the variability of a failure, calculate an average RPN value for the Failure Group and if needed for the Failure Cause. It is also possible to group all the failures of all produced products to see the variability and calculate the average RPN value for a product or a common production system. The result will show us the frequency of a specific failure per operation and the quantity of failures in common production system that enables making further decision about a current production system.

Example. The following example, shown in Table 2 (Appendix 1) represents the core of our case study. We have the production process or the PR card with 12 operations

where 20 failures occurred. Theoretically the RPN value equal to (20x1000) 20000. Of course in a real production it guarantees the total product and production process crash. In our case we will use this maximum RPN value to define the scope of the entire production process performance level, for a product, as well as for a common production system. This maximum value will equal to 100%. Further we shall calculate the actual percentage of the process capability for every produced product and also the average capability percentage of the common production system. As the result of it, according to the calculated percentage, we will define the sigma performance level for each product and for the common production system. To calculate the RPN percentage for both cases, we shall use the following Equation 1:

$$P_{PR} = \frac{\sum RPN_{PC}}{\sum RPN_{Total}} \times 100\% \quad (1)$$

where:

P_{RP} – per cent of a failure by operation or a failure group/cause,

$\sum RPN_{PC}$ – RPN value for a particular failure group/cause or operation,

$\sum RPN_{Total}$ – total RPN value by product or by a common production system.

As the result, according to the calculations for the given case (product) we have got the current/total RPN value of 2000 that makes 10% of the total theoretical RPN value of 20000. In order to calculate a sigma performance process yield we use the following calculation: 100% - 10% = 90%. Hence, we have got 90% of the process yield or positive performance level from the current product.

From the statistical point of view, the term Six Sigma is defined as having less than 3.4 defects per million opportunities or a success rate of 99.9997% where sigma is a term used to represent the variation about the process average [26].

On the basis of received results (10%) and according to the Table 1 that represents the sigma performance scale that can be translated into measures of process capability Cp and Cpk [27], we can define the sigma performance level for the current example. Here we have got 2.78 δ or around 3 δ. Based on developed calculation it is possible to monitor on which sigma level a company operates at the current moment.

Table 1. Sigma performance scale [27]

Sigma Performance Level	Defects per Million Opportunities	Process Yield	Process Capability (Cp)	Process Capability (Cpk)	Estimated Cost of Poor Quality (% Revenue)
1.0 δ	670000	33%	Not capable	Not capable	>40%
2.0 δ	308537	69.2%	Not capable	Not capable	30-40%
3.0 δ	66807	93.32%	1.0	0.5	20-30%
4.0 δ	6210	99.38%	1.33	0.83	15-20%
5.0 δ	233	99.9767%	1.67	1.17	10-15%
6.0 δ	3.4	99.99966%	2.0	1.5	<10%

If a company is operating at three sigma levels for quality control, this is interpreted as achieving a success rate of 93.32% or 66807 defects per million opportunities. Therefore, the Six Sigma method is a very rigorous quality control concept where many organizations still performs at three sigma levels [28].

The result of current example shows the "dashboard", on what sigma performance level company operates. Thanks to this result, the company management should decide which Six Sigma project should be started in order to improve the way of costs decreasing, product quality, On-Time Delivery (OTD) and a customer's satisfaction.

4 Conclusion

Customers are placing increased demands on companies for highly qualified products, reliable production processes and order On-Time Delivery (OTD). Many companies do not have a detailed review of their activities, the rate of their losses caused by unreliable production processes and their current production level.

In this paper we have presented a new Knowledge Management (KM) framework where were considered the item/product data flow between Product Data Management (PDM) and Enterprise Resource Planning (ERP) systems using a special middleware that integrated and synchronized the data between those systems and Data Mart (DM). The integration provided to unite different data that generated the needed information for a product manufacturing. The framework allowed engineers to refuse from every day routine work and pay more attention to specific production issues. In DM we succeeded to combine the Production Route (PR) card, Faults Classification standard DOE-NE-STD-1004-92 and Failure Mode and Effect Analysis (FMEA) methodology into a common database. This combination ensured us to perform a bi-directional assessment and analysis of production operation and the failures that occur during a production process on a product, and on a common production system level. The collected data into ERP system from production floor was transferred into our DM where it was processed according to the specified algorithm. It allowed us to get Severity, Occurrence and Detection rating in online mode. Thanks to these ratings we have got RPN that was converted into the percentage value that showed us production process yield which enabled to estimate sigma performance level for the product and common production system. The application of such approach can be very useful for a company when it is to be audited by a customer and for managers. This case study can be like a "dashboard" in a cockpit that shows on which performance level a company is operates. The higher sigma level the higher a customer's trust.

Acknowledgements. This research was supported by Estonian Ministry of Education and Research for targeted financing scheme SF0140113Bs08 and Grant ETF9460 and European Social Fund's Doctoral Studies and Internationalisation Program DoRa, which is carried out by Foundation Archimedes.

References

1. Koch, R.: Living the 80/20 Way: Work Less, Worry Less, Succeed More, Enjoy More illustrated edition. Nicholas Brealey Publishing (2004)
2. Matzler, K., Hinterhuber, H.H.: How to make product development projects more successful by integrating Kano's model of customer satisfaction into quality function deployment. Technovation 18(1), 25–38 (1998)
3. Gunasekaran, A.: Agile manufacturing: enablers and an implementation framework. Int. J. Prod. Res. 36(5), 1223–1247 (1998)
4. Lõun, K., Riives, J., Otto, T.: Evaluation of the operation expedience of technological resources in a manufacturing network. Estonian Journal of Engineering, 51–65 (2011)
5. Teece, D.J.: Capturing Value from Knowledge Assets. California Management Review 40(3), 55–79 (1998)
6. Becerra-Fernandez, I., Gonzalez, A., Sabherwal, R.: Knowledge Management. Pearson Education, Inc., Upper Saddle River (2004)
7. Davenport, T.H., De Long, D.W., Beers, M.C.: Successful Knowledge Management Projects. Sloan Management Review, 43–57 (1998)
8. O'Brien, J.A., Marakas, G.M.: Management Information Systems. The McGraw-Hill Companies, Inc. (2008)
9. Blake, I., Gerard, P.L.: The information system as a competitive weapon. Communications of the ACM - Special Section on Management of Information Systems 27(12), 1193–1201 (1984)
10. Saaksvuori, A., Immonen, A.: Product Lifecycle Management, 3rd edn. Springer, Heidelberg (2008)
11. Shevtshenko, E., Karaulova, T., Kramarenko, S., Wang, Y.: IDSS used as a framework for collaborative projects in conglomerate enterprises. Journal of Achievements in Materials and Manufacturing Engineering, 89–92 (2007)
12. PDM-ERP Middleware (December 2012), http://plmware.tesis.de/index.php?page=1043
13. Berson, A., Smith, S.J.: Data Warehousing, Data Mining & OLAP. McGraw Hill (1997)
14. Arnrich, B., Walter, J., Albert, A., Ennker, J., Ritter, H.: Data Mart based research in heart surgery: challenges and benefits, September 7-11. MedInfo, San Francisco (2004)
15. Wang, K.: Applying data mart to manufacturing: the nature and implications. Journal of Intelligent Manufacturing 18, 487–495 (2007)
16. Production Route Card (October 2012), http://www.enotes.com/american-scholar/q-and-a/what-meant-by-job-card-route-card-used-production-99397
17. Sharon, K.J.: Combing QFD and FMEA to optimize performance. ASQC Quality Congress 52, 564–575 (1998)
18. Mcdermott, R.E., Mikulak, R.J., Beauregard, M.R.: The basics of FMEA. Resources Engineering, Inc., USA (1996)
19. Stamatis, D.H.: Failure mode and effect analysis: FMEA from theory to execution, 2nd edn. ASQ Quality Press (2003)
20. DOE-NE-STD-1004-92, Root Cause Analysis Guidance Document, US (1992), http://www.everyspec.com/DOE/DOE+PUBS/DOE_NE_STD_1004_92_262 (November 2012)
21. Anbari, F.T.: Six Sigma Method and Its Applications in Project Management. In: Proceedings of the Project Management Institute Annual Seminars and Symposium, San Antonio, Texas. Project Management Institute, Newtown Square (2002)

22. Rancour, T., McCracken, M.: Applying six sigma methods for breakthrough safety performance, pp. 31–34. American Society of Safety Engineers (2000)
23. Antony, J.: Some pros and cons of Six Sigma: an academic perspective. The TQM Magazine 16(4), 303–306 (2004)
24. Sahno, J., Opik, R., Kostina, M., Paavel, M., Shevtshenko, E., Wang, Y.: Knowledge Management Framework for Production Route Selection in Manufacturing Enterprises. In: 8th International DAAAM Baltic Conference Industrial Engineering, pp. 567–572 (2012)
25. Kimball, R., Ross, M.: The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2nd edn. John Wiley & Sons, New York (2002)
26. Antony, J., Banuelas, R.: Key ingredients for the effective implementation of six sigma program. Measuring Business Excellence 6, 20–27 (2002)
27. Gregory, H.W.: Six Sigma for Business Leaders, 1st edn. Business Systems Solutions, Inc., USA (2004)
28. McClusky, R.: The Rise, fall, and revival of six sigma. Measuring Business Excellence 4, 6–17 (2000)

Appendix 1

Table 2. Combination of PR card, Fault Classification standard and FMEA table in DM

ID/Item Name	Designation Code	Work Centre	Operation Sequence	Operation Name	Operation Time, H	Failure Group	Failure Name	Failure Cause	Failure Cause Name	Severity	Occurrence	Detection	Risk Priority Number (RPN)
									Maximum RPN value per product				20000
													100%
Item 1	ABC	WI	1	OpA		7	Supplier problem	7C	Defective material	6	5	3	108
			2	OpB		4	Design problem	4A	Inadequate design	6	3	4	72
						5	Training deficiency	5A	No training provided	8	3	3	72
			3	OpC		1	Equipment problem	1E	Bad equipment work	8	5	3	120
						5	Training deficiency	5E	Inadequate content	6	5	4	120
						3	Personnel error	3A	Inadequate work environment	8	2	6	96
			4	OpD		1	Equipment problem	1C	Software failure	7	5	3	105
						2	Procedure problem	2B	Lack of procedure	8	3	3	72

Table 2. (*continued*)

ID/Item Name	Designation Code	Work Centre	Operation Sequence	Operation Name	Operation Time, H	Failure Group		Failure Cause		Severity	Occurrence	Detection	Risk Priority Number (RPN)
						Failure Group	Failure Name	Failure Cause	Failure Cause Name				
Item 2	DEF	W2	5	OpB		6	Management problem	6B	Work organization	8	3	4	96
						2	Procedure problem	2A	Defective or inadequate procedure	5	3	6	90
			6	OpC		3	Personnel error	3B	Inattention to detail	9	4	4	144
			7	OpD		4	Design problem	4C	Dimensions related problems	6	4	3	72
						5	Training deficiency	5A	No training provided	5	5	4	100
			8	OpE		1	Equipment problem	1C	Software failure	7	3	5	105
			9	OpF		2	Procedure problem	2C	Error in equipment or mat. selection	6	3	3	54
Assembly	XYZ	W3	10	OpG		7	Supplier problem	7A	Time delivery error	7	4	5	140
						4	Design problem	4B	Drawing, specification, or data error	6	3	4	72
			11	OpH		2	Procedure problem	2B	Lack of procedure	7	5	4	140
						6	Management problem	6D	Improper resource allocation	10	2	3	60
			12	OpI		5	Training deficiency	5A	No training provided	9	6	3	162
										Total RPN value			2000
										RPN % from entire product			10 %
										% of Sigma process yield			**90 %**
										Sigma performance level			**2.78 δ**

Android, GIS and Web Base Project, Emergency Management System (EMS) Which Overcomes Quick Emergency Response Challenges

Atif Saeed*, Muhamamd Shahid Bhatti, Muhammad Ajmal,
Adil Waseem, Arsalan Akbar, and Adnan Mahmood

COMSATS Institute of Information Technology Defence Road off Raiwind Road,
Lahore, Pakistan
{asaeed,msbhatti,rajmal}@ciitlahore.edu.pk
{mhd.adil,arslan.akbar134,adnan_mahmood90}@hotmail.com

Abstract. The golden time and measure of agility during the emergency is that hour when the incident has occurred. Recovery and response should be served promptly to minimize the risk of losing life and property. Therefore this paper examines(analyze) and determines(clarifies) the concurrent(real time) application of Android, GIS, and web base system that is focusing on those exceptional areas where the emergency system is not very much efficient or is not very much agile (effective). Most emergency situation occurs due to fire, road accidents or natural disasters, this paper analyzes GIS data of transport systems, incident location, hospitals, ambulance and observing teams, police, fire brigades etc, spatial data analysis can also done for past incident scrutiny (analysis) by using ArcGIS server manager at the backend designed by all the basic map interaction WCF services in ASP.NET which can be able to respond on all sort of needs required by the responder by using his/her smart android phone.

Keywords: Android, GIS, WCF, Restful Services.

1 Introduction

Emergency Management System (EMS) has been the major issue recently. Emergency can point towards all those events that threaten life styles of the people living in a state, especially the resources of that state prescribely for its sustainability and stability. All such situations might be carried out from fires, explosions, and traffic accident, terrorism, natural disasters all comprises of hazardous events.

In order to view certain emergency management systems there are many tidings (information) & management systems established, such as Advanced Emergency Response System [1] and DCV-Decision Centered Visualization [1]. Most of these

* Corresponding author.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 269–278.
DOI: 10.1007/978-3-642-36981-0_26

systems are mainly endorsed (support) for the purpose of communication, resources management whereas these systems were installed on the user's computer. Nowadays mobile devices, phones and portable digital assistants (PDAs) have become the tools that everyone has its access on a daily routine. Almost everyone holds such a sort of buttoned (handheld) device due to which desk top application has been gone too far behind. People want to have everything to be done prompt (quick). So as a result this application will help such affected people to have agile response whereas there exact location will be positioned on the map by the GPS built in mobile devices; which is installed on the server site of this application built in GIS technologies (Arc Server) in the back end where GIS Server database administrators will be managing the server and will provide the type of services requested by the service responder.

1.1 Problem Statement

From the past analysis it has been observed that there were no accurate system that have been providing any kind of quick response services to the victimized person, if there were any system that were not very much effective and life saving. Such systems mainly works in a way that the victimized person will call the rescue service team's helpline form his/her mobile or landline and will tell them his/her position where he/she requires the rescue. Meanwhile there might be a chance that the helpline services are temporarily down or the operators are very much busy or there are possibilities of fake calls too. During these circumstances the person who is really in need of rescue might get unnoticed and loses his/her life.

Population is increasing day by day and in modern era mostly people use their own vehicles due to this, large amount of vehicles are moving on the road which causes traffic congestion on the mostly part of the city. In this heavy traffic sometimes the rescue services (mostly ambulances) were oftenly delayed to reach at the emergency spot. Another reason is the construction works going on, this also causes traffic congestion and ultimately can cause delay in response services. With using EMS the exact location of the victimized person can be shown whereas the rescue services can uses the shortest path to the victimized person using the map which is uploaded on the EMS server end.

While during any natural disaster it is impossible to provide rescue services at the spot because such disaster ultimately destroys everything. So the rescue services are only possible after the disaster has occurred. With the help of EMS the location of the victimized person can be located and such location can be saved for further analysis with which we can find out that where the disaster have caused much damage in case of life and property etc..

1.2 Role of GIS

GIS data is mostly present in the form of spatial sources, which includes scale/resolution, dimension (2D, 3D), schemas, geodata, geo location, geo references etc. [4], [6] shows related to this area, several phases are to be discussed in the process of emergency management:

- Planning,
- Mitigation,
- Preparedness,
- Response
- Recovery.

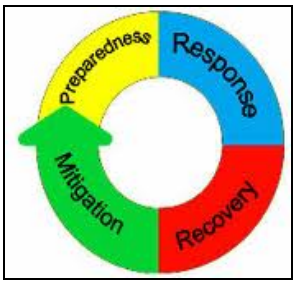


Fig. 1. Emergency management cycle

All the above phases of emergency management depend upon the data from different sources. The appropriate data has to be collected, organized and displayed rationally to determine the scope of emergency management system [5]. During an emergency it is very essential to have accurate data and take prompt actions. So that's why the surveillance teams should have appropriate and exact data about the location where the incident occurred. Therefore, by utilizing such GIS sources all such teams can update and use the information from the databases that shows location on the map.

1.3 Reasons to Use Mobile Phones

Keeping in context to the development of latest technologies of smartphone and tablets it is significant development for emergency management system (EMS), bringing authenticated functionality to it [7].

Such type of explosive application will provide powerful alerting functionality and enhance the rescue providers from providing the quick response to the emergency responder from fires, explosions, and traffic accident, terrorist attacks, natural disasters and all other hazardous events. The EMS is enabled by net-centric [7] technology and architecture connected to a centre server.

Now comes the time of activation of this application (EMS) on the field. Whenever there comes any disaster, the person that has encountered the problem, will run this application on his/her smartphone or tablet where following [figure 2] text fields will be shown which should be field by the responder then after entering all the text fields. The application will be submitted to the back end at the sever [1] where the responder's position (longitude & latitude) will be shown to the rescue administrator on the map via GPS/GSM built in the mobile.

Fig. 2. Interface that should be filled by responder

1.4 Use of Sensors in EMS

Sensors are nowadays very growing technologies that are being introduced in the smart phones. These technologies can help a lot for the rescue providers to provide services to the responder when he/she is in need of rescue.

Similarly, certain built in sensory are added in this application such as gyroscope, air pressure sensor, temperature sensor, fast fall algorithms [11]. These sensors will be helpful to the person if he/she is away from his/her mobile phone and is unable to execute this application, and then some text will be sent to his/her SOS as for the sake of rescue.

Suppose there is an incident occurred where the person is unavailable to get to his/her mobile. During certain condition mobile senses the situation and pops up a message for the user that he/she is in a situation of emergency. If the user does not respond to the message than after three minutes the same message pops up, such system will be carried out for three times relatively. If the user is not in state of responding to the popped message then at the fourth attempt the mobile will automatically send a message "NEED SOME HELP PLEASE RESCUE ME" to the person whom that person had declared his/her SOS. Such type of system is most possible during the condition when the person is having heart attack, trauma, rise in blood pressure, driving bike or car alone, mental distress (migraine), asthma attack etc.

- **Gyroscope and Accelerometer**

Accelerometers in mobile phones are used to detect the orientation of the phone. The gyroscope, or gyro for short, adds an additional dimension to the information supplied by the accelerometer by tracking rotation or twist.

An accelerometer measures linear acceleration of movement, while a gyro on the other hand measures the angular rotational velocity. Both sensors measure rate of change; they just measure the rate of change for different things. This type of sensor

will balance the equilibrium of the mobile such sensor will be used during the condition of earthquake, vehicular accidents etc. [12].

- **Air Pressure Sensor (Barometer)**

Most of the pressure sensors used are designed for a high resolution (as low as 0, 02 hPa) and high accuracy. The overall relative altitude accuracy is about factor 10 better than GPS. The accuracy of absolute altitude is depending on atmospheric pressure changes. For this case GPS data can help. Using these precise sensors a very accurate altimeter function can be implemented. Such sensor is mostly reliable when there is a condition of tsunami, storm etc... [13]

- **Temperature Sensor**

This type of sensor is a build in application known as AndroSensor [14] [15] which is available in android market. Whereas this application can be readily used with EMS. Such type of sensor is mostly effective when there is a fire, volcano eruption, or similar situation occurred.

1.5 Reasons to Use Web Services

In the emergency management system (EMS) traffic scenario, accident site, hospital locations and situation, rescue helicopters, fire brigades, ambulances and other local emergency information should be identified on actual basis(real-time) so do the geographical base map[Fig. 3].

From the [Fig.1] it is clear that all the relevant information is present in the server, therefore for such application we have to establish a local database for the relevant data. In which the GIS map will be loaded and all the rescue service providers will be in contact to the server administrator.

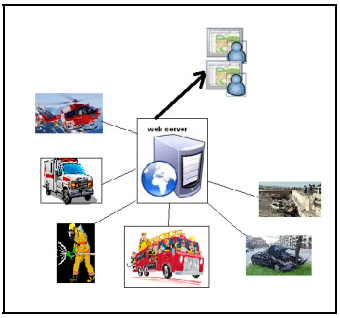


Fig. 3. EMS web server sketch

A tangible definition of web services is that the information needed to implement for its extreme practical usage. So that the main aim of the web services is to be applied to the traditional Web programming model. A web service is the last defined area of communication which will mollify (satisfy) the users of those services [1].

In this paper the web services is a software system tailored to scaffold (support) practical mobile-to-server collaboration over the network. Web services communicate with WCF (Windows Communication Foundation), RESTFULL SEVICES (get method, set method) with the mobile. In addition it will use XML from the mobile device and display the location of the user via GPS/GSM technology on the map that is present on the server.

1.6 The Function of GIS in Web Service

Web services provide a procedure that is being used by other applications. The assistance of using Web services is that the user is not familiar of how these services are constructed. The only thing they need to know is how to avail and use them, because the services interfaces are rigorously followed by industry principles. Web Services, are although different from Web applications, Services are not required to use by the user; but software applications is just to call. Though, the Web service does not contain any user interface, whereas Web applications provide the user interface. Usually, for the users, software containing Web Service are completely hidden, but ArcGIS Server web services is different from them, they will provide all necessary GIS functionalities to the system, this system will grant the necessary user experience to elaborate.

GIS Server provides the stage for sharing the usage of GIS information across the user's (client) such as map within your population. It can create maps, geographical data, spatial analysis data, geodata, geoinformations, mobile data, WMC (Web Mapping Services), KML (Keyhole Markup Language [1].

1.6.1 Restful Services in GIS Web

Representational State Transfer (REST) it is an architectural style of large-scale networked software that takes advantage of the technologies and protocols of the World Wide Web. REST illustrate how concentrated data objects, or resources, can be defined and addressed, stressing the easy exchange of information and scalability. It is an architectural style for building distributed hypermedia driven applications by defining resources that implement uniform interfaces using standard HTTP verbs (GET, POST, PUT, and DELETE), and that can be located/identified by a Uniform Resource Identifier (URI).

Geoserver has been providing a very efficient REST based interface to allow access to geoserver's admin functionality. Geoserver's REST defines the following key concepts:

- **Workspace** – A group of data stores and feature types. This is equivalent to a namespace.
- **Datastore** – A source of vector based spatial data. It may be a datastore, shape file or even a WCF server.
- **Feature types** – A vector spatial resource that originates from a data store.
- **Coverage** store – A source of raster based spatial data.

- **Coverages** – A raster based dataset with a coverage store as the data store.
- **Styles** – describe how a resource should be symbolized or rendered.
- **Layers** – it is a published resource, equivalent to a feature type or coverage.
- **Layer groups** – grouping of layers accessible under a single resource name.

2 Design and architecture of EMS

2.1 Platform Design

The design of the manifesto server we designed will be consisting of following parts [Fig. 4]: (1) Web server (GIS), (2) data server and (3) Android smart phone (PDA).The PDA will ignite the request to the web server hosts that containing GIS renewals, such as maps, globes and location of locators where they will be expose responder's request on the client side of the applications; the Web server(GIS) will be hosting the Web services that will use the sources running on the Web server(GIS); the data server containing the GIS source will be published as services on the Web server. Therefore the responder can be mapped documented, located on the globe, geodatabases, and toolboxes.

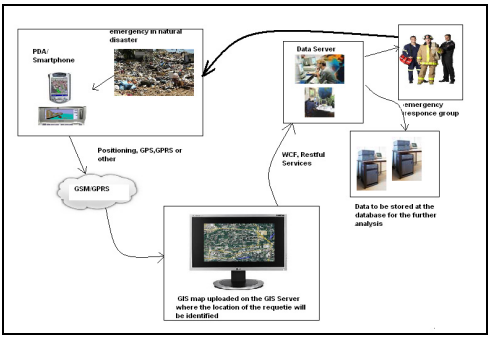


Fig. 4. Architecture of EMS

2.2 System Development

The system implementation will be very much easy in such way that the region where any kind of emergency response will be needed this application can be readily used.

The first step for this application is that the responder whoever wants to avail this services, the responder must have the PDA/android smart phone with him because with this the responder will be identified his/her exact location using GPS/GSM where the rescue services should be provided. As in Pakistan there is no proper technology through which the rescuer's can find the exact location of the people who are in need of rescue. With the help of this application these people can be easily detected via GPS/GSM.

As [Fig. 1] the responder has entered details on the front end of application on the smart phone, the responder will be asked what kind of help do you need? With all the required information [Fig. 1] obtained the location will be detected on the map build on the Arc Map, to the GIS Database Server manager on the back end. Where the DBS manager will analyze the responders location, where the manager will be having in contact with the emergency rescue provider agencies, which he will send to the affected area where the responder will get the initial treatment or will be rescued.

2.3 Algorithm

```
Incident occur
{
    User opens application;
    //  where  victimized  person  inputs  entities  from
    smart phone
    If (input == correct entities)
    Get location via GPS ();
    Store location (var lat, long);
    Send location (post method);
}
Get location via GPS ()
{
    //connect to satellite
    Get the current location via GPS;
}
Store location ()
{
    Var lat;
    Var long;
    String (lat, long)
    Store the location get via satellite;
}
Send location ()
{
    Method (post);
    // where services used are restful post services
    Lat, long sent to the server;
}
Store location in DB ()
{
    Get  lat,  long  from  the  smart  phone  using  restful
    services;
    Store in DB;
    Show position on the map;
}
```

```
// identifying the location of the user, rescue
teams   are sent towards the victimized area using
shortest path
```

3 Flow Chart

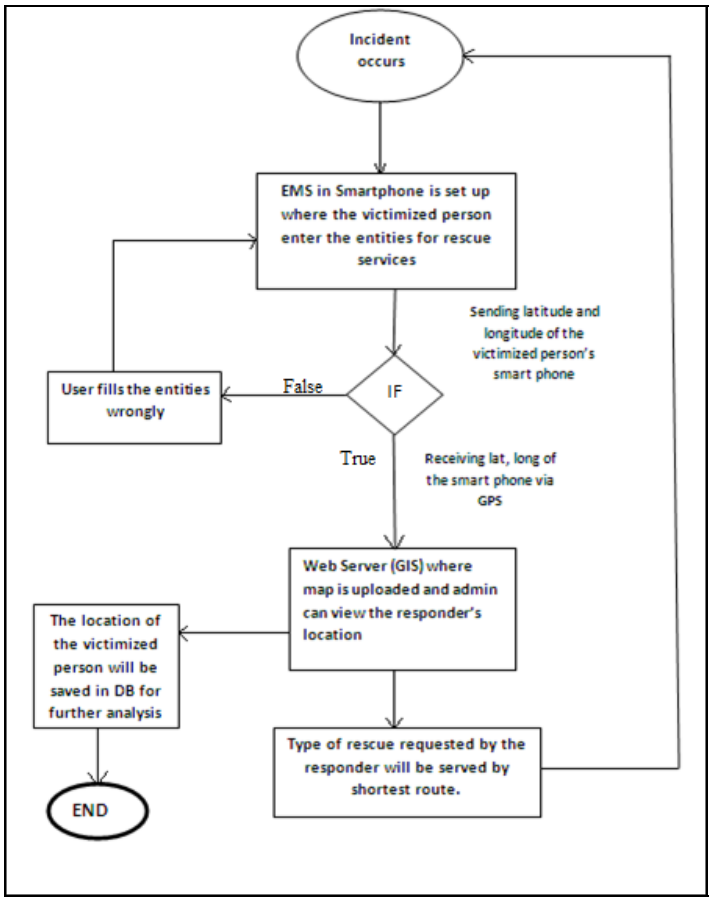


Fig. 5. Flow Chart of EMS

4 Conclusion

By the establishment of emergency management system using Android, Geographic Information System (GIS) and Web platform, clearly shows that the merging of multiple technologies can be very helpful in saving the life and properties of those people which are victimized. Such system can boost up greatly in daily life rescue operations and amongst agencies that are dealing with emergency response

challenges. Ultimately EMS will give the Emergency response teams to make their services better and much effective and by the involvement of smart phones victimized person will be easily approached by the response teams.

References

1. Tiana, H., Liu, Z., Zhang, J., Luo, C., Liu, B.: Design and development of emergency response geographic information platform based on web service
2. Zerger, A., Smith, D.I.: Impediments to using GIS for real-time disaster decision support (Centre for GIS and Modelling, Department of Geomatics, The University of Melbourne, Victoria 3010, Australia), Centre for Resource and Environmental Studies, Australian National University, Canberra 0200, Australia
3. Ganesh Kumar, B., Ramesh, D.: Emergency Response Management and Information System (ERMIS) – A GIS based software to resolve the emergency recovery challenges in Madurai city (PSNA College of Engineering & Technology, Dindigul, and Tamilnadu, India)
4. Zlatanova, S., Holweg, D.: 1GISt, 3D GEO-Information in emergency response: A framework (Delft University of Technology, Delft, The Netherlands, 2Fraunhofer IGD, Darmstadt, Germany)
5. GIS Technology for Disasters and Emergency Management. An ESRI White Paper (Russ Johnson Public Safety Industry Manager, ESRI)
6. ESRI White Paper. GIS for emergency management (July 1999), http://www.esri.com/library/whitepapers/pdfs/emermgmt.pdf
7. How smartphone and tablet technologies enhance emergency mass notification systems, http://www.gsnmagazine.com/node/24080
8. Zhang, X., Chen, H., Hong, H.: Service-Oriented Geographic Information Sharing and Service Architecture for Digital City. Advances in Intelligent Web Mastering 43, 381–386 (2007)
9. Coors, V., Flick, S.: Integrating Levels of Detail in a Web-based 3D-GIS. In: 6th ACM Symposium on Geographic Information Systems (ACM GIS 1998), Washington D.C., USA (1998)
10. Zlatanova, S., van den Heuvel, F.: 3D GIS for outdoor AR applications. In: Proceedings of the Third International Symposium on Mobile Multimedia Systems & Applications, Delft, The Netherlands, December 6, pp. 117–124 (2002) ISBN 90-9016467-7
11. Li, Q., Stankovic, J.A., Hanson, M., Barth, A., Lach, J.: Accurate, Fast Fall Detection Using Gyroscopes and Accelerometer-Derived Posture Information. University of Virginia
12. http://www.gsmarena.com/glossary.php3?term=sensors
13. http://bugs.xcsoar.org/ticket/1367
14. http://www.appbrain.com/app/androsensor/com.fivasim.androsensor

Pervasive Intelligent Decision Support System – Technology Acceptance in Intensive Care Units

Filipe Portela[1], Jorge Aguiar[1], Manuel Filipe Santos[1], Álvaro Silva[2], and Fernado Rua[2]

[1] Algoritmi Center, University of Minho, Guimarães, Portugal
cfp@dsi.uminho.pt, jorge.phena@gmail.com, mfs@dsi.uminho.pt
[2] Intensive Care Unit, Centro Hospitalar do Porto, Portugal
moreirasilva@clix.pt, fernandorua.sci@hgsa.min-saude.pt

Abstract. Intensive Care Units are considered a critical environment where the decision needs to be carefully taken. The real-time recognition of the condition of the patient is important to drive the decision process efficiently. In order to help the decision process, a Pervasive Intelligent Decision Support System (PIDSS) was developed. To provide a better comprehension of the acceptance of the PIDSS it is very important to assess how the users accept the system at level of usability and their importance in the Decision Making Process. This assessment was made using the four constructs proposed by the Technology Acceptance Methodology and a questionnaire-based approach guided by the Delphi Methodology. The results obtained so far show that although the users are satisfied with the offered information recognizing its importance, they demand for a faster system.

Keywords: TAM, INTCare, Technology Acceptance, Intensive Care, Decision Support System, Pervasive, Technology Assessment.

1 Introduction

Decision making in Intensive Medicine (IM) is a crucial process because deals with critical condition patients. Nothing can fail and if something wrong happens the patient can die. It is a specific area of Medicine and their knowledge it is practiced in the Intensive Care Units (ICU). ICU is recognized as a critical environment where the decision needs to be performed fast and with a high level of accuracy [1]. In the ICU the patient care is always the main concern and tasks like is patient documentation are relegated for a second plane [2]. The introduction of intelligent decision support system (IDSS) in the support of decision process is claimed by many of the nurses and physicians which work in ICU. This type of support can be addressed by a pervasive system which operates automatically and in real-time. This system can give to the ICU staff a better comprehension about the patient condition and at the same time predict future situations. INTCare is framed in this type of system. It is a system developed by this research team and which has as main goal the prediction of the patient organ failure and patient outcome in real-time for the next 24 hours. With the

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 279–292.
DOI: 10.1007/978-3-642-36981-0_27

development of the project, other types of sources were deepened and as result some new knowledge were obtained. Currently, INTCare it is considered by the ICU staff a very useful and complete platform, being composed by a set of pertinent information for the Decision Making Process (DMP). To this work, a list of requirements was defined based on the needs of ICU and the goal to make the system more suitable to the environment. They may be summarized as:

R1. To implement an online data acquisition component;
R2. To make available the laboratory results in an open format;
R3. To allow an open access to prescriptions, interventions and therapeutics;
R4. To dematerialise the nursing records;
R5. To integrate the main systems used in ICU in a single platform;
R6. Develop an automatic system to process and transforming the data.

Taking advantage of the modifications introduced (R1 to R6) it is possible to determine automatically and in real-time, using online learning:

a) ICU medical scores [3];
b) ICU critical events [4];
c) Probability of occur an organ failure probability and patient die [5].

In order to assess the results achieved, the technology, the INTCare functionalities and their importance to ICU, a questionnaire was developed. This questionnaire is based in the Technology Acceptance Methodology (TAM) [6] and it is concerned to the evaluation of four aspects: perceived usefulness (PU), perceived ease of use (PEOU), behavioural intention (BI) and use behaviour (UB). Despite of the questionnaire used be composed by a high number of questions, in this paper only are presented the TAM results associated to the decision making process, i.e., the results related to DMP.

This paper is divided in seven sections. The first and second sections introduce the work and make an overview of the concepts and work previous performed. The third section presents the improvements attained in the Intensive Care according to INTCare features. Then the fourth section introduces the PIDSS at the level of results achieved. The fifth and sixth sections are related to the TAM; the questionnaire performed and results achieved. Finally, some remarks and future work are considered.

2 Background

During 2009 when this project started the ICU information system was composed by a set of information silos. The ICU professionals need to access to more than five hospital applications to obtain important information and to make their decisions. Now with the introduction of INTCare they have the most important information available in the Electronic Nursing Record (ENR) and they can obtain new knowledge automatically moments after the patient documentation. This situation only was possible with the modification made by INTCare project and the introduction of new knowledge into the ICU DMP.

2.1 INTCare

INTCare [7, 8] is an IDSS to predict organ failure and patient outcome for the next 24 hours in real-time using online learning. This system is result of a research project. The research and modifications done allowed to obtain new types of data in an electronically format and in real-time [9, 10]. The new reality and the new environment created [11] allow for obtaining new knowledge fundamental to the decision process predicting patient condition, scoring the ICU measures and tracking critical events automatically and in real-time. The data is obtained through a streaming process and the knowledge attained is disseminated in situated devices.

2.2 Decision Making Process in Intensive Care Units

Making decisions in ICU is a complicated and danger process, because all tasks need to be performed quickly and accurately [1]. The ICU professionals deal with patients in serious life-risk. The use of technologies to support this type of process is welcome [2] however, normally this type of systems aren't helping, i.e., don't present the accurate information in the right time and in the right place. These types of situations complicate the decision process. At the same time, there is a problem associated to the patient documentation because it is always relegated to a second place. In order to overcome this situation some modifications were made in the ICU environment [11] and in the DMP.

2.3 Technology Acceptance Methodology and Delphi Methodology

The evaluation of a technology application is crucial to comprehend its suitability in a specific environment and also to measure the satisfaction level of its users. One of the most used models in this area is the Technology Acceptance Methodology. "TAM is adapted from the Theory of Reasoned Action (TRA) model which describes human behaviours in a specific situation" [12]. The main purpose of TAM is to present an approach to study the effects of external variables towards people's internal beliefs, attitudes, and intentions [6]. This model is also important because it gives an understanding about the acceptance of the decision support by the ICU staff and how can be useful in the course of their daily work. The goals of TAM can be achieved by using methodologies based on questionnaires. As a support tool it is important to use some aspects/characteristic of the Delphi method. The principles of the Delphi method involves the use of questionnaires being one of its key features [13] the preservation of anonymity of the participants. A questionnaire was prepared by a coordination team, composed by professionals of ICU and Information System, and sent to a set of participants (a group of experts from the ICU nurses team). The questionnaire was prepared taking into account the constructs of TAM [14, 15]. The correlations of the answers were evaluated through the Kendall's tau (τ) coefficient. Kendall's tau is a measure of rank correlation. The values range from -1 (inversion) to $+1$ (perfect agreement). A value of zero indicates the absence of association.

2.4 Related Work – Results Obtained in the First Approach

In order to make a first assessment of the technology, a quick and short questionnaire was produced [16]. The main goal was to have a first idea about the usefulness and ease of use of the system in superficial way. This questionnaire was the starting point of the second questionnaire (with tam) and it had a short scope. The questions were divided into two groups: Functional characteristics (data registration, information access and proactive performance) and Technical characteristics (efficient consulting, response time, system security, usability, and interoperability). Finally, a last question evaluate if the system suits the needs. The questionnaire was answered using a five-scale metric: Does not meet / in complete disagreement (<20% of cases) (1) until fully meet / fully agree (> 80%) (5) [16]. In terms of results only two questions were answered with less than 4 points: one question about the registration system and other question about the understanding of the system and their benefits.

Concluding, in the first phase of assessment the users revealed to be comfortable with the system. These results motivated: i) to continue the development of the project; and ii) perform a more extensive and deep questionnaire having the objective to understand the technology acceptance by the ICU users.

3 Research Propose – Improvements Introduced in ICU

The improvements made are according to the INTCare requirements (R1 to R6) defined in the introduction and can be summarized as:

3.1 Data Acquisition System (R1)

The first requirement was resolved with the implementation of a gateway. The gateway is connected to the vital signs monitors, reads the patient information and stores it on a database through the data acquisition agent. This is an autonomous process and it is always in a continuous collecting process (streaming).

In this phase two problems appear: missing patient identification (PID) and the acquisition of bad values. To overcome these problems two triggers were developed. One trigger to verify on the Electronic Health Record (EHR) system the PID of the patient admitted in bed where the values are provided and other trigger to validate the values. This second process uses the range of values pre-defined by ICU. Both the procedures are executed in the moment of the values are collected.

3.2 Laboratory (R2)

Regarding to the laboratory, an effort was made to have the lab results in an open format, i.e., accessible electronically and able to be handled without restrictions. The main objective was making the results available for ICU immediately after the patient exams are concluded. This change gives the possibility to have the results in a comparative format during the patient stay in ICU. Those exams have different types and are executed by different services and at different hours [16].

3.3 Open Access to Prescriptions (R3)

In this point the objective was deal with pharmacy and to study the possibility to construct an easy access to patient prescriptions. These prescriptions were totally controlled by pharmacy and whenever someone needed to consult the patient therapeutic plan had to open a too slow platform. Now, the interaction between the pharmaceutical system and ENR is made by an agent. Periodically, the ENR agent sends a request to the pharmacy drugs system and then, the requested data is sent to a database table [16].

3.4 Electronic Nursing Record (R4 and R5)

Electronic Nursing Record (ENR) is a platform that it was developed with the objective to receive all medical data and put it available electronically and in real-time to the physicians and nurses in an hourly mode. ENR can achieve two requirements because being it electronic can dematerialize the processes and due the interoperability mode can interoperate with all of others ICU data sources. Currently the ICU staff using the ENR has more vital information about the patient in order to help to make their decisions. ENR is a touch and web-based platform and it is composed by different screens. The data is grouped by the information provenance.

3.5 Automatic Data Processing and Transformation (R6)

After obtain all the essential data to the decision making process it was necessary introduce new features to the transformation process. The uses of intelligent agents allowed automate the whole process. Now the tasks associated to data preparation process are performed automatically and in real-time without human effort. These changes increase the speed in getting new knowledge being they useful and available in the right time, i.e., in the moment of the decision is taken.

4 Pervasive Intelligent Decision Support System

A pervasive intelligent Decision Support System (PIDSS) is recognized as a system that helps the decision making process and it is accessible anywhere and anytime. In the health care arena there are two concepts related to PIDSS as is the pervasive healthcare and the pervasive computing [17]. Due their pervasive features, INTCare can produce three different types of knowledge. This knowledge is available anywhere and anytime.

4.1 ICU Medical Scores

The objective of PIDSS component is to behave as an Intelligent Scoring System (ISS). The ISS [3] is incorporated into the Electronic Nursing Record (ENR). Nowadays, the ICU professionals can record and consult the scores in real-time.

This application allows for the automatic calculation in real-time of a set of scores: simplified acute physiology score (SAPS) II [18], SAPSIII [19], Sequential Organ Failure Assessment score (SOFA) [20], Glasgow Coma Score (GSC) [21], Therapeutic Intervention Scoring System (TISS-28) [22] and Modified Early Warning Score (MEWS) [23]. At the same time it is possible to analyze the patient evolution in terms of the scores through interactive graphs (in a hourly and daily base). As mentioned before [3]: this approach makes possible to provide a set of scores calculated / updated in real time. The ISS proposed processes automatically the scores and adapts the results according to the new values collected generating new knowledge. The main gains in using this approach can be summarized as:

- The data acquisition, scores calculation and results are made in real-time;
- All values are considered - no missing values;
- The data is displayed in a new way – real-time charts to compare trends;
- Less human intervention in the scores calculation – less errors;
- The scores are available anywhere and anytime;
- Help decision making process through a continuous scores monitoring.

4.2 ICU Critical Events

Critical Events (CE) are very important to the development of Data Mining (DM) models. In order to develop DM models in a real setting it was necessary to define procedures to automatically compute CE for five variables: Urine Output (Diuresis), Blood Pressure, Heart Rate, Respiratory and Temperature [4]. The procedure calculates according some rules the number and elapsed time of an event. Then, the value is characterized as critical or not. As result it is possible determine a number of critical events for the patient by hour and category [4]. In complement it is calculated the Accumulated Critical Events (ACE) [4]. The CE system is composed by a grid and a system similar to a traffic light. This system is used as a way to alert about the patient condition. The grid shows: the number of critical events by hour, the number of ACE, the time in critical event by hour and the total time in critical event. The implementation of this new approach allows to the physicians have better understanding of the patient's condition.

4.3 Ensemble Based Models

Data Mining (DM) is the centre of the PIDSS. The objective of DM system is to predict the patient organ failure (cardiovascular, hepatic, coagulation, respiratory and renal) and patient outcome for the next hour. To achieve this goal an ensemble DM was developed. To evaluate the ensemble three measures were considered: *Sensibility, Accuracy* and *Terror.* For each measure the average of 10 runs was taken. The selected models are used only if they satisfy the following conditions (quality measure): *Total Error* <= 40%; *sensitivity* >= 85% and *Accuracy* >= 60%. The use of ensemble helps to choose the best model in the cases where more than one model

presents good results. From the six targets, only three satisfy the quality measures defined: outcome, cardiovascular and coagulation. The low level of results verified in the other targets it is associated to the dynamic characteristics of the environment. Currently, it is possible induce DM models in real-time using online-learning and an ensemble approach in order to adapt the predictive models automatically. The doctors can use the predictions to save lives and avoid complicated situations to the patient.

5 Technology Acceptance Questionnaires

For this study it was elaborated a questionnaire based on the four constructs of TAM 3. This questionnaire was elaborated by taking into account some scientific articles that report similar processes of technological implementation and are framed in the hospital environment and the first results obtained. It means they were aggregated into several groups to represent all the aspects of TAM. The main purpose is to obtain a better understanding about the user's intentions on the use of this system in the long run as well as the functionality for them. The questionnaire is composed by 96 questions. However, in this paper only were considered the questions related to the decision process. In this questionnaire was applied the Likert Scale [24] to evaluate the results. This scale was chosen because the use of short scales (scales that goes between three and four) can better constrain results into close type of answers such as a simple yes or no; and secondly, by applying a higher scale this could fall into a dispersion of results that lead the answers to inaccurate results. As a consequence the chosen scale follows a range from one to five points similar to the used in the previous work / questionnaire. It allows for giving two values for each side and at the same time finding a neutrality point [24]. The considered levels are the following:

1) Not satisfies/in complete disagreement (< 20% of cases);
2) Satisfies a bit/in some level of disagreement (20-40%);
3) Satisfies/under some level of agreement (40-60%);
4) Satisfies a lot/strongly agreement (60-80%);
5) Satisfies completely/full agreement (> 80%).

The level of results collected from this questionnaire vary by the fact of the participant answer in a properly manner (with consciousness) or not. The answers always depend on the goodwill of each participant by answering in a balanced way to the questions of a certain group. This hypothesis does not verify when the participant evaluates a specific characteristic as a whole and gives the same answer to the group questions. To avoid this problem it was added to the questionnaire three screening questions to understand the level of the user's consciousness (e.g. 3+2). Table 1 crosses the questions with the constructs: Perceived Usefulness (PU); Perceived Ease of Use (PEOU); Behavioural Intention (BI); Use Behaviour (UB).

Table 1. Questions of Functional Characteristics

	Functional Characteristics	***PU***	***PEOU***	***BI***	***UB***
1.1	It allows the efficient registration of the information?		X		X
1.2	It allows obtaining efficient information for decision support?		X		X
1.3	It shows the prevision of Adverse Effects in an efficacy way?		X	X	
1.4	It shows usefulness when predicts the Scores?		X	X	
1.5	It improves the proactive performance of the professionals?	X	X	X	X
1.6	It allows tasks to be performed with greater precision?	X	X		
1.7	Can help to mitigate situations of an excessive workload?	X	X	X	X
1.8	Can allow a major control of several tasks?	X	X		X
1.9	Can help to have a better decision making based in best evidences?	X	X		X
1.10	Potentiates an improvement delivery of patient's health care?	X	X		X
1.11	It allows monitoring the patient's condition?	X	X		X
1.12	It promotes automating tasks?	X	X		X
1.13	It allows answering with the appropriate information to perform the task?	X	X	X	X
	Technical Characteristics	***PU***	***PEOU***	***BI***	***UB***
2.1	Can promote quality of the information?	X		X	X
2.2	Can access to information quickly?	X			X
2.3	It allows access to information in a secure way?	X	X		X
2.4	Can operate simultaneously with other hospital systems?	X	X	X	X
2.5	Can facilitate an operation by having a tactile interface beside to patient's beds?	X	X	X	X
2.6	It allows an efficient use based on the available technical support?	X	X	X	X
2.7	**Evaluate the potential of each registration presented on the Intervention's panel:**				
2.71	Can facilitate obtaining information regarding the realized interventions?		X		
2.72	Can facilitate obtaining information regarding the therapeutic attitudes?		X		
2.73	Graphic aspect?		X		
2.74	Registration of the work plan?		X		
2.75	Utility of TISS28?	X	X	X	X
2.76	Graphic aspect of TISS28?	X	X	X	
2.77	Global Evaluation of TISS28	X	X	X	X
2.8	**Evaluate the potential of each registration presented on the Score's panel:**				
2.8.1	The records made automatically, present similar values relatively to the manuscripts ones?	X	X		
2.8.2	Utility of SOFA CHART?	X	X	X	X
2.8.3	Utility of GLASGOW CHART?	X	X	X	X
2.8.4	Graphic aspect is intuitive?	X	X	X	X
2.8.5	The graphics can help to a better understanding of the real patient's condition?	X	X	X	X
2.8.6	By using the automation registration of Scores it facilitates the registration of SAPS II?	X		X	X
2.8.7	By using the automation registration of Scores it facilitates the registration of SAPS III?	X		X	X

Table 1. (*continued*)

2.8.8	By using the automation registration of Scores it facilitates the registration of GLASGOW?	X		X	X
2.8.9	Global Evaluation of Scores?	X	X	X	X
2.9	**Evaluate the potential of each registration presented on the Vital Sign's panel:**				
2.9.1	Utility of information?	X	X	X	
2.9.2	Utility of consulting information (hourly, daily, continuous)?	X	X	X	
2.9.3	Graphic aspect?	X	X	X	
2.9.4	MEWS – Utility of system?	X	X	X	
2.9.5	Adverse events – Utility of system?	X		X	X
2.9.6	The early warning system for Adverse Events is useful?	X	X		X
2.9.7	Graphic aspect?		X		
2.9.8	Global evaluation of the vital signs?	X	X	X	X
2.10	**It is advantageous to use this system in intensive care units?**	X		X	X

6 Results

After collecting answers from 14 questionnaires sent by email (35% total number of nurses in ICU) an analysis of the results was performed. First a processing was done to avoid invalid or inconsistent answers given by the participants. Then was noticed that only one participant out of the 14 nurses answered the questionnaire in an inconsistent way for the proposed questions. This situation leaded to only consider 13 surveys. Table 2 presents the technology experience of the respondents.

Table 2. Level of experience in Information Technology

Question	*Answer*	*Percentage*
What is your experience in technology – How much time do you spend at the computer	Less than 2 hours/day	0%
	Between 2 to 4 hours/day	57%
	More 4 hours/day	36%
Type of User?	Full Autonomy	62%
	Rarely need technical support (less than 3 times/month)	38%
	Need regular technical support	0%
Uses computer preferably for?	Application of production staff (email, text processing, spreadsheet)	62%
	Handling/Consulting administrative information	31%
	Handling/Consulting clinical information	77%
	Handling/Consult management information	8%

6.1 Respondent Analysis

For a better perception of the answers made by each respondent, one analysis (average and mode) was carried out by the person questioned and TAM construct (fig 1 to 4). In this analysis, the persons (X axis) are represented by numbers (1 to 13).

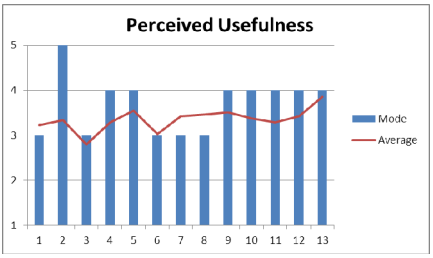


Fig. 1. Evaluation of PU

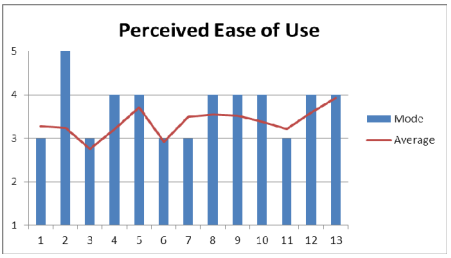


Fig. 2. Evaluation of PEOU

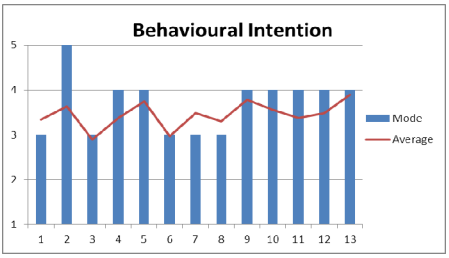


Fig. 3. Evaluation of BI

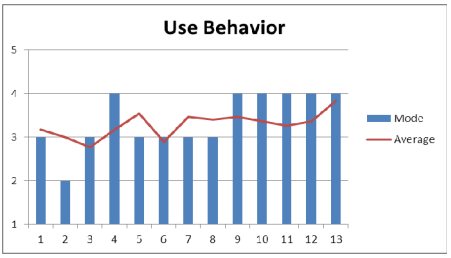


Fig. 4. Evaluation of UB

At a high level of analysis, in figures 1, 2 and 3, is possible to observe that the second person answered most of the questions with 5 values; however the average is between 3 and 4 points. This means that this person is quite pleased with some aspects of the system and not with others. In general the evaluations are above 3 points. At same time some correlations techniques were used to understand if the users are in accordance with the answers. In a global way they are in relative accordance in some of the questions, being the overall Kendall's tau: 0,158224.

6.2 Question Analysis

In this sub-section, instead of doing an analysis by respondent, an analysis was made for each one of the question and TAM 3 construct. In the Y axis they are the possible answers of the questionnaire (1-5) and in the X axis they are the questions numbers.

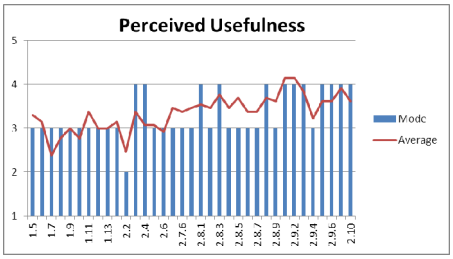


Fig. 5. Analysis for questions (PU)

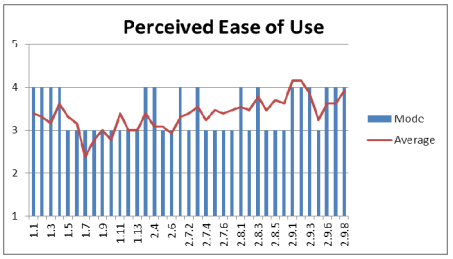


Fig. 6. Analysis for questions (PEOU)

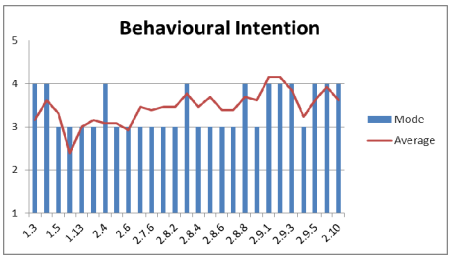


Fig. 7. Analysis for questions (BI)

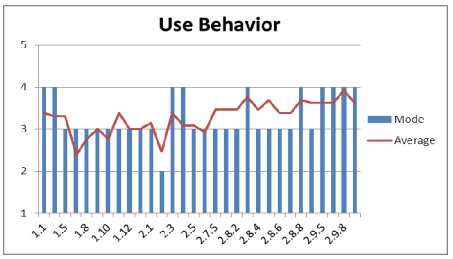


Fig. 8. Analysis for questions (UB)

In figure 5 and 8 it should be stressed that question 2.2 has a lowest score. On the other hand, the average of the answers for the questions related to these constructs is situated between 2 and 4 points. This result means that sometimes it is difficult to have access to the data. This happens due to hospital connectivity problems in the network. This problem represents the biggest barrier to the success of INTCare. In Figure 6 and 7, the vast majority of the answers of the questions to this constructs were stood by 3 points.

6.3 Global Analysis by Question

A global analysis was done in order to understand the best features, the average, the mode and the standard deviation (stdev) for each one of constructs. Table 3 offers a quick view of the answers obtained by the questionnaires. This table shows that the ICU staffs are satisfied with the system. All the constructs present positive results being the best, the Perceived Ease of Use and the worst, the Use behaviour.

Table 3. Summary of mode and average for each construct and analysis overall

	PU	*PEOU*	*BI*	*UB*	*Overall*
Mode	3	3	3	3	3
Average	3,35	3,67	3,44	3,28	3,36
Stdev	0,18	0,21	0,20	0,21	0,21
Min	1	1	1	1	1
Max	5	5	5	5	5

Table 4 offers a quick understanding of each construct. This table presents the three best and the three worst results. In the positive side they are some information (charts) to the decision process and the importance that the graphics have in order to have a better understand of the patient condition. This characteristic has as average (avg) 4.15. In the opposite side they are some functional characteristics. It is possible to observe that there are two negative characteristics (1.7 and 2.2). Both of them are related to the system speed. Despite all that, only one question has as mode 2 points. All of the other questions have a mode between the 3 and 4 points.

Table 4. The three best and worst results

	PU	*Mode*	*Avg*	***PEOU***	*Mode*	*Avg*	***BI***	*Mode*	*Avg*	***UB***	*Mode*	*Avg*
Best Results	2.9.1	4	**4,15**	2.9.1	4	**4,15**	2.9.1	4	**4,15**	2.9.8	4	3,92
	2.9.2	4	**4,15**	2.9.2	4	**4,15**	2.9.2	4	**4,15**	2.8.3	4	3,77
	2.9.8	4	3,92	2.9.8	4	3,92	2.9.8	4	3,92	2.8.5	3	3,69
Worst Results	1.7	3	2,38	1.7	3	2,38	1.7	3	2,38	1.7	3	2,38
	2.2	2	2,46	1.8	3	2,77	2.6	3	2,92	1.8	3	2,77
	1.8	3	2,77	1.10	3	2,77	1,13	3	3	1.10	3	2,77

7 Conclusions and Future Work

First, it should be noted that the initial objective proposed regarding to the junction of the Technology Acceptance Model (TAM 3) with the Delphi's method in order to evaluate the acceptance by users, their perceptions and the impact on the INTCare's system usage behaviour, it is totally innovative and can be considered a success. This is the first approach to assess the impact of such type of solutions in ICU environment.

In order to understand the acceptance it is possible to conclude that the ICU staff it is very comfortable with the system INTCare. They pointed the data access as the biggest problem and the utility of the information generated for the decision process as the biggest gain. The user acceptance was very positive (average upper than 3 points) for the four constructs assessed: Perceived Usefulness, Perceived ease of use, Behavioral Intention and Usage Behaviour.

Concluding, the ICU professionals are receptive to the INTCare system and to the new knowledge provided because it can help them in the decision making process.

The results obtained by the questionnaires allow for some concluding remarks:

- ✓ These results encourage further development and optimization of the solutions designed, as well as a deeper assessment of all the resources available;
- ✓ It is required an improvement of physical resources (e.g. memory, hardware) of the ICU.

Acknowledgements. This work is supported by FEDER through Operational Program for Competitiveness Factors – COMPETE and by national funds though FCT – Fundação para a Ciência e Tecnologia in the scope of the project: FCOMP-01-0124-FEDER-022674.

The authors would like to thank FCT (Foundation of Science and Technology, Portugal) for the financial support through the contract PTDC/EIA/72819/ 2006. The work of Filipe Portela was supported by the grant SFRH/BD/70156/2010 from FCT.

References

[1] Baggs, J.G., Norton, S.A., Schmitt, M.H., Dombeck, M.T., Setters, C.R., Quinn, J.R.: Intensive care unit cultures and end-of-life decision making. Journal of Critical Care 22, 159–168 (2007)

[2] Mador, R.L., Shaw, N.T.: The impact of a Critical Care Information System (CCIS) on time spent charting and in direct patient care by staff in the ICU: a review of the literature. International Journal of Medical Informatics 78, 435–445 (2009)

[3] Portela, F., Santos, M.F., Machado, J., Silva, Á., Rua, F., Abelha, A.: Intelligent Data Acquisition and Scoring System for Intensive Medicine. In: Böhm, C., Khuri, S., Lhotská, L., Renda, M.E. (eds.) ITBAM 2012. LNCS, vol. 7451, pp. 1–15. Springer, Heidelberg (2012)

[4] Filipe Portela, P.G., Santos, M.F., Silva, Á., Rua, F.: Intelligent and Real Time Data Acquisition and Evaluation to Determine Critical Events in Intensive Medicine. In: HCist 2012 - International Conference on Health and Social Care Information Systems and Technologies, Portugal (2012)

[5] Filipe Portela, F.P., Santos, M.F.: Data Mining Predictive Models For Pervasive Intelligent Decision Support In Intensive Care Medicine. Presented at the KMIS 2012 - nternational Conference on Knowledge Management and Information Sharing, Barcelona (2012)

[6] Chooprayoon, V., Fung, C.C.: TECTAM: An Approach to Study Technology Acceptance Model (TAM) in Gaining Knowledge on the Adoption and Use of E-Commerce/E-Business Technology among Small and Medium Enterprises in Thailand (2010)

[7] Santos, M.F., Portela, F., Vilas-Boas, M., Machado, J., Abelha, A., Neves, J.: INTCARE - Multi-agent approach for real-time Intelligent Decision Support in Intensive Medicine. In: 3rd International Conference on Agents and Artificial Intelligence (ICAART), Rome, Italy (2011)

[8] Gago, P., Santos, M.F., Silva, Á., Cortez, P., Neves, J., Gomes, L.: INTCare: a knowledge discovery based intelligent decision support system for intensive care medicine. Journal of Decision Systems (2006)

[9] Portela, F., Gago, P., Santos, M.F., Silva, A., Rua, F., Machado, J., et al.: Knowledge Discovery for Pervasive and Real-Time Intelligent Decision Support in Intensive Care Medicine. Presented at the *KMIS* 2011- International Conference on Knowledge Management and Information Sharing, Paris, France (2011)

[10] Portela, F., Santos, M., Vilas-Boas, M., Rua, F., Silva, Á., Neves, J.: Real-time Intelligent decision support in intensive medicine. Presented at the *KMIS* 2010- International Conference on Knowledge Management and Information Sharing, Valência, Espanha (2010)

[11] Portela, F., Santos, M.F., Silva, Á., Machado, J., Abelha, A.: Enabling a Pervasive approach for Intelligent Decision Support in Critical Health Care. Presented at the *HCist 2011* – International Workshop on Health and Social Care Information Systems and Technologies, Algarve, Portugal (2011)

[12] Fishbein, M., Ajzen, I.: Belief, attitude, intention and behaviour: An introduction to theory and research. Addison-Wesley (1975)

[13] Zackiewicz, M., Salles Filho, S.: Technological Foresight–Um instrumento para política científica e tecnológica. Parcerias Estratégicas 6 (2010)
[14] Venkatesh, V., Bala, H.: Technology acceptance model 3 and a research agenda on interventions. Decision Sciences 39, 273–315 (2008)
[15] Venkatesh, V., Davis, F.D.: A theoretical extension of the technology acceptance model: Four longitudinal field studies. Management Science, 186–204 (2000)
[16] Filipe Portela, M.F.S., Machado, J., Abelha, A., Neves, J., Silva, Á., Rua, F.: Intelligent Decision Support in Intensive Care –Towards Technology Acceptance. Presented at the 26th European Simulation and Modelling Conference - ESM 2012, Essen, Germany (2012)
[17] Varshney, U.: Pervasive Healthcare Computing: EMR/EHR, Wireless and Health Monitoring. Springer-Verlag New York Inc. (2009)
[18] Le Gall, J.R., Lemeshow, S., Saulnier, F.: A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA 270, 2957–2963 (1993)
[19] Metnitz, P.G.H., Moreno, R.P., Almeida, E., Jordan, B., Bauer, P., Campos, R.A., et al.: SAPS 3—From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. Intensive Care Medicine 31, 1336–1344 (2005)
[20] Vincent, J., Mendonça, A., Cantraine, F., Moreno, R., Takala, J., Suter, P., et al.: Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: Results of a multicenter, prospective study. Critical Care Medicine 26, 1793–1800 (1998)
[21] Jones, C.: Glasgow coma scale. AJN, The American Journal of Nursing 79, 1551 (1979)
[22] Reis Miranda, D., de Rijk, A., Schaufeli, W.: Simplified Therapeutic Intervention Scoring System: the TISS-28 items–results from a multicenter study. Critical Care Medicine 24, 64 (1996)
[23] Gardner-Thorpe, J., Love, N., Wrightson, J., Walsh, S., Keeling, N.: The value of Modified Early Warning Score (MEWS) in surgical in-patients: a prospective observational study. Annals of the Royal College of Surgeons of England 88, 571 (2006)
[24] Johns, R.: Likert Items and Scales. Survey Question Bank: Methods Fact Sheet 1 (2010)

GLORIA: The First Free Access e-Infrastructure of Robotic Telescopes for Citizen Science

Carlos Jesús Pérez-del-Pulgar[1], Raquel Cedazo[3], Juan Cabello[1], Esteban González[2], Víctor F. Muñoz[1], Fernando Serena[2], María C. López[1], Fernando Ibáñez[2], Francisco M. Sánchez[2], Alberto Castro[4], and Ronan Cunniffe[4]

[1] Automation and System Egineering Department, Universidad de Málaga, Málaga, Spain
carlosperez@uma.es
[2] Computer Architecture and Information Technology Department, Universidad Politécnica de Madrid, Madrid, Spain
rcedazo@fi.upm.es
[3] Electronics, Automation and Industrial Computing Department, Universidad Politécnica de Madrid, Madrid, Spain
egonzalez@ciclope.info
[4] Stellar Physics Department, Instituto de Astrofísica de Andalucía, Granada, Spain
ajct@iaa.es

Abstract. This paper describes the GLORIA system (GLObal Robotic telescope Intelligent Array), the first worldwide network of robotic telescopes, which covers four continents with seventeen telescopes, and can be extended. The network architecture has been designed taking into account the functionalities offered to the community, the number of telescopes, users that will be managed and network compatibility with all kinds of robotic telescopes. Different kinds of experiments have been designed to support the various requirements of astronomers. In order to manage the user access to the network resources, an evaluation system, called karma, has been defined which will operate according to some established policy.

Keywords: robotic telescope, social network, web 2.0.

1 Introduction

Research in astronomy poses two main challenges; 1) the immensity of the sky and 2) the huge quantity of astronomical data being gathered. In fact, astronomers are nowadays facing great difficulty in finding the resources to analyze the increasing flood of data generated by modern astronomical instruments. To meet these challenges, an increasing number of astronomy projects have turned to the public, via the Internet, e.g. GalaxyZoo [1][2] and Zooniverse [3]. These projects use the collective intelligence [4] of hundreds of thousands of users to analyze huge amounts of data and achieve results that are otherwise impossible. On the other hand, in recent years, various robotic telescope networks have emerged including some that allow access to users such as students and amateur astronomers as well as professional

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 293–304.
DOI: 10.1007/978-3-642-36981-0_28

researchers. Examples include the eStar [5] project, continued by RoboNet-I [6] and Robonet-II [7] projects (pure research); Las Cumbres Observatory Global Telescope Network [8] (for both research and education, with two operational telescopes of 2m and 1m diameter); GTN ("Global Telescope Network") supported by NASA and Sonoma State University (gtn.sonoma.edu); SkyNet [9] which has around eleven telescopes in their network and is intended for both research and educational purposes; the Australia Telescope National Facility [10], which involves an array of six 22-m antennas used for radio astronomy research; the BOOTES network [11], which includes 4 telescopes worldwide for researching purposes; and finally, the TAROT network [12] which has two robotic ground-based telescopes. All these networks are based on their own closed architectures and it is very difficult to integrate a robotic telescope into these networks without using devices compatible with the software of the network. Moreover, although these networks can manage their observing time using an advanced scheduling system, they do not take into account the quantity of information generated and how to process it. On the other hand, for pure educational purposes, there are currently several telescope networks that are trying to develop software specifically for integrating telescopes into their network. An example is the SAROS software system used by APTA [13][14].

The worldwide network of robotic telescopes proposed in this paper is based on the idea: "the more eyes we would put on the sky the more and greater the scientific discoveries that will be achieved". Thus, in order to try to improve the way of doing astronomy research, this work aims to build the first open access worldwide network of robotic telescopes, helping users to get involved in astronomy research, to collaborate in making new discoveries and enabling telescope owners to integrate their instruments into the network using free and open source software.

The GLORIA project is based on experience gained from *Cíclope* [15] and *rts2* [16] and is being developed by a consortium of thirteen members from different countries around the world, which will make their seventeen robotic telescopes available to users.

2 Overview of the System

This section explains the design and network architecture of the GLORIA system, as are the services that will be provided. Later, the concept of a GLORIA experiment is introduced, as well as its classification and description.

2.1 Network Architecture

The GLORIA network consists of three main components: telescopes, users and the core logic that connects them together as a network.

Following a device-driver approach, we decided to split the software into two distinct areas, as shown in Fig. 1. The bottom area, called Robotic Telescope Network, corresponds to the subset of software that is functionally and topologically located in each separate telescope. This software (Robotic Telescope System or RTS) takes the role of a device according to the proposed approach, and offers the possibility of teleoperation of any telescope from anywhere. This telescope network will grow as owner users put their telescopes available in GLORIA.

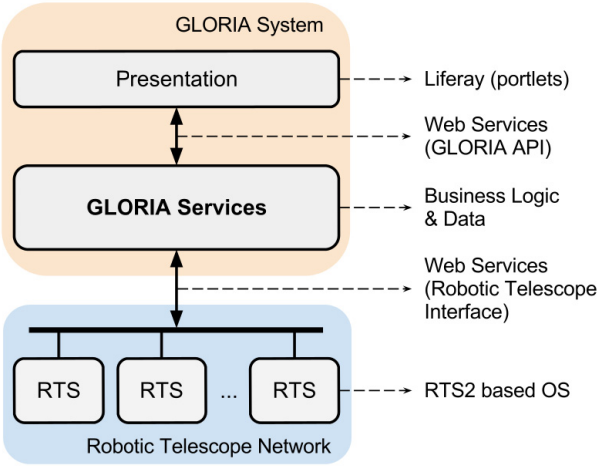


Fig. 1. GLORIA Network architecture

The upper area, called the GLORIA System, assumes the corresponding role of driver over the Robotic Telescope Network. Communication between these two parts is done via the web, using a high level protocol defined in the GLORIA standards (Robotic Telescope Interface).

The GLORIA System in turn consists of two elements with quite different purposes. The first corresponds to the Presentation layer of the system, i.e. the subset of software that is the interface between the users, and the second element are the services. These services bring together all the business logic of GLORIA and are identified as the core of the whole system. Anything a user may want to do, i.e. control a telescope, run an experiment or vote for an image, will be done through a web service operation included in some service inside the GLORIA Services layer.

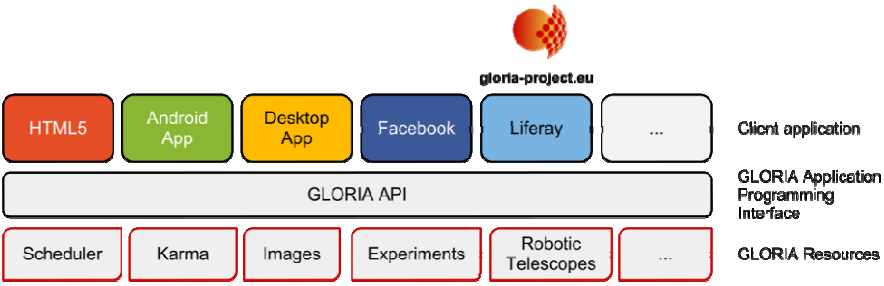


Fig. 2. GLORIA Application Programming Interface

The architecture of GLORIA Network has been designed according to service oriented methodology (SOA) [17]. The main reason for applying this paradigm is to make GLORIA functionality independent from the various paths or interfaces used by the community. In this way we achieve independence from any particular technology

other than the web transport. Therefore, the only thing that requires any client application, which is part of the Presentation layer, is to use the GLORIA API to consume any of its services (see Fig. 2).

2.2 Telescopes Network Scheduler

In order to manage all observation requests by users, it is necessary to develop a scheduler that caters for all the robotic telescopes in the network [18], taking into account that each telescope has its own scheduler and observation policy. This core scheduler arranges observations taking into account the distinct telescope properties, the nature of the target to be observed and some constraints set by the astronomers. The scheduler solution is based on these parts:

- Observing plan: this is a package, in the form of an XML file, containing everything the scheduler (and eventually a telescope) needs to know to carry out the observation. An Observing Plan includes constraints (specific time window, specific telescope, specific filters, height above horizon, moon separation, time after dusk/before dawn, etc.) and a sequence of instructions in a basic scripting language of five verbs: Target (what (object or coordinates) the telescope should point at now), Camera (configuration settings such as windowing, binning, etc.), Expose (defining timing, repetitions or duration and filters), Label (marks the beginning of a logical block in the program, but has no physical effect and Repeat (used after a block of instructions to indicate repetitions of the entire block).
- A local scheduler: each telescope has a scheduler of its own (possibly with far more than GLORIA on its mind – i.e. a telescope may be offering GLORIA only a part of its time, and this may be by interleaving GLORIA tasks with its private affairs and not necessarily with a dedicated time slot).
- A central scheduler: this provides the global robotic network schedule.

In this model, considering a job as an execution request of an Observing Plan, the central-scheduler advertises this to the local-schedulers of telescopes that can do the job, and each local-scheduler will reply with either a proposal (for how and when it can do it), or a refusal. The central scheduler accepts the best proposal (determined by policy), and waits for confirmation that the job is complete, or for an apology that the job was not or will not be carried out according to the original proposal. The central-scheduler manages all jobs in the GLORIA system, receiving new jobs from the authoring tool, advertising them (and re-advertising as necessary), accepting the best proposal (taking into account a global policy) and retiring tasks when they are done.

The scheduler policy is based on the Karma (described in section 4.4), but an additional priority layer is required to prioritize some Observing. This layer enables the system to execute alerts i.e. GRB or supernovae observations.

3 Telescopes Management

GLORIA is a heterogeneous network, but all the telescopes must present a common interface in order to integrate itself into the system. This interface is the outer layer of the modular and scalable architecture that all robotic telescope systems (RTS) must fulfill to be included in the network. This section describes this architecture and all steps that a RT has to take to complete the registration process.

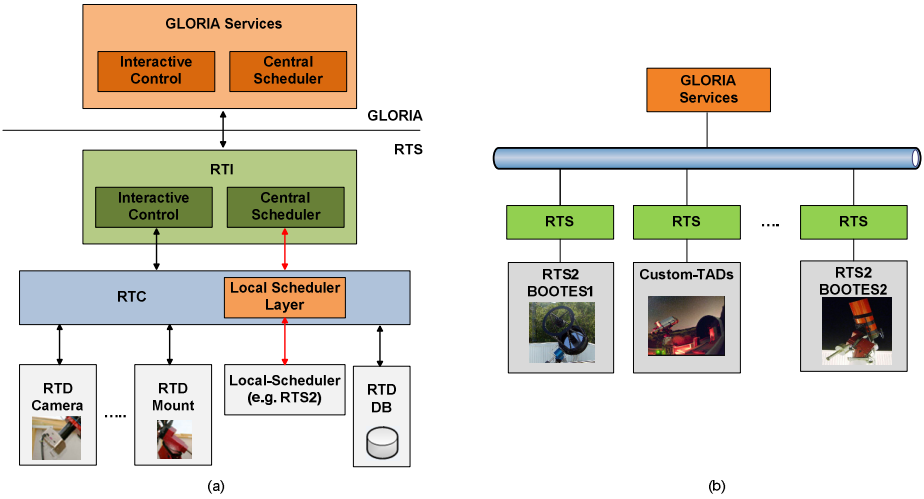


Fig. 3. a) RTS architecture. b) Robotics telescopes in GLORIA Network.

3.1 RTS Architecture

Fig. 3 shows all the components of the robotic telescope system and the communications between them, and they are described in detail below.

A separate Robotic Telescope Device (RTD) controls and monitors each physical device in the robotic telescope (RT), and belongs to one of a number of different classes, taking into account the different devices that a telescope commonly has, i.e. dome, mount, etc. Each RTD type supports the standard properties of the corresponding device, and a distinct interface is defined for each RTD type. These interfaces are used in turn by the Robotic Telescope Controller (RTC). In addition to the RTDs for hardware control, there is also a specific RTD for managing the Robotic Telescope Database (RTDB) which contains the images the system acquires.

The RTC manages the devices (RTD/RTDB) and the communication between the Robotic Telescope Interface (RTI) and the RTDs. It is also responsible of monitoring device state and for security. Using the RTD state information, the RTC is responsible for undertaking all necessary actions to keep the telescope in a stable and undamaged state. Finally, the Robotic Telescope Interface (RTI) is the communication interface between the Gloria Services (GS) and RTS.

All interfaces within the RTS (RTI and RTDs) are technology independent (Java, C++, etc.). Like others telescopes systems [6], these interfaces have been defined using WSDL (Web Services Description Language). Thus, a telescope owner can develop the entire RTI using any technology that supports web services interfaces.

3.2 RT Registration

In order to make easier the integration of new telescopes into the Gloria network, all the RTS software will be available for owners as well as detailed instructions. The only modifications owners have to make are those related to the HW dependent part of the RTDs. Once the telescope owner has the system installed in his telescope and he is already logged in the Gloria network, the RT has to undergo various stages or phases before being fully integrated in GLORIA network:

- Test: At this step, the RT owner has already fulfilled the GLORIA Network Integration Request, where all the telescope data, the devices and implemented operations are detailed. When the RT enters this phase it becomes Pre-Alpha version, and the system runs some tests to verify critical aspects of the communication with the RT and its basic characteristics. Once the automatic tests have been passed, the telescope is ready to be validated from Gloria System.
- Administrative validation: The owner and a subset of users selected by him/her have exclusive access to verify the operation of the RT. At this step, the RT becomes Alpha version. They compare the available operations offered by the owner to the ones actually implemented and its results. The objective of this phase is to refine the functionality of the RT to the owner's desired quality level, before opening it to the community.
- Collective validation: The use of the telescope is open to the community, but there will be no direct impact on its karma. Every user can operate the RT and report incidences about the problems they have encountered on it. The RT reaches the Beta state. The goal of this phase is to obtain, through all the information gathered from individual experiences, a global validation report that helps Gloria Administrators to decide the final integration of the RT.
- Complete: The RT is completely ready to be used by the community with its karma enabled. Here, the telescope reaches the Release state. An initial value of karma is assigned based on the results given by the last global validation report generated during the Beta phase.

4 User Management and Website

The final users of the network are both the owners of the RTs and anyone who wants to execute astronomical experiments. They have a common point of access to the system based on the Web.

4.1 User Authentication

Authentication and user management is one of the most important things for creating a social network. In the case of GLORIA, which has involved several universities and

research centers, and in the future schools, associations and other, this becomes more important because it's necessary to give access to the network for a large number of users [13].

To make this work easier, a study about the different technologies for authentication (user management and authentication in different web applications) has been made, concluding that the use of a federated authentication service helps for this purpose. Federated authentication allows any organization to deploy its own authentication system, and to connect it to different web applications. The advantages of this type of authentication features are many: easy to implement, scalable approach, integrating a large volume of users, use the same authentication system in different web applications that use different technologies, etc. [19].

There are many examples in which federated authentication have been successful. At European level, the Stork project [20] allows to users to authenticate using their national eID, and nowadays it's using to establish relationships with the public administration. The initiative eduGain [21] is a service that identifies users around the world from the research and education community.

In the case of this project, an authentication provider (SP) that will allow organizations to authenticate users who have made agreements with the GLORIA project will be deploy, for example: you are a secondary education center that has a telescope and wants to share it in GLORIA network, and you want to give access to the network to all the students. In this case the center would install an identification provider (IdP), and they connect it to the GLORIA Service Provider. In the case of users who do not belong to any organization, GLORIA will have its own identification provider that contains the database of registered users.

Fig. 4 shows an example of authentication system architecture that will be implemented in the GLORIA project. User will log into any web application project, it will direct user to GLORIA Federation Service, and the user will be able toauthenticate using his center service provider or the GLORIA user database. Once identified by the system, he will be able to browse between different web applications with his role, and without having to re-identify.

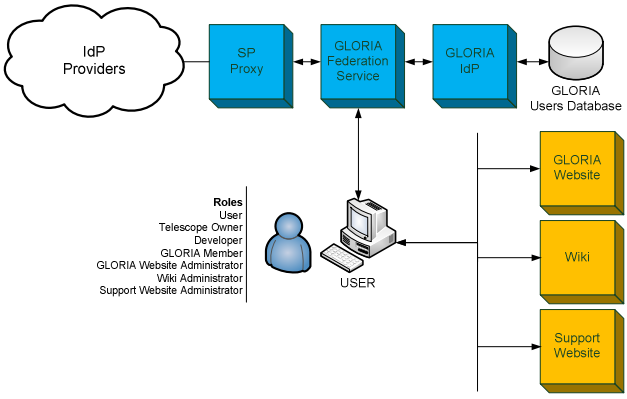
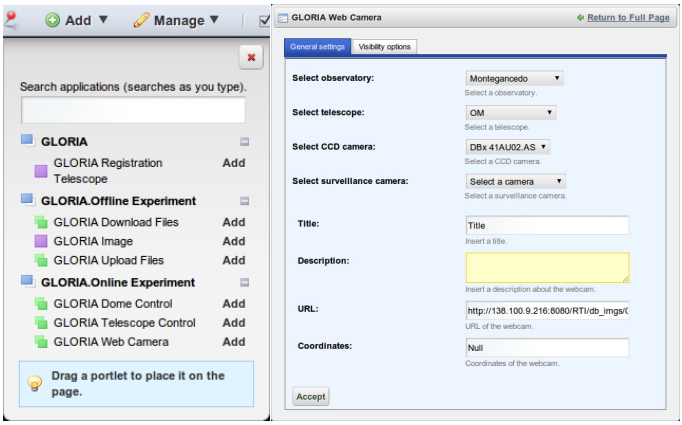


Fig. 4. Authentication architecture

4.2 Authoring Tool

The GLORIA website provides an Authoring Tool, which integrates the different web components, called portlets, for the on-line and off-line experiments. This Authoring Tool provides a user-friendly web interface for the creation of new experiments.



(a) (b)

Fig. 5. Authoring tool

The Authoring Tool is mainly a palette of web components organized by category, as can be seen at Fig. 5.a. There are currently three different categories, general purpose on-line and off-line. Experiments are created by simply adding web components can be to any page using drag-and-drop.Furthermore, each web component has a configuration that can be modified by the experiment designer, as can be seen in Fig. 5.b.

New web components can be written and added to the palette in any category, but the web administrator is responsible for deploying them through the Liferay [22] control panel.

An example of how to create and configure an experiment can be watched in a video on the GLORIA YouTube channel [23].

4.3 Experiments

GLORIA is going to take science to citizens, by enabling research of a specific astronomical issue via an experiment, in which users will be guided through the different tasks the research requires. These experiments are of two kinds: on-line (which require a telescope) and off-line, which work on existing data produced by the GLORIA network or derived from other databases, such as the European Virtual Observatory.

On-line experiments are based on Observing Plans, defined in Section 2.2, and are classified according to when (Fixed, Scheduled or Alert) and how (Batch or Interactive).

Fixed experiments execute at a time chosen by the user, Scheduled one at a time chosen by the system, and Alerts execute when triggered by an external signal. As to how the experiment is carried out: Interactive mode allows teleoperation of the telescope, with semi-direct (mediated by the software on the telescope) access to individual devices such as cameras or filter wheels, while batch mode refers to the execution of a pre-programmed Observing plan. This yields six combinations, for all of which there are valid uses.

The Off-line experiments allow users to do it with the information stored in the GLORIA database. There are several types of off-line experiments: the "Find your star" experiment allows a user to get the object description and an image for high quality printing; the target of the "Solar activity monitoring" experiment is the study of the solar activity (sunspots), applying a methodology for calculating astrophysical parameters; the user can verify if a star brightness decreasing is due to the occultation of the star by a Solar System object (e.g. asteroid), requesting a "Occultation of stars" experiment; the user can search for optical transients in a proper sky part requesting an "Optical transients" experiment; the "Prediction of occultation of stars" experiment can predict possible occultation of stars thanks to a Solar System object ephemerides; The "Search for variable stars evolution with time" experiment allows user to decide if a star shows significant changes in variability period and/or shape for time scales of the order of years; the user can classify an object attending its light curve; in this case, the requested off-line experiment is called "Classification of variable objects".

4.4 Karma

Karma is an index that measures the reputation of a user in GLORIA. Users participate in GLORIA by voting for images, including commentaries, etc.., i.e, increasing science contribution. It is used in numerous social networks like Digg, eBay, etc. This index is automatically calculated at intervals by the GLORIA System as a function of each user's participation and the application of specific policies that can be established dynamically.

The Karma component in GLORIA is based on the Social Activity [24] module of Liferay. This module permits valuation rules to be defined as a function of user actions over the website.

Every piece of presented information (called a resource) can be evaluated by users through the execution of actions on over them. For instance, a user can vote an image, write a comment, etc. Valuation rules use two metrics: participation points are awarded to the user doing the action, and contribution points are awarded to the creator of the resource, with a user's total score being the sum of these. The Karma module transforms these points into a karma value in the range 0-10.

Once users are evaluated, GLORIA can assign observation time, but establishing a set of rules to assign this time according to karma is a complex and resource demanding task. Another point of view is needed and the real world can give us an answer.

What do we use in the real world to trade resources? Money, so we propose to introduce a virtual currency, called GLORIUS, to pay for observation time. The quantity of GLORIUS a user earns will depend on their participation in the network, i.e, its karma. They can then spend GLORIUS whenever they make a reservation,

thus preventing individual users from taking all the available observation time for themselves. This implies a definition of an economic model. This approach is being used for resource management and scheduling in other fields like Grid computing [25].

This system of reservations and time allocation will be implemented through the use of auctions. This way, users will be able to compete for observation time. Auctions are used in many platforms, such as ebay, to automatically set prices based on valuations made by a community of users. The price of a time slot of observation time on a specific RT will be fixed using this method. All of this process is shown in Fig. 6.

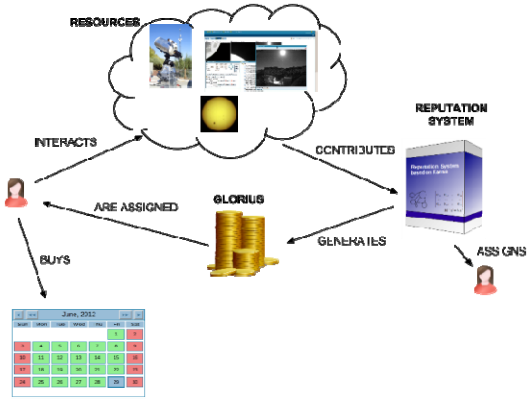


Fig. 6. Interactions between Karma and economic model

5 Results

During the three years of the project, the GLORIA consortium will integrate 17 telescopes, with 12 of them already commissioned and working in various scientific and public outreach fields, three more beginning operation with GLORIA, and two more that will be installed during the lifetime of the project. Once all RTS software has been developed and tested, a progressive deployment plan will be defined with the aim of integrate it into every RT. At the end, all RTs will be managed by the GLORIA network using the same interface. Moreover, a set of the experiments explained in section 2.2 will be performed to evaluate the overall system's operation.
In the same manner as RT deployment (and earlier), the GLORIA system side will be also progressively launched, with step by step testing and deployment of all the modules of the system.

Up to now, only one experiment has been created through the Authoring Tool described above. This is a solar experiment which is composed by the following web components: CCD camera, surveillance camera, mount and dome (see Fig. 7).

Currently, four telescopes have been integrated into the GLORIA network: OM [26], TAD [27], BOOTES-1 and BOOTES-2 [11]. The solar experiment is running right now through the OM and TAD telescopes, as these are the only solar telescopes of the network.

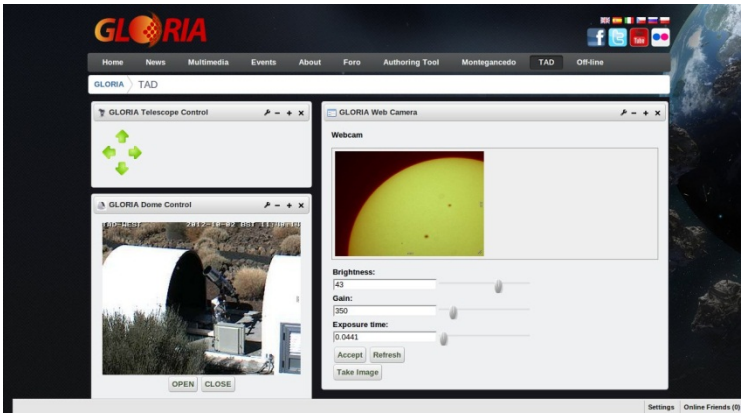


Fig. 7. TAD solar experiment interface

6 Conclusions and Future Work

This paper has described the first worldwide network of robotic telescope, which intends to give users free access to a network of 17 robotic telescopes spread across 4 continents and both hemispheres. Furthermore, GLORIA will enable more telescopes to join the system. The users of the network will be able to execute two kinds of astronomical experiments: on-line experiments (those that require a telescope) and off-line experiments (those that work on data produced by the GLORIA network or derived from other databases). Thus, the GLORIA websites offers different functionalities, including: a user evaluation system (karma), experiments and telescope management facilities, user authentication, and others tools which allow users to make different actions.

Acknowledgment. The authors acknowledge the partial financial support from the GLORIA project. GLObal Robotic telescopes Intelligent Array for e-Science (GLORIA) is a project funded by the European Union Seventh Framework Programme (FP7/2007-2012) under grant agreement number 283783.

References

1. Jordan Riddick, M., Bracey, G., Gay, P., et al.: Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. Astronomy Education Review 9(1), 010103 (2009)
2. Galaxy Zoo, `http://www.galaxyzoo.org`
3. Zooniverse, `http://www.zooniverse.org`

4. Smith, J.B.: Collective Intelligence in Computer-Based Collaboration. Lawrence Erlbaum Associates, Inc. (1994)
5. Allan, A., et al.: Operating a heterogeneous telescope network. In: Silva, D.R., Doxsey, R.E. (eds.) Proceedings of the SPIE, vol. 6270, p. 6270-0H (2006)
6. Mottram, C.J., Fraser, S.N.: Robonet-1.0. Astronomical Notes 329(3), 317–320 (2008)
7. Tsaprasn, Y., Street, R., et al.: RoboNet-II: Follow-up observations of microlensing events with a robotic network of telescopes. Astronomical Notes 330(1), 4–11 (2009)
8. Brown, T., Rosing, W.E., et al.: Temporal Variability, and the Las Cumbres Observatory Global Telescope. In: American Astronomical Society Meeting 210, #66.01; Bulletin of the American Astronomical Society 39, 173 (2007)
9. SkyNet, http://skynet.unc.edu
10. Bock, D.C.: The Australia Telescope National Facility. In: General Assembly and Scientific Symposium, 2011 XXXth URSI (2011)
11. Castro-Tirado, A.J., Jelínek, M., Mateo Sanguino, T.J., et al.: BOOTES: A stereoscopic robotic ground support facility. Astron. Nachr. 325, 679 (2004)
12. Bourez-Laas, M., Vachier, F.: CADOR and TAROT: A Virtual Observatory. In: Advanced Software and Control for Astronomy II. Proc. of SPIE, vol. 7019, p. 701918 (2008)
13. Duncan, A.R.: Federating access to small-aperture telescopes. In: International Conference on eScience, pp. 8–14 (2010)
14. APTA Telescopes network, http://www.telescope-net.com
15. Cedazo, R., López, D., Sanchez, F.M., Sebastian, J.M.: Ciclope: foss for developing and managing educational web laboratories. IEEE Transactions on Education 50(4), 352–359 (2007)
16. Kubánek, P., Jelínek, M., et al.: The RTS2 protocol. In: Proc. SPIE 7019, Advanced Software and Control for Astronomy II, p. 70192S (2008), doi:10.1117/12.788623
17. Earl, T.: SOA Design Patterns. Prentice Hall (2009)
18. Denny, R.B.: Dispatch Scheduling of Automated Telescopes. In: The Society for Astronomical Sciences 23rd Annual Symposium on Telescope Science, p. 35. Society for Astronomical Sciences, CA (2004)
19. Chaowang Shang, Z.Y.: SAML Based Unified Access Control Model for Inter-Platform Educational Resources. In: International Conference on Computer Science and Software Engineering, pp. 909–912 (2008)
20. European eID Interoperability Platform, http://www.eid-stork.eu
21. eduGain, http://www.geant.net
22. LifeRay Portal, http://www.liferay.com
23. GLORIA Youtube Channel, http://www.youtube.com/user/gloriaproject
24. Sezov Jr., R.: Liferay in action. Manning, 167–175 (2012)
25. Buyya, R., Abramson, D., Giddy, J., Stockinger, H.: Economic models for resource managament and scheduling in Grid computing. Concurrency Computation: Practice and Experience, 1507–1542 (2002)
26. OM Telescope website, http://om.fi.upm.es
27. TAD Telescope website, http://www.ot-tad.com

Evaluating Web Site Structure Based on Navigation Profiles and Site Topology

Alberto Simões[1], Anália Lourenço[2], and José João Almeida[3]

[1] Centro de Estudos Humanísticos, Universidade do Minho
ambs@ilch.uminho.pt
[2] Departamento de Eng. Biológica, Universidade do Minho
analia@deb.uminho.pt
[3] Departamento de Informática, Universidade do Minho
jj@di.uminho.pt

Abstract. This work aims at pointing out the benefits of a topology-oriented wide scope, but differentiated, profile analysis. The goal was to conciliate advanced common website usage profiling techniques with the analysis of the website's topology information, outputting valuable knowledge in an intuitive and comprehensible way. Server load balancing, crawler activity evaluation and Web site restructuring are the primary analysis concerns and, in this regard, experiments over six month data of a real-world Web site were considered successful.

1 Introduction

Usually, Web usage profiling is performed by specialised analysts that are capable of mining clickstreams and interpreting different kinds of outputs. Yet, even do it is not always possible to make Webmasters participate in Web analysis activities, they should be primary end-users and thus, results should be made more comprehensible.

Profile reporting is somewhat challenging due to the complexity and detail of the information. Site topology provides a valuable help by tagging the sequence of requests to site structure, i.e., mapping profiles into the site's oriented graph representation [5].

Graphs are a natural representation of topologies. Nevertheless, site topologies tend to embrace a high number of Web pages and hyperlinks, leading to high dimensional graphs that are not easily rendered or analysed. While providing similar navigation and analysis capabilities, layered graph visualisation is considered a better approach.

In this scenario, we propose an integrated approach to topology-based differentiated profiling. Data visualisation was the primary design concern and addressed both data comprehensibility and readability and graph navigability.

2 Topology-Oriented Profiling

The development of the proposed approach involved three main activities: topology automatic retrieval and processing, differentiated profiling analysis and graph

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 305–311.
DOI: 10.1007/978-3-642-36981-0_29

layered visualisation. Topology information is issued as the most adequate support to provide profiling insights. Differentiated profiling ensures focused processing and analysis of the two foremost groups of users, i.e., regular users and Web crawlers. Finally, the visualisation of the topology, i.e., the site graph, takes into consideration site's dimension and structuring, preserving data comprehensibility and readability.

2.1 Case Study

Natura Web site, property of the Language Specification and Processing Group[1] of the Computer Science Department of the University of Minho, complied with all these requirements and thus, was considered suitable for this experiment. Natura is a Natural Language Processing (NLP) research project focused on Portuguese language and its Web site supports the project research activities, the group's academic Web pages related to Natura and general NLP and the homepages of some of the project's members. Although this is a non-commercial site, the diversity of its contents is quite appealing in terms of differentiated usage profiling. Scientific publications, academic events, software and other NLP resources are mainly visited by students and researchers. Yet, the music repository embracing poems, lyrics, accords, music scores and karaoke files, attracts not only regular users and general searchers, but also focused music-related retrievers.

These experiments used six month server log data (Table 1).

Table 1. General information about server log data

Metric	Value
Unique user agents	30 324
Unique hosts	61 0127
Unique requests	44 673
Unique referrers	2 855
Total number of requests	7 198 999
Volume of traffic	1 795.09 GB

2.2 Topology Extraction and Processing

Topology extraction is supported by breadth-first crawler harvesting whereas the user may specify multiple seeds and harvesting stop points. These parameters specialise broad crawling in order to prevent crawling into irrelevant or problematic areas. For example, Concurrent Version System (CVS) access points and mailing list archives. The crawler outputs two classes of graph entities: nodes and edges. Nodes represent the URLs of site documents while edges stand for the hyperlinks relating those URLs.

[1] Natura Project Web site `http://natura.di.uminho.pt`

Topology processing aims the simplification of the crawled topology and includes three main processes: node clustering, graph simplification and node depth calculation.

Node Clustering. There are some specific website areas or kind of web applications that have hundred of similar pages, all based on a similar template. These pages share the same set of inbounds and outbounds, as shown in figure 1.

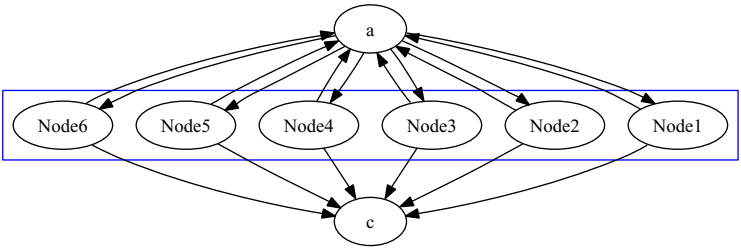


Fig. 1. An example of a cluster candidate

This means that all these nodes can be collapsed in a single node, representing them all, without losing any kind of information, but reducing the size of the graph, making it easier to analyse and to visualize. Therefore, the graph is inspected in order to find areas that can be simplified.

Graph Simplification. There are some pages that include a big set of inbounds and outbounds from and to the same web page. Also, some pages include a big set of hyperlinks to themselves, like the usual index presented in the top of webpages, that link to different sections in that same page. Figure 2 exemplifies how simplification is performed regarding multiple, or self-reference, hyperlinks.

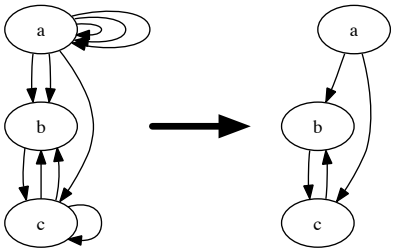


Fig. 2. Example of graph simplifications

Depth Computation. It is not plausible to assume that visits always start at the root or the first level nodes nor to expect that their pattern will follow a straightforward depth-first or breadth-first pattern. Depending on previous visits and Web site indexing, visits can be initiated at any level and may look into different (related or non-related) levels of contents. Therefore, the calculation of node depth (Figure 3) was considered an important profiling asset.

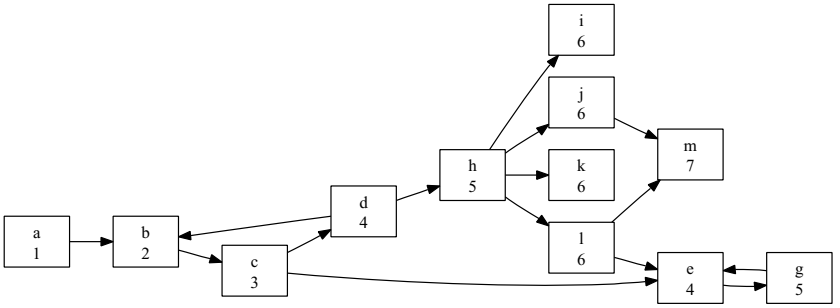


Fig. 3. Graph depth calculation based on SPPP Dijkstra's algorithm

2.3 Profile Analysis

Conventionally, Web analysis focuses on regular users' most common traverse patterns. Yet, regular users are not the only users whose attention Webmasters wish to capture. Nowadays, Web site visibility is greatly dependent on general and focused purpose search indexes. Also, there are a considerable number of other crawlers traversing the Web sites gathering all sorts of information. Even though they are not primary users, their presence indicates that someone is interested in that kind of contents and such interest may return some profit if Webmasters take into consideration those profiles as well.

Due to the ever growing similarities between crawler and regular user patterns, usage differentiation is a challenging task. Web crawlers are widespread and standard detection heuristics are unable to cope with the continuous evolving of the technology. Navigation pattern mining seems to be the most reasonable approach to the problem as it naturally encompasses changes in navigation patterns and does not imply the maintenance of any catalog.

The present work used the pattern mining approach introduced in [4,3], involving the semi-automatic labeling of a training set of Web sessions and tree model induction. Besides crawler and regular user sessions there were identified browser-related application sessions (Table 2).

2.4 Topology-Based Profile Evaluation

A web application was conceived in order to support topology-based profile visualisation. This tool is written in Perl and uses GraphViz software [2] to perform on-the-fly graph rendering while profiles are retrieved from a MySQL database.

Table 2. Statistics about differentiated Web sessions

Month	Total	Crawler	Regular	Application	Unknown
january	166 490	59 879	98 875	6 592	1 144
february	175 192	66 091	103 163	4 670	1 268
march	256 649	120 829	130 126	4 187	1 507
april	222 203	102 445	115 041	3 376	1 341
may	339 413	135 151	196 535	5 621	2 106
june	318 937	151 324	161 511	4 067	2 035

As it is very hard to visualise any site graph as a whole [1] and profile analysis would not be comprehensible, the tool uses a layered visualisation approach. Each layer includes the current node and all nodes that are directly connected to it (Figure 4).

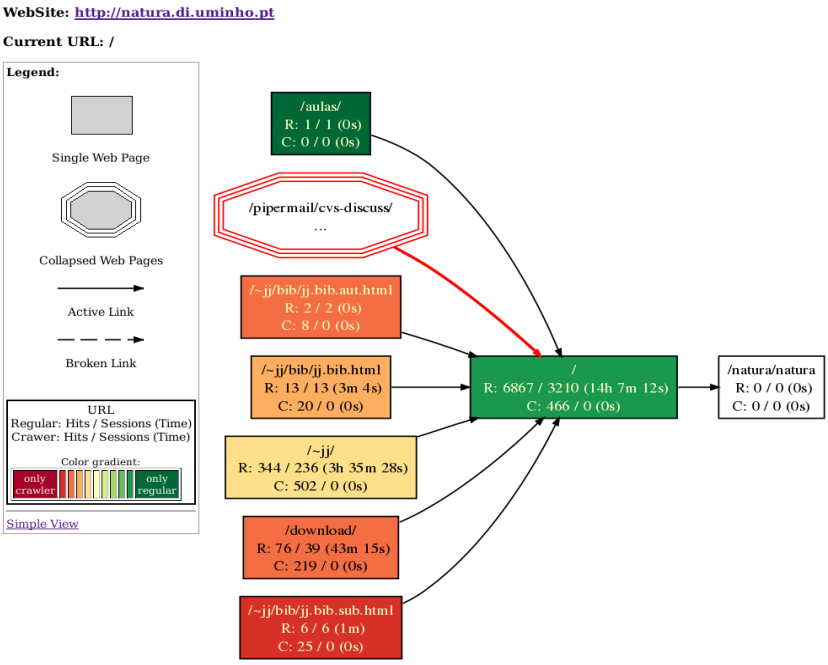


Fig. 4. Layered perspective of Natura's Web site root

Node representation includes the corresponding URL and profile statistics. There are two profile views: the general view and the expanded view. The general view presents the number of hits for crawler or regular users, while the expanded view also includes session counts and estimated time spent. Also, a colour gradient provides an intuitive view of node's crawler vs user load balance.

Node colour is based on a gradient from green to red, where greener nodes are mostly accessed by regular users and redder node are mostly accessed by Web crawlers. Node clusters are represented by an octagon shape with a sample non-clickable URL. On the other hand, edge representation is twofold: plain arrows for active links, and dashed arrows for broken links (Figure 5).

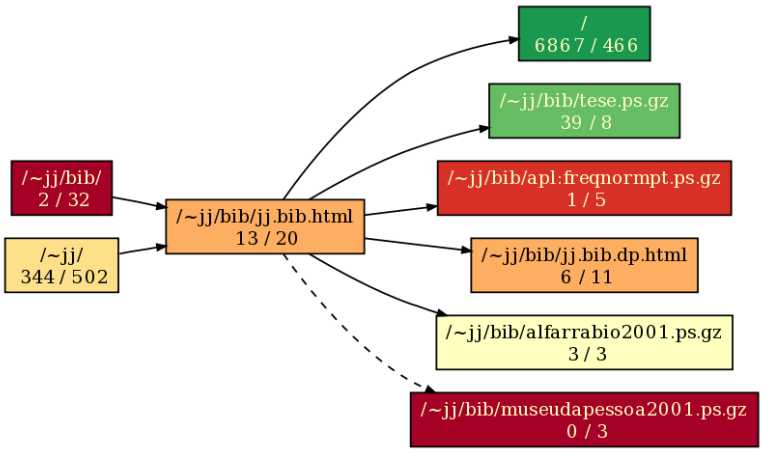


Fig. 5. Example of hyperlink representations

Looking at the generated topology Webmasters can easily identify hot and dead site areas and broken links. Besides, they may evaluate differentiated profiling metrics tracking down overlapping traverse areas, i.e., site areas that attract both users and crawlers. Large overlapping areas are interpreted as good indexing indicators, while distant usage areas indicate poor or inadequate indexing. Based upon the list of crawlers that were profiled, it is also possible to argue about indexing flaws, both in terms of crawling purposes and site structuring.

As the period of analysis is specified by the user, it is possible to assess server load over distinct periods of time as well as compare the impact of restructuring actions.

3 Final Remarks

Profile analysis provides relevant insights about user interests and current navigation patterns. Yet, regular users and crawlers are not alike and should be analysed separately aiming at their particular traverse purposes.

The proposed approach extended profiling with site topology. Layered graph viewing supported profile visualisation and graph features represented site's most common semantics. Intuitively, node and edge shapes represent different site resources and their associations while node colour balances differentiated usage.

Data visualisation was the primary design concern and addressed both data comprehensibility and readability and graph navigability. At first, if not aware of the real structure of the Web site, a naive user may consider topology navigation confusing, but the Webmaster will quickly find out his way.

In terms of future work, both profiling and data visualisation may be enriched. Node information can be more thorough. Clusters should identify the list of collapsed URLs along with the associated regular/crawler profiling metrics. Also, it would be relevant to provide information about the most common navigation patterns highlighting the preferred hyperlinks. Regarding the web application, the usage of GraphViz is an overhead to the server running the application, and therefore we should opt for an HTML canvas-based rendering tool.

Acknowledgments. This work was partly supported by project CROSS (PTDC/EIA-CCO/108995/2008), funded by the Portuguese Foundation for Science and Technology.

References

1. Cugini, J., Scholtz, J.: VISVIP: 3D visualization of paths through web sites. In: Tenth International Workshop on Database and Expert Systems Applications, Florence, Italy, pp. 259–263 (1999)
2. Gansner, E.R., North, S.C.: An open graph visualization system and its applications to software engineering. Journal of Software — Practice and Experience 30(11), 1203–1233 (2000)
3. Lourenço, A., Belo, O.: Applying clickstream data mining to real-time web crawler detection and containment. In: Lenz, H.-J., Decker, R. (eds.) Advances in Data Analysis — Proceedings of the 30th Annual Conference of the Gesellschaft für Klassifikation, vol. XVI, pp. 351–358 (2006)
4. Lourenço, A., Belo, O.: Catching web crawlers in the act. In: Sixth International Conference on Web Engineering, ICWE 2006 (2006)
5. Oikonomopoulou, D., Rigou, M., Sirmakessis, S., Tsakalidis, A.: Full-coverage web prediction based on web usage mining and site topology. In: IEEE/WIC/ACM International Conference on Web Intelligence (WI 2004), pp. 716–719 (2004)

The Problems of the Insolvency Register in the Czech Republic from the Perspective of Information Technology

Luboš Smrčka

University of Economics, Faculty of Business Administration,
W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic
smrckal@vse.cz

Abstract. The study concerns the situation in the Czech Republic, where the new *Insolvency Act* was accepted in 2006 and took effect from the beginning of 2008. One of the aims of the acceptance of the legislation was the creation of conditions for improved use of information technology in the area of insolvency proceedings. It was intended to bring about a state where a third party would easily acquaint itself with details concerning individual cases so as to compile sufficiently substantiating statistics on insolvency proceedings. While the first goal has been accomplished and it is now possible to access a range of information online and follow a concrete case practically simultaneously to the development of the insolvency proceedings, the second goal has remained unfulfilled. The fault lies in inadequate technological and jurisdictional solutions which have not paid sufficient attention to the possibilities of using information technology in insolvency practice in the Czech Republic.

Keywords: insolvency, information technology, statistics, bankruptcy, reorganisation.

1 The Significance of Information on Insolvency Proceedings

Insolvency proceedings are economic events that are frequently overlooked by economic theory. We cannot avoid the impression that they are often considered marginal events. This is evidently the real reason for the fact that relatively little attention is devoted to them in professional literature and when their procedures and impacts are examined more broadly, it is undertaken mostly by attorneys, law theorists or judges. But economic theory merely takes their existence into account for the most part.

Despite this, the last ten years in particular have seen the increasingly frequent emergence of studies concerning the significance of insolvency proceedings on the basic phenomena of national economy. In particular, research has been conducted on the relationships between the efficiency of insolvency proceedings and realities such as the average interest rate in loans provided to commercial companies, dependant on the risks undertaken on the creditors' sides and on their position in the actual insolvency proceedings, and similarly, the influence of the efficiency of insolvency proceedings on other aspects of the behaviour of financial institutes [1]. Another

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 313–322.
DOI: 10.1007/978-3-642-36981-0_30

area is research regarding the degree to which the regulations on bankruptcy of commercial companies influence the development of enterprise from the perspective of globalisation [2].

It is interesting to note that these studies used and use as their departure point knowledge of general legal norms and detailed analyses of mere samples of bankrupt companies as our knowledge of complex statistical data on these lawsuits is markedly limited. As can be seen from the diction of the previous sentence, this fact does not concern the situation in the Czech Republic only; rather, we must assume that this is a phenomenon usual in most developed countries. For instance, in their studies, authors Lee, Yamakawa, Peng and Barney [2] appeal to the comparative statistics of *Doing Business* prepared by the World Bank together with the IFC [3]. However - as we shall show in the following passages – we cannot consider these statistics reliable as the resultant data does not command much trust upon closer scrutiny. By contrast, Davydenko and Franks [1] clearly do not trust the statistics provided by multinational institutions and have based their unusually inspirational studies on the samples of 2600 small businesses. They were able to do this primarily thanks to the fact that ten large banks representing significant market shares in markets of researched states provided a range of specific data.

But the question remains: Why do we actually need to know certain data on insolvency proceedings? The answer could be a remark made by Frank Borman, American astronaut and later director of Eastern Airlines: "Capitalism without bankruptcy is like Christianity without hell." This is because the demise of a commercial company due to its inability to achieve adequate economic results, due to the incompetence of the management, or owing to criminal activity is a thoroughly logical part of the whole economic system. In most cases (let us leave aside criminal acts for the time being) it is a question of correct allocation of resources and especially the free allocation of resources. We can express the thesis that the bankruptcy of a debtor and insolvency proceedings (which is essentially bankruptcy, thus the demise of a business and the liquidation of its assets in favour of its creditors) are mechanisms for freeing assets from the structure of an economic subject who is no longer able to fulfil its function. This is followed by providing these assets to such subjects who will put these assets to better use. From the creditor's point of view, the simpler, more controlled, faster and more profitable the process, the less the creditor (let us say a financial institution) will implement its risk of loss to a third party – in the given context, to other debtors. A logical and proven approach in the case of banks in an environment of poor and slow insolvency law is that they operate restrictively with a larger number of their clients and limit their loaning or require further collateral for loans. In all cases this means that creditors (investors, banks) assume higher potential expenses when enforcing receivables in the event of a debtor's default and also a higher risk that the whole transaction will make a loss. Let us take it as proven that high risks during the process of enforcing receivables and ineffective insolvency proceedings (that is, ineffective from the creditor's point of view) lead to a general price increase for loans [1-2], [4-5]. Davydenko and Franks [1] arrive at a clear conclusion arising from analysis of the research results of their sample of businesses – in countries where the creditor is better protected and in which the right of creditors to decide about insolvency

proceedings for a debtor in default is upheld, there are generally more liberal conditions on the financial asset markets and interested parties thus have less difficulty finding and utilising outside resources. It is of substantial importance that lower prices for these resources are also set (pp. 605 – 607).

We can therefore close this introductory passage by asserting that the quality and especially the efficiency of insolvency proceedings are basic attributes of the economic environment of a given country; when these processes are ineffective and of poor quality, an increase in the cost of outside resources for all economic subjects results. This is because all (if they use outside resources) who take loans must pay not only the price conditioned by the economic reality of the country and by the objective state of the relevant market, but must also pay a sort of "regional risk surcharge". Its amount (among other aspects) depends also on the extent of risk undertaken by a creditor in insolvency proceedings.

From the perspective of governments and states, the matter can be viewed from a different angle. Preferences expressed by the mechanisms and regulations of insolvency proceedings are at the same time a certain political problem, whereas like many political problems, this question does not have a single, specific economic answer. From the point of view of national economy, it is highly probable that in the long term the most effective procedure for insolvency proceedings would be to place the bulk of decision-making authority into the hands of the actual creditors within the boundaries of certain general guidelines laid down in a relevant act.

(We could lead a relatively sophisticated discussion as to whether the actual principle of collective enforcement of receivables as represented by insolvency proceedings is sufficiently effective. That is, whether it would not actually be more convenient to leave this area solely to the individual activities of the creditors regardless of, or with a certain non-dogmatic consideration of, the question of the prisoner's dilemma or to the problem of the "common pool" [6]. But let us take it as a given that this doubt is solved politically and a change of viewpoint on this area is impossible within the context of developed countries in the historically relevant time.)

In reality, however, this purely rational solution is not dominant as its economic basis surprisingly does not succeed in competition with other variants which are also able to pursue political goals. The most usual and frequently attacked of these goals is the retention of employment [7]. The result of the endeavour to include such a goal into insolvency law, then, is several more necessary steps. The authority of the courts must be strengthened and it is necessary to assign them the task of preserving the operation of the debtor's business for as long as possible. But this means that it is necessary to limit the rights of the creditors as the interest of the creditor is often completely different than cooperating to accomplish a political aim. On the contrary, it is also necessary to strengthen the rights of debtors - it is in the interests of a debtor in bankruptcy that the creditor be forced to cooperate in the retention of the business. Insolvency law then becomes an instrument for politically shaping economic reality, which necessarily leads to the growth of transactional costs of creditor subjects and, among others, to these costs and risks being implemented into the cost of money.

In general practice, most preferences incorporated into relevant laws (and possibly regulations) during the legislative process are far less manifest. We refer to the

favouring of certain creditors. Employees (whose receivables are most often unpaid wages) in particular are at issue; in several insolvency systems employees' receivables have a special status which is often defended in the language of the 19th century and theses on the "powerlessness" of employees. Let us grant that we can understand this at least from a human perspective insofar as the insolvency of the employer can in several cases subject employees to truly unreasonable losses. In reality, however, this is part of the social system created by the government, whereas this part is involuntarily prescribed to other creditors, which is otherwise nonsensical as these creditors have no responsibility towards the state social system and do not bear any responsibility for the debtor not having paid its employees their wages. The preferential treatment given to the receivables of state institutions is totally unfair. Where tax or insurance is concerned, for example, there is a possibility to get a better ranking for satisfying receivables compared to other creditors through the institute of judicial right of lien.

As a result, a range of models for organising insolvency law and the adjustment of the whole insolvency system (which more or less fulfils a political goal) are used instead of a clean economic solution of the problem. This mostly means that short-term economic goals are favoured; for instance, retention of employment positions and the debtor's business operations rather than opting for a long-term strategy of lowest possible creditor expenses, minimisation of risks for creditors and thus reduction of capital prices. Whereas a focus on short-term goals has an advantage from a political point of view insofar as its results can be quantified and proven, and that in a certain, brief period such a course of action leads to the retention of employment and production, opting for a purely economic solution brings clear success only after sustaining a particular adjustment of the system over a longer period. In comparison to the usual political cycle of four years, the medium-term horizon of ten or twenty years is at a significant disadvantage.

2 The State of Use of Information Technology

If states favoured a purely economic solution of the insolvency problem, we would probably not have any major difficulties with statistical data. Such a government itself would have to assemble the necessary data to a reasonable degree and verify the results of its policy. In reality, however, we see the exact opposite in many developed countries. In fact, there are no other official statistics on the course of insolvency proceedings other than those which probe the number of cases, number of individual methods of settlement, usual duration of proceedings and other essential albeit (in reality) descriptive information. In truth, we have no knowledge of insolvency proceedings that would enable systematic analysis of the course of these proceedings and an evaluation of their efficiency.

The situation in the Czech Republic can serve as an example. In the context of discussions regarding the concept of insolvency law, which took place at the turn of the last and this century, one of the main themes was also the implementation of information technology into the whole insolvency system. The aim was primarily to

strengthen the knowledge of participants and non-participants regarding its procedures. Statistics were a further theme of discussion, that is, the creation of a method of providing information about the course of insolvency proceedings as a whole, which would enable the quantification of results from proceedings into outputs of a statistical character.

On the primary level, legislation assented to the complete publication of data on all insolvency proceedings regardless of their general societal relevance. Act 182/2006 Coll., *on Bankruptcy and Ways of its Solution* (the *Insolvency Act*) implemented the institute of the insolvency register [8] as a publically accessible "stockpile" of all operations registered by insolvency courts in every individual insolvency proceeding. As of the coming into effect of the act (1 January 2008), it has been possible to find all relevant steps in this information system via the internet without having to register or take any further steps. From an awareness point of view, this represents significant progress and, without doubt, a textbook example of the use of information technology for increasing the transparency of insolvency proceedings. Interested and professional public can now follow proceedings "online", and thanks to the existence of deadlines the danger of information being misused by insiders, although not entirely eliminated, has certainly been considerably reduced. All information is accessible to all who are prepared to outlay the relatively negligible expenses (for instance, in terms of time) necessary to access the information. In comparison to previous practice, this represents a fundamental change.

Procurement of information is very simple from the angle of information technology – the search form makes it possible to enter the name of the company or name and surname of a natural person in insolvency, the identification number (Id. No.) of the company or date of birth (or birth certificate number) of a natural person, or file number; it is also possible to search according to individual legal acts during insolvency proceedings (actions) or according to the state of the proceedings (last actual defined shift given by a court ruling). So one can, for instance, search commercial companies or individuals which the court has declared bankrupt within a specific defined period. Yet the search system itself is very problematic as we shall soon demonstrate.

After searching for a concrete case, the interested party can open all documents the court has stored in the insolvency register, that is, not only documents compiled directly by the court, but also those submitted to the court by participants of the proceedings or other relevant parties connected with the proceedings in question. In the vast majority of cases, these are documents in pdf. format. Where more complex cases are concerned, this leads to the opposite extreme – in the flood of data it is almost impossible to find relevant information. For instance, on the 19 November 2012 the insolvency register contained a total of 4454 documents in the case of the proceedings with the Sazka lottery company, whereas 3182 of these were applications filed by creditors and steps concerning these applications. The difficulty lies in the fact that these documents are generally divided only into those concerning the proceedings before declaration of bankruptcy, proceedings after declaration of bankruptcy, documentation of incidental lawsuits, a section of "other" and finally "receivables". Inside these individual sections the system offers no further search

possibilities in terms of organising the insolvency register. Searching is therefore possible only by using the regular Ctrl+F browser function used by the party interested in the information. But owing to the fact that document titles placed in the insolvency register are often only partially standardised, searching by terms is problematic. Individual steps of proceedings have their own identification code. But in order to use these, the interested party would have to have at its disposal a list of these codes and their specifications.

Most importantly, however, the name of the participant (who filed the document and whom the document concerns) of the proceedings is not stated in the actual document titles. The name of the creditor is only exceptionally stated in the column of valid creditors, but this is only the case in a minimum out of the entire number of entries in the insolvency register. It is not possible to identify creditors from code identifications. In the case of Sazka this then means that it is unusually difficult to find a specific application for a receivable and the correspondence around this application (for example, its partial or complete withdrawal) as it is impossible to use a full-fledged term search. Searching according to the date of filing the item into the insolvency register works well, so finding the appropriate documents will be far easier if we know the date of the steps being searched.

Although the insolvency register was intended to be a common, modern and innovative information channel, even a perfunctory survey of the way it works shows many ways in which the information technology used does not allow the public to utilise the potential of the insolvency register in an effective way. Here we find a rather incomprehensible divergence between intention and result, whereas the result simply does not correspond with the possibilities offered even by a cheap technological design of the information system. A somewhat aggressive explanation for this divergence is offered – that it serves to fulfil a political task, that is, that the system on the face of it does indeed fulfil the demand for openness; this openness is, however, to a certain extent annulled by other circumstances. The intention would then be that the information provided should prevent effective assessment of the true state and quality of insolvency proceedings. This is obviously a tempting construction which is, however, almost certainly untrue. Pointing out standard bureaucratic procedure in similar cases offers a far more likely explanation. Political will in the area of the legislative process ended in the expression of the requirement that the insolvency register objectively and very quickly make public all information on investigations underway. In reality no clearly defined political assignment existed as to how such awareness should be, what parameters it should have, and to what extent is should enable the analysis of the real state of insolvency proceedings as a whole. The *Insolvency Act* itself contains no details that would more precisely define the properties of the insolvency register. In §2 letter i), it is only stipulated that for the purposes of this law "the insolvency register is understood as the information system containing data according to this law." There are in numerous other places requests for speed in entering new information into the insolvency register and several other details concerning its functioning; we nevertheless find no other request for information over and above the making public of files from insolvency proceedings or other processing of information entered.

The result is thus a system which fulfils the political assignment in the literal sense, but in reality offers practically nothing more. In the terminology of project managers we could say that this project was never pulled through to the end. An objective observer must admit that the bureaucratic apparatus acted highly economically when it clearly chose the cheapest solution which would incur the lowest expenses in terms of cost and effort for its implementation.

Hence the result of fulfilling requirements at a secondary level, defined by the discussion underway before the acceptance of the *Insolvency Act*. In it, the requirements were defined for collecting information and processing it into statistical data of a type which would enable the meaningful analysis of the development of insolvency proceedings as a system. The insolvency register fulfils absolutely none of these requirements. When searching according to the state of the proceedings or action, it is not possible to enter a period longer than two weeks. This means that if we wanted to ascertain all proceedings during the calendar year which progressed to the approval of the insolvency administrator's closing report, we would have to repeat the search at least 26x. But the situation would be still worse – the system of search is underdimensioned, so it is capable of selecting only a rather limited amount of cases from the entire number of proceedings. With the majority of actions or states of proceedings it is therefore necessary to select an interval still shorter than 14 days as the amount of outputs during an interval of 14 days is too high and the system cannot handle it. Most of the time, therefore, the possibility of search is limited even to a mere week, but if we were to search certain quite regular steps which courts or participants of proceedings frequently perform, we would have to shorten the search interval even more.

Concrete cases attained in this way, where approval of the insolvency administrator's closing report from the sides of the creditors and the insolvency court was obtained, cannot be processed at all within the system. It is possible to study them, but further searching, specification of selection and so on is not made possible. The obtained file, moreover, contains both commercial companies and natural persons – entrepreneurs listed in the commercial register on the one hand, and also natural persons in insolvency on the other hand. In this way, the results of searching become markedly unclear.

As we can see, the whole mechanism clearly prevents generalisation of available information and its aggregation; this is especially so when we become aware that government institutions do not have any further data on insolvency proceedings, and if they do, this data is not made public according to a clearly defined and reliable standard. From the point of view of the insolvency administrator and also the courts, no obligations are set which would allocate one or another group the task of gathering and providing data on individual cases in such a format that would enable further work with this data.

We again find a marked divergence from public interest, which would be much better handled if reliable statistical data existed regarding the course and results of insolvency proceedings. These would enable far better definition of the need to modify insolvency law so that it could be clear to the highest degree and cheapest for the creditor, so that risks could at the same time be reduced and a contribution be made towards the expedient solution of insolvency proceedings.

3 Departure Points for the Czech Republic

In the same way that insolvency law is constantly developing, it is clear that it will be necessary to revise the activity of the Czech insolvency register and also of similar institutions in other countries. The example of the Czech Republic is very significant and it is possible to consider it as a model. Numerous developed countries have in recent years moved towards the liberalisation of access to information on insolvency proceedings; in few places, however, has it been accompanied by the generation of mechanisms which allow information gained to be used otherwise than for research of (sometimes even rather complex) concrete cases.

As far as our knowledge on the course and efficiency of insolvency proceedings is concerned, we rely on the statistics issued by the World Bank and IFC in the regular publication *Doing Business* [10]. In reality, however, we are highly sceptical of these numbers, both regarding the environment of the Czech Republic and also as far as the data for other states are concerned. Although this publication presents itself as a set of information of a statistical type, in numerous matters these are mere estimates; estimates, moreover, the methodology of which could be cast into doubt. As regards data on insolvency proceedings, this is precisely the case.

Table 1. Duration of proceedings from declaration of bankruptcy and creditor yields in the CR

Year	Duration of insolvency proceedings (in years)	Creditor yields from debtor bankruptcy (% receivables)
2002	**9,2**	**15,4**
2003	**9,2**	**15,4**
2004	**9,2**	**16,8**
2005	**9,2**	**17,8**
2006	**9,2**	**18,5**
2007	**6,5**	**21,3**
2008	**6,5**	**20,9**
2009	**6,5**	**20,9**
2010	**3,2**	**55,9**
2011	**3,2**	**56,0**

Source: Doing Business 2012, http://www.doingbusiness.org/custom-query, [11].

As we can see from the table, according to the statistics of the World Bank and IFC, the yields for creditors from insolvency proceedings in the Czech Republic at the given time should be 56 percent, which must be considered a highly improbable possibility. In reality, the yield from insolvency proceedings are much lower – even though no relevant statistical substantiation exists at the given time for this, all practical experiences with the development of bankruptcies and reorganisations suggest that this is the case. It is improbable that yields would be higher than thirty or forty percent as a yield less than fifty percent is attained only in exceptional cases, whereas in a marked number of cases there is in fact zero financial fulfilment for the creditor.

If we were to take the data of the World Bank and IFC for the assessment of the development of insolvency proceedings in the Czech Republic over the last ten years,

we would have to deduce considerable progress and a positive improvement of the situation. In reality, though, there has been no such development.

These misleading signals about positive development are somewhat confusing and could lead to completely erroneous conclusions in the area of other political decisions. Similar assessments are totally irrelevant from the angle of economic practice, however, given the fact that subjects operating on the Czech markets are acquainted with the true situation. Without needing any more precise statistics, they know from personal experience that the data given in international comparisons is not relevant.

What does occur, however, is a broadening of the divergence between world economics and political decisions which unfortunately are not usually classified on the basis of economic practice, but far more on the basis of statistical data.

4 Conclusion

For a correct and detailed knowledge of the course and efficiency of insolvency proceedings it will in the future be necessary to reform the insolvency register to a certain degree, improve its search and selective capabilities, but it will especially be necessary to change certain duties of insolvency administrators and entrust them (on pain of penalty) with statistic-type information duties. This would enable finding a range of data in a separate report, especially the following information: identification of the debtor; identification of the insolvency administrator; date of filing the petition for bankruptcy; date of ruling on bankruptcy; date of decision on the manner of settling the bankruptcy with denotation of the manner in which the bankruptcy is to be settled; date of possible changes in the manner of settling the bankruptcy with a variant identification of the type of change; amount of deposit for costs of the insolvency proceedings if it was paid, and who paid it; the number of days that elapsed between the making public of the insolvency proposal and the making public of the ruling on the bankruptcy (a ruling which has come into legal force); the number of applications of secured creditors and the number and amount of their receivables (whether recognised or unrecognised); the amount of satisfaction for secured creditors; the number of applications from unsecured creditors; the amount of receivables beyond the property base or receivables deemed equal to them and their volume; the number of incidental lawsuits conducted by the insolvency administrator; the volume of receivables in the lawsuit; duration of individual lawsuits; the volume of receivables arising after the announcement of the manner in which the bankruptcy is to be settled (including costs of the preliminary insolvency administrator); estimate of the property base: cash + estimate of liquid assets + estimate of fixed assets + other; yield of liquidated assets (liquid and fixed), remuneration and costs of the administrator. Subsequent processing of this data by means of information technology would enable one to gain outputs which would analyse the course and efficiency of insolvency proceedings in a relevant way.

Owing to the ever-increasing complexity of economic relationships and the fact that cooperative links are extremely unclear, developed countries will sooner or later have to significantly change their approach to gaining information from insolvency

proceedings and especially to gaining information on the course and efficiency of insolvency proceedings. Insolvency law is a highly sensitive mechanism where every intervention into its functioning could lead to unexpected and especially to unintended effects. It would be highly irresponsible to decide for a change in insolvency law without knowing the mechanism and efficiency of the present system very precisely at the same time. This will entail substantially more operative and, most of all, more thorough integration of information technology into the whole system, whereas (using the example of the Czech Republic) it can be proven that the present state is unsatisfactory and thoroughly inadequate.

Even though we have been witnesses in the past few years of a considerable facilitation in sharing information about insolvency proceedings, thanks among others to the internet and other modern media, inadequate political assignments have resulted in these endeavours ending in a mere general declassification of data on individual insolvency proceedings. As in other areas, information technology could also enable further reductions of transactional costs for creditors in the insolvency system, which could potentially lead to a marked increase in the efficiency of these processes.

Acknowledgments. The article was written as an output for the research project *"The development of transaction costs among Czech economic subjects in insolvency proceedings, the possibilities of their reduction to a level standard in the EU with the aid of improved legislation, the possibilities of improving insolvency proceeding statistics and creating a household financial fragility model",* registered at the Technological agency of the Czech Republic (TA CR) under the registration number GT309022.

References

1. Davydenko, S.A., Franks, J.R.: Do Bankruptcy Codes Matter? A Study of Defaults in France, Germany, and the U.K. The Journal of Finance LXIII(2), 565–607 (2008)
2. Lee, S.-H., Yamakawa, Y., Peng, M.V., Barney, J.B.: How do bankruptcy laws affects entrepreneurship development around the world? Journal of Business Venturing 26(5), 505–620 (2011)
3. International Finance Corporation and The World Bank: Doing Business, http://www.doingbusiness.org/
4. Richter, T.: Insolvenční právo. ASPI Walters. Kluwer, Prague (2008)
5. Brealey, R.A., Myers, S.C.: Principles of Corporate Finance, 9th edn. McGraw-Hill (2007)
6. Smrčka, L.: Filosofické aspekty vztahu insolvenčního a exekučního práva. Komorní Listy 1(4), 27–33 (2012) (in Czech)
7. Blazy, R., Chopard, B., Fimayer, A., Guigou, J.D.: Employment preservation vs. creditors' repayment under bankruptcy law: The French dilemma? International Review of Law and Economics 31(2), 126–141 (2011)
8. https://isir.justice.cz/isir/common/index.do
9. Kotoučová, J. a kol.: Zákon o úpadku a způsobech jeho řešení (insolvenční zákon) - komentář. C.H.Beck, Praha (2010)
10. http://www.doingbusiness.org/
11. http://www.doingbusiness.org/custom-query

Improving Public Transport Management: A Simulation Based on the Context of Software Multi-agents

Marcia Pasin and Thiago Lopes Trugillo da Silveira

Centro de Tecnologia, Universidade Federal de Santa Maria
Av. Roraima 1.000 - Cidade Universitária - CEP 97.105-900 - Santa Maria - RS, Brazil
{marcia,thiago}@inf.ufsm.br

Abstract. The Simulation is a key element in efficient management and operation of public transport. This activity includes processing of dynamic data possibly obtained through sensors, as well dealing with static information including map digitalization. A fundamental problem is that, although available, systems that perform these activities are proprietary solutions, have high cost of acquisition, are difficult to be used by non-experts, or offer a limited number of features, without considering important aspects such as changes in travel demand and traffic flow. To address this problem, this paper presents a friendly simulation for public transport systems implemented with support of open technologies. Dynamism to modeling bus and travelers behavior is offered by a MAS (multi-agent system). Our simulation was experimentally evaluated using as scenario a medium-sized city with different travel demand and traffic flow.

Keywords: simulation, multi-agent system, public transport.

1 Introduction

Traffic is an issue that has been gaining incrementally attention in the daily. And it is known that the complexity of traffic management only tends to increase due to the substantial increase of demand in recent years and the trend of growth of demand expected in the coming years. Increasing complexity of traffic management has been required by responsible agencies and users the conversion and use of improvements, best practices and rules to ensure more efficient operation. A recurring question, in the case of transit, is the lack of system and tools, which can be used by non-expert users (mayor, politics, managers) to enable low-cost planning, monitoring and executing simulations of the process of urban public transport. Lack of investment in these tools results in management difficulties, which impacts on expected quality of service (QoS).

The lack of an efficient public transportation is a major obstacle to the development of smart cities. As solutions to these difficulties and aiming towards sustainable mobility, information and communication technology have been applied, with the primary outcome the Intelligent Transport Systems (ITS). Aiming the

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 323–329.
DOI: 10.1007/978-3-642-36981-0_31

provisioning of suitable computing support, ITS are being increasingly developed. ITS enables analysis, prediction and diagnosis to existing problems in traffic process. An important aspect concerning ITS design is the development of tools, which allow transit operators and municipalities friendly and accurate access to information on urban public transport. These tools should provide information to allow operational and managerial planning and improved QoS.

In the case of transit operators, interesting information to allow planning and management includes: *(i)* current position of the vehicles in the system, *(ii)* number of travelers transported daily, *(iii)* highlighting paths (bus lines) more burdened (or more idle), etc. It is also important to provide features to facilitate planning such as those that measure the effects of a service level (which is a function of traffic volume). For instance, given a certain level of service, which is the number of cars to be placed in a particular bus line? For other actors in the system, for instance, an emergency service may be interested to know what is the average time that the rescue vehicle needs to reach the scene of an accident, given a certain level of service on roads.

However, a major drawback concerning public transport is the unavailability of solutions to allow planning, information provisioning, simulation and real-time monitoring. There are exceptions ([3] [4] [5] [6]), but in general, systems and tools lack in not considering important aspects (convoy detection, delayed cars, overload lines) or offer a limited number of features (without considering changes in travel demand and traffic flow). Lack of investment in these systems and tools results in difficulty to management and to travelers usability, that impacts on the QoS. Considered an important element in this process, software tools enable evaluation of situations, allowing the observation of different scenarios and anticipation of problems before implementing solutions into practice.

In this context, this article describes the implementation of a graphical simulation to provide information and experimentation about urban public transportation mainly to transit operators. We aim to hand three main constraints: *(i)* the simulation is target to the public transport scenario and specific requirements, which require to deal with static information about cars (such as identification) and their associated lines (name of the line, source and destination) and real maps as well to deal with dynamic information about car location, average car speed, fuel consumption and the number of travelers entering or leaving a car, etc., *(ii)* the simulation can be executed in different simulation times (faster and slower than real time), *(iii)* reports are generated with regard to convoy detection, the most delayed cars, the most overload lines, and other QoS metrics, such as number of sitting and standing travelers. To modeling transit components in this dynamic scenario, we are using as support a simulator following a multi-agent system (MAS). In fact, there are many opportunities to use MAS in traffic and transit systems [1].

In a general way, interesting features of this simulation tool include: *(i)* the use of free software and *(ii)* usability because the tool needs to be easy to use by non-experts. Once implemented completely the tool, simulations can be made facing traffic through the *(i)* addition and removal of a bus stop on the way, *(ii)* addition and removal of a bus on the lane, or *(iii)* stopping a vehicle for simulating accident. Through this tool, an ordinary user can perform simulations and forecasts to help in

the process of urban public transport management. The resulting software can also be used by government institutions and municipalities to assist in planning and monitoring of urban public transport.

In general, the tool proposed here aims to serve as a mechanism to enable better understanding and improving the process of transportation in the cities of medium and small size. Furthermore, such information may be used by licensees companies in order to anticipate demands and take proactive actions to improve quality of transport services and fleet management.

The rest of the paper is organized as follows. Related works are presented in section 2. Section 3 presents the architecture of the simulation tool to public transport. Section 4 details aspects of implementation using as scenario a number of lines serving the city of Santa Maria (in Southern Brazil) and a simulator called Siafu as support. Finally, Section 5 describes conclusions and future work.

2 Related Works

In the context of public transport management, existing simulation tools are academic solutions ([3] [5] [6]) with limited features, or are mainly proprietary solutions (to public transport see [4], and more general see VisSim available at http://www.vissim.com, SUMO available at http://sumo.sourceforge.net, and ITSUMO) and, therefore, are not user friendly. We agree that systems and tools must be based on open technologies, aiming low cost of deployment, flexibility, user friendly issues and must deal with important aspects such as changes in travel demand and traffic flow (e.g. presence of rush hours in some periods of a journey). Finally, efficient performance is also a required feature. Another solution that needs to be commented here, due its general nature (not only concerning public transport) and popularity, is the Google Traffic service, which is the widely available service offered by Google Maps (http://maps.google.com.br). This service estimates traffic behavior is a sort of metropolises in almost real time. Traffic information is provided with the support of the concept of crowd sourcing. Basically, users of transport system while accessing Google services on their mobile devices, implicitly send information about their current position and speed (captured via GPS) to Google data center. Based on the received data and historical data, the tool provides an estimation of the traffic, which is displayed using different colors on a map from Google Maps, via web interface. However, this solution lacks mechanisms to manage and observe a certain fleet in a more specific way. It only offers a transit and traffic macro vision.

3 Providing Information to Transit Operators

Aiming to provide information about fleet status (in post-operation) to transit operators, we developed a simulation over a real map. The main elements of the simulation include data entry, processing (an implementation supported by Siafu simulator) and visualization of information (using a graphical interface and reports). They are shown in Fig. 1, in a context of an infrastructure, described in previous work [2] which also support the information system to transit operators and an advanced traveler information system (ATIS) to end users.

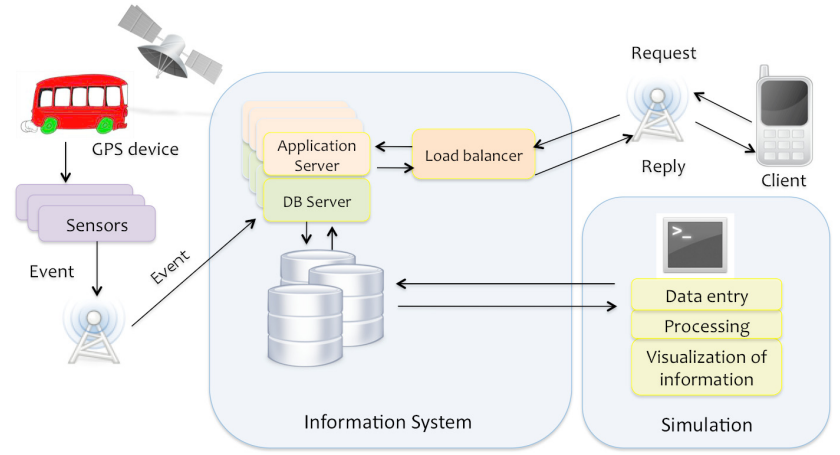


Fig. 1. System architecture and simulation components

The data entry module is supplied with static information about the bus stops and routes which are previously stored in a file using CSV format (and in the future in a database). Sensors placed anywhere in the infrastructure gather information about bus positioning, velocity, volume in the fuel tank and send it in real time to update the CSV file. This information is retrieved every pre-defined interval of time. Alternatively, the same infrastructure can be used to provide information to travelers (toward an ATIS), in displays or mobile devices.

4 Implementation

The simulation tool was implemented in two essay versions. A first implementation essay was conducted using Google Maps API version 3 and PHP language aiming to build a web available tool. However, the interface has many dynamic elements (in movement on the screen) and it was difficult to play in a remote browser. Thus, we decided to use the support of a simulator toolkit (Siafu). The used simulator toolkit and simulation features are described in the following. In both implementations, vehicle location is executed using information previously stored in a CSV file. However, in future, the idea is to have an implementation to work with the support of real-time obtained data.

Siafu Simulator. Siafu (http://siafusimulator.sourceforge.net) is a simulator toolkit developed in the context of MAS. To programming using Siafu, three basic classes must be extended: BaseAgentModel, BaseWorldModel, and BaseContextModel. These classes address respectively the agents' behavior (agents that may have different actions within the simulation, e.g. cars), places (there are specific places where cars can circulate, e.g. lines) and overlays (i.e. regions on the map that are intended to cause changes in simulation). The child classes (classes that extend the

three aforementioned base classes) need to override the methods create*x* (where *x* represents agents, places or overlays) and *doIteration*. The first is responsible for initializing each of the models (agent, and world context) and the last updates each of them every time an iteration happens in the simulation.

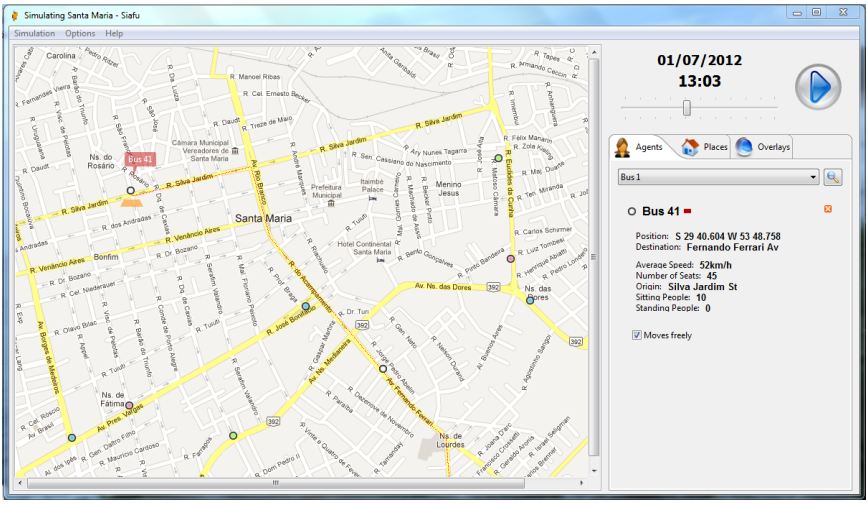


Fig. 2. Tool screenshot of the current implementation

The Simulation Features. In fact, the simulation in question aims to show how the operation of the bus lines service on a given micro-region (i.e. the city of Santa Maria) works. The map of the micro-region was obtained via Google Maps. Therefore, we have been implemented cars (agents) that travel along this micro-region as an extension of the class Agent (Siafu *). Any car holds information about them such as: origin, destination, number of sitting travelers, number of standing travelers, speed mean, etc. The simulation of the process of a car arriving on a bus stop is done by assigning the *bus stop* agent to certain places in Siafu simulator. Similarly to *bus stop*, it can be also identified other locations such as departure and arrival points. Thus, we have implemented our bus line service creating multiple agents starting from a departure point and going to the same final point (the bus garage), and go through the *bus stop* along the way. In each Siafu *place*, each *bus stop* has information on how many people are waiting for a bus. This amount of people changes over time and also changes when a car is in that place. For simplicity, in the current implementation of the simulation, the bus has constant speed. They remain at constant speed and always going from one place to another, using as start a departure point and ending at an arrival point, or a *bus stop* when one finds a *bus stop,* and wait for embarkation and disembarkation, as occurs in reality. This part of the simulation is done through a *waiting call* (in this case the bus speed in question is zero), multiplying the number of people who will get on the bus by a constant (time average it takes each person to get to the bus). Fig. 2 shows a screenshot of the current implementation. In the screenshot, the graphical tool has two different areas: *(i)* the

map and its elements and *(ii)* an information area with a mechanism to manually change the simulation speed.

Convoy Detection. The proposed tool also implements the automatic convoy detection and provides reports that are described in the following. A convoy consists of a group of cars (two or more) assembled to travel together in accordance with a certain destination. The goal of a convoy is to increase the flow of embarkation and disembarkation in a regular *bus stop* to avoid congestion on dedicated bus lanes. Typically, congestion at a point of interest is caused by the time required for loading and unloading passengers in a car. The efficiency of the convoy depends on the average size of the convoy (TMC) and also on the flow of users it can serve. With features provided by Siafu, convoy detection is an easy service to be implemented. Siafu provides a method called *isNear,* which returns true if an agent (i.e. a car) is closed to another. Since the method *isNear* is applied to all cars in the simulation, a convoy of any TMC will be eventually detected.

Reports. Finally, another feature provided by the our simulation is the generation of a set of reports such as a list that describes the most delayed cars, the most overload lines, and other QoS metrics in terms of number of sitting and standing travelers. These reports are generated at the end of a simulation journey as a HTML file or as graphical information. For instance, Fig. 3 shows a graphical information with the number of sitting and standing travelers with regard to the bus service frequency every 30, 60 and 90 minutes respectively.

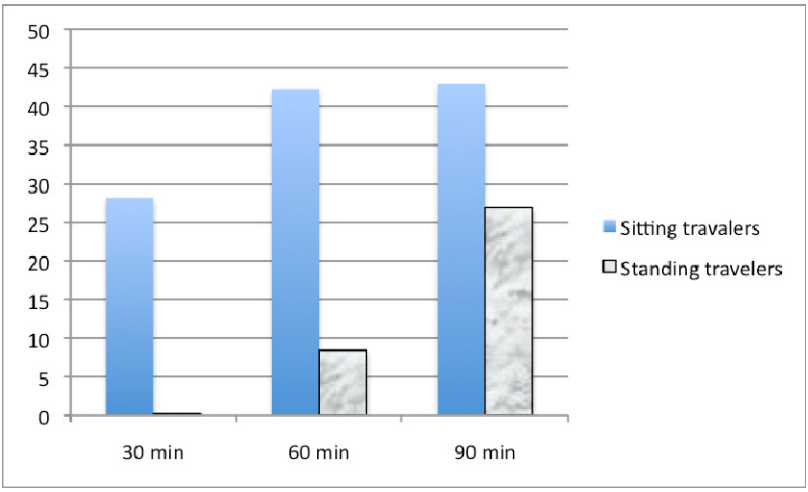


Fig. 3. QoS regarding to sitting travelers versus standing travelers

It should be noted that to the input scenario, to provide a high QoS to travelers (in terms of reduced of no standing people at the bus), a bus service must be started every half an hour.

5 Conclusions and Future Perspectives

This paper presented the current implementation status of simulation tool to monitoring and planning of public urban transport process. The features provided by the tool are convoy detection and a sort of reports with statistical information about the most delayed cars and the most overload lines, for instance. The tool is supported by the concept of software agents. Once completed the implementation of the simulation tool, the idea is to use different cities as scenario and to incorporate different behavior to drivers by MAS. Furthermore, such information and detailed information about positioning of cars can be utilized for traffic behavior analysis in future demand modeling. Finally, we believe the simulation presented in this paper can be used in conjunct with other ITS to improve QoS in public transport service.

Acknowledgment. This work has been conducted in the context of the projects CTIC SIMTUR Cidades Inteligentes project that is financially supported by the Rede Nacional de Pesquisa (RNP)/Brazil and CNPq/FAPERGS RS-SOC PRONEX under grant 10/0049-7.

References

1. Bazzan, A.L.C.: Opportunities for multi-agent systems and multi-agent reinforcement learning in traffic control. Autonomous Agents and Multiagent Systems 18(3), 342–375 (2009)
2. Jaques, P.A., et al.: Provendo informações para atores do sistema de transporte público: um passo na direção de sistemas inteligentes de transporte. In: XXVI ANPET - Congresso de Pesquisa e Ensino em Transportes, pp. 1792–1803 (2012)
3. Gruer, P., Hilaire, V., Koukam, A.: Multi-agent approach to modeling and simulation of urban transportation system. In: IEEE Systems, Man, and Cybernetics Conference, pp. 2499–2504 (2001)
4. Ladeira, M.C., et al.: Monitoramento da operação de transporte público: o caso de Porto Alegre. In: Congresso de Pesquisa e Ensino em Transportes (2009)
5. Molina, M.: An Intelligent Assistant for Public Transport Management. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005, Part II. LNCS, vol. 3645, pp. 199–208. Springer, Heidelberg (2005)
6. Meignan, D., Simonin, O., Koukam, A.: Simulation and evaluation of urban bus networks using a multi agent approach. Simulation Modeling Practice and Theory 15(6), 659–671 (2007)

Social Networks Mining Based on Information Retrieval Technologies and Bees Swarm Optimization: Application to DBLP

Drias Yassine and Drias Habiba

USTHB, LRIA, Department of Computer Science, Algeria
yassinedrias@gmail.com, hdrias@usthb.dz

Abstract. Unlike the previous works where detecting communities is performed on large graphs, our approach considers textual documents for discovering potential social networks. More precisely, the aim of this paper is to extract social communities from a collection of documents and a query specifying the domain of interest that may link the group. We propose a methodology which develops an information retrieval system capable to generate the documents that are in relationship with any topic. The authors of these documents are linked together to constitute the social community around the given thematic. The search process in the information retrieval system is designed using a bee swarm optimization method in order to optimize the retrieval time. Our approach was implemented and tested on CACM and DBLP and the time of building a social network is quasi instant.

Keywords: social network, knowledge mining, information retrieval, BSO, DBLP.

1 Introduction and Motivation

Social networks have emerged with the advent of the Internet and are nowadays of unquestionable importance. The main objective of this study is the development of a methodology for discovering social networks from a collection of documents. Information retrieval technologies are exploited to reach the target, which is to build social networks from textual documents. The system we developed generates a group of individuals who have the same center of interest that is specified in a query. Such cluster stimulates its members to constitute an effective social network that may provoke interaction between each other. The problematic of searching authors 'who work on the same topic as me' is one of the recent phenomena that we observe these last years. Indeed researchers often try to build their proper social network by their own. Our contribution helps in that sense and offers a tool that is capable to yield such community in an automatic way.

Of course, since the idea of dealing with such issue is based on information retrieval technologies, the system basis has to be implemented in an effective way. We know that information retrieval distinguishes two main stages, indexing and retrieval. Indexing consists in translating the documents in a format that will allow

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 331–343.
DOI: 10.1007/978-3-642-36981-0_32

making a quick access to their words. Lex, the lexical analyzer generator most commonly spread, notably in compilers construction, is used for the implementation of the index and the dictionary.

For the design of documents retrieval, we used a robust method that can handle a large collection of documents and which is based on a bio-inspired approach called Bee Swarm Optimization (BSO) [3]. The whole system was implemented and tested on two benchmarks, namely CACM, which is a collection of articles published by ACM and DBLP, which is a database that contains bibliographic information on journals and conference proceedings in computer science.

1.1 Related Works

Social networks mining has been widely studied this last decade and many works have reached important advancements in the area. Applications to domains like biology, economics and marketing have been also undertaken and the results were promising. Among the main issues that interest the researchers are the mining methods, community identification and modeling social rating networks. Lots of investments such those of [2, 5, 6, 9, 11] have been devoted to community identification. All of these articles build the social network from a given large graph and differ from each other by the method designed to extract the community structure. In [5], the authors focused especially on web communities and in [10], the authors considered the marketing application. Other axes that were also investigated concern the analysis of the social network as in [7] and the scoring and evaluation of the social community as in [9, 10]. The community scoring function quantifies how 'efficient' is the community.

On the other hand, information retrieval has known extremely interesting developments for more than four decades. The general concepts and techniques are well described in [1, 8, 12, 13].

Finally BSO, the bee swarm optimization approach is introduced in [3] and one of its applications to web information retrieval is published in [4].

2 A Brief Review on Information Retrieval

In this section, the main concepts used in the design of our system are presented in a brief manner. Indexing and searching documents are especially reviewed.

2.1 Lex as a Tool for Documents Indexing

Indexing documents consists in recognizing the significant words of the documents and insert them in data structures namely, the inverted file and the dictionary for use in the matching process between the documents and the queries. Lex, a lexical analyzer [8] is a tool for automatically and rapidly implementing a lexer for a programming language or a language defined by any language designer. Lex is widely used in compilers but it is also prevalent in many areas that require the recognition of words, such as word processing and natural language. As the crucial indexing step is based on the recognition of words to classify them according to their importance, it

was necessary to find a fast, effective and robust way to implement it. This reason motivates our choice of Lex as a working tool. The latter generates automatically the indexing program from a source of specifications of words and actions that must be performed after their recognition.

Afterwards, the indexing program when executed produces three output files: the dictionary, the documents file and the inverted file, which will be used in the matching process. The dictionary contains all the relevant terms that appear in the documents as well as their total frequency and their positions in the two other files. The document file includes the documents with their terms, each document with its associated terms and their frequencies. Finally, the inverted file contains all the relevant terms, each one followed by the list of the documents in which it appears in addition to its frequency in each document.

2.2 Traditional Matching Techniques

Searching for a document is a matching process between the document and the query. Such operation is handled by the means of a measure that calculated the similarity that may exist between the content of a document and the one of the query.

Similarity Measure. The information retrieval model that we consider in our study is the vector model because it fits well with the approach BSO that is used in addition to the matching step to accelerate the search process for the huge datasets. In this model, documents and queries are represented by vectors of words. The similarity between a document and a query is the degree of similarity between both entities in terms of the number of common words they share. The higher value of the similarity, the better is the response quality for a query. Several formulas exist to calculate the similarity between a document *d* and a query *q*, among them we chose the cosine described as follows:

$$\mathrm{Sim}(d, q) = \sum_i (a_i * b_i) / (\sum_i (a_i)^2 * \sum_i (b_i)^2)^{1/2}$$

ai and *bi* represents respectively the weight of the i^{th} term in the document and the weight of the same term in the query. The weight of a term in the document is often determined by the expression *tf* * *idf* where *tf* is the term frequency in the document and *idf* represents the frequency of the term in the other documents.

Search Process. There is a straightforward technique that can be thought to search for the document that is the most similar with the query. It consists in crossing all the documents of the collection and calculating the similarity between the documents and the query. At the same time the process sorts the documents in decreasing order according to the computed similarity between the query and the document. The search algorithm in this case uses the documents file. At the end the relevant documents are displayed thereafter.

This method is effective in terms of solution quality, but it is expensive in terms of execution time as it crosses sequentially all the documents in the collection. If m is the number of documents and if n is the number of words, the worst case complexity of the naive search is Θ (n * m).

Another more effective way of performing the same task is to keep the same search principle but executing it on the inverted file. As the latter is organized such that we can access to any term of the documents, the algorithm will access to only the terms that appears in the query. As we have already seen in the previously the inverted file associates with each term of the collection the list of documents in which it appears. Therefore it suffices to find the terms shared with the query and calculate the similarity between the documents associated with the terms and the query without having to go through all the documents. And this way, it will exclude implicitly the consideration to all the other documents. Therefore the search time will be significantly reduced. This technique has been used for many years because of its efficiency. The algorithm is therefore as follows:

```
Input:       the inverted file and the query q
Output: the relevant documents that are the most similar to the query
begin
     initialize queue to empty;
     for each term i of the query
     do
            for each document d associated with i
            do      compute Sim (d, q) ;
                    insert d in queue if it is not in it according to
     the
                    decreasing order of the computed similarity;
            done ;
     done;
     display the documents in queue;
end

Procedure Sim(d, q)
begin
     Sim :=0 ;
     for each term i of d
     do
            for each term j of q
            do
                    if ( i = j ) then Sim := Sim + (Pi * Pj) ;
            done;
     done;
end;
```

3 Bee Swarm Optimization BSO

BSO (Bee Swarm Optimization) is an intelligent approach designed for problem solving. It is inspired by the natural behavior of bees and reflects the process of searching for food to solve complex problems. It was described in [3] and applied for the first time to the satisfiability problem. Then it was also adapted for web information retrieval [4].

BSO is based on an experiment that was conducted by the biologists Seely, Cazmazinz and Sneys in 1991. The results stipulate globally that the bees when searching for food consider the richness of the source despite the long distance that may separates it from the hive. The second important income is that bees communicate between themselves through a dance that indicates the direction, the

concentration in food and the distance to the source. This phenomenon is translated to an algorithm whose objective is to search for solutions for a given problem. The good solutions simulate the interesting sources of food. The searching process translates the bee behavior and the selection of good solutions through the dance table the communication of interesting regions between bees.

To solve a given problem, BSO considers a solution space in which the search is performed. The search starts with a single bee called BeeInit which is located randomly in the solution space. It begins by determining strategic points in the search space where to perform the searches. The set of these points constitute the search area. The technique used to define the latter is based on a diversification generator, which is a process of finding positions that are as distant as possible from each other. These dispersed positions are then assigned to *k* bees in order to explore the neighborhood. In other words, each bee will seek the best possible solution in its mining area. The solutions found by the *k* bees that have good characteristics (measured by an evaluation function f) are then stored in a table called Dance. At the end, the best solution is considered to behave like BeeInit and the whole process is iterated until a solution with a good quality is found.

3.1 BeeInit

The search space or solution space is a set which contains all the possible solutions that are good or bad. *BeeInit* simulates the scout bee in the algorithm. It corresponds to a randomly chosen solution and allows the initialization of the process. Afterwards, this particular bee is replaced by another bee called Sref, which has the same role but which is chosen as the best solution found by the whole swarm at a given generation.

3.2 The Search Area

The search area consists of *k* bees well dispersed in the search space. In other words, k bees are at equal distances from each other. The idea behind the determination of scattered solutions is to better browse the search space in order to catch the most promising solutions. One possible code that determines such positions where bees are supposed to start the exploration of the regions is as follows:

```
Search area creation algorithm
Input:
     Sref: solution which is represented by a vector of n attributes;
Output:
     k dispersed solutions;
begin
     h = 0;
     SearchArea = empty;
     while (size of SearchArea <= MaxSize) and (h < flip)
     do
        begin
                    s = Sref;
                    p = 0;
                    repeat
                    change the element that is at position flip*p+h
             in s;
                    p = p +1;
```

```
                    until flip*p+h ≥ k;
                    SearchArea = SearchArea ∪ {s};
                    h = h+1;
        end;
end.
```

The parameter flip depends on the problem nature and thus is set by experiments.

Each bee explores its assigned region from a starting point of the search area. It proceeds to a local search in order to improve the current solution. The algorithm is as follows:

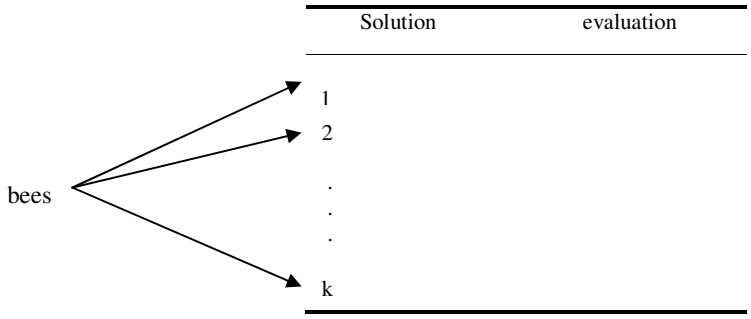
```
Region Exploration Algorithm
Input:
  -   A solution s encapsulated in a bee
Output:
  -   The best solution of the region

begin
     while there are still solutions to explore in the region
     do
        begin
            if the current solution is better than s
                    then s := current solution;
        end
 end.
```

The evaluation function is capable to appreciate the quality of any solution. It is defined differently from one problem to another one. It intervenes for instance in the region exploration algorithm where it allows the improvement of the quality of the current solution.

3.3 The Dance Structure

The Dance structure is a table with *k* entries and two columns. Each entry corresponds to a bee which exhibits its best solution in the corresponding first column once it fully explored its assigned region. The second column shows the evaluation of the solution that appears at the same line. The evaluation allows the selection of the best solution to be encapsulated in a bee to behave like Sref in order to iterate the search process. Table 1 illustrates the Dance structure.

Table 1. The Dance Table



3.4 BSO Algorithm

The algorithm BSO can be now written as follows:

```
begin
     let Sref be the solution found by BeeInit;
     while (MaxIter is not reached)
        do
          begin
            insert Sref in a taboo list;
            determine SearchArea from Sref;
            assign a solution of SearchArea to each bee;
                for each k bee
                      do
                       begin
                            improve the assigned solution;
                                    store the result in Dance;
                       end;
                  select the new reference solution Sref from Dance;
         end;
end.
```

The algorithm operates with empirical parameters such as *MaxIter*, which is the maximum number of iterations and *k* the number of bees. It uses a taboo list to insert the reference solution in order to avoid performing an iteration with the same solution.

4 BSO for Information Retrieval

In this section we show how to adapt BSO to information retrieval and then how to use the results to construct social networks. As we have already seen from the previous section, BSO is an approach of solving complex problem. In the information retrieval framework, the solutions are represented by the documents, that is, each document is viewed as a potential solution. The search space is defined by the whole collection of documents or the inverted file. The use of the inverted file is of course more advisable because, as we have seen in section 2 the number of documents to be consulted in the search process is much smaller than if we consider the whole set of documents. As for the scout bee BeeInit, we assign a document randomly drawn among the documents of the search space.

The algorithm BSO will be applied to the search process of information retrieval as follows:

A solution corresponds to a document while the evaluation function measures the similarity between the document and the query. First the details of the process of generating SearchArea that contains the *k* bees are as the following:

```
begin
     h = 0;
     while size of SearchArea not reached and h< Flip do
            s = Sref;
            p = 0;
            repeat
                   if the term Flip*p+h exists in s then remove it
```

```
                    from s else insert it in s;
                    p = p+1;
                until Flip*p+h ≥ k;
                SearchArea = SearchArea ∪ {s};
                (* set of all solutions s *)
                h = h+1;
        endwhile
end ;
```

Let $n=20$ be the number of terms and $Flip = 5$. If the terms are subscripted from 1 to 20, then the strategy consists in flipping the terms that occupy positions (1,6,11,16), (2,7,12,17), (3,8,13,18), (4,9,14,19) and (5,10,15,20).

5 BSO for Mining Scientific Social Nets

5.1 Textual Sources and Scientific Social Networks

The social network that we intended to build is a group of individuals that may be connected by links to a special interest. It is constructed from documents where the individuals are authors of some articles. The system we developed can be applied to newspapers, libraries and scientific documents such as conference proceedings and journals.

The benchmarks considered in this study are the collection CACM and the database DBLP. This contains publications including researchers in the field of computing.

- **The CACM collection**

CACM is a collection of titles and abstracts of the ACM journal and contains 3204 documents.

- **The DBLP database**

The Digital Bibliography and Library Project (DBLP) is a website publishing a catalog of bibliographies in computer science. Hosted by the University of Trier, Germany, it has existed since the 1980s. It was originally conceived as a catalog of bibliographies on databases and logic programming. In January 2010, DBLP lists more than 1.3 million articles in computer science.

- **Query and social networking**

The idea is to build a social network of authors engaged in the same field of research specified as a query.

5.2 Social Net Mining

The authors retrieved documents is the social network of scientists working on the same issues specified by the query. Fig. 1 illustrates the idea: in the center of the graph, we find the topic and around it all the authors who work on it.

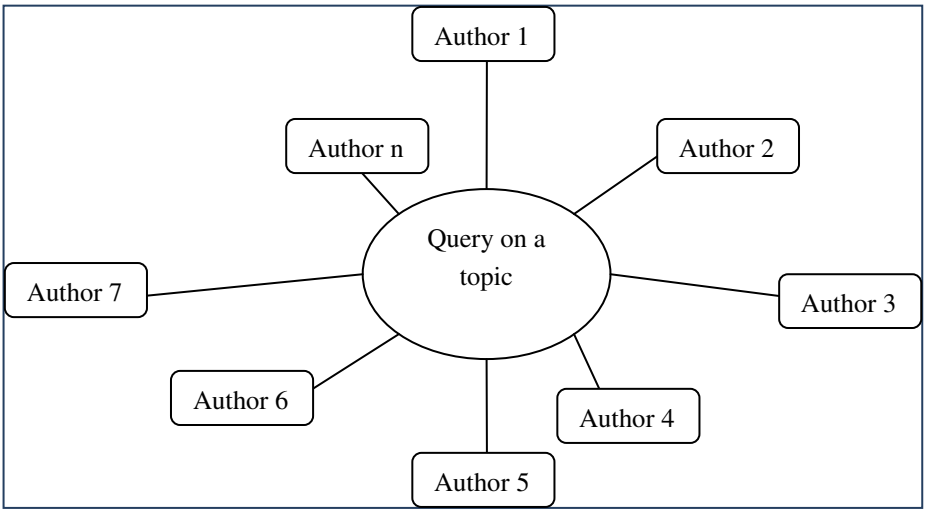


Fig. 1. A social net of authors working on a specified topic

6 Experimental Study

We separated the experimental treatment in three different tasks, which are:

- Indexing of documents and query
- Retrieving documents similar to the query
- constructing the social network

We tested the developed programs on CACM and DBLP. The latter indexes more than one million articles and contains more than 10,000 links to web pages of computer scientists. We have considered in this work one sub-base under DBLP containing 50,000 documents as shown in Table 2.

Table 2. Characteristics of the studied sub-collection of DBLP

Size of the sub-base	(documents)	**50 000**
	(Mb)	4.47
Number of terms of the dictionary		80000
Number of terms of the sub-base		600000

6.1 Extraction of Documents

We have implemented the retrieving methods based respectively on the inverted file and BSO. The results are shown in Table 3.

Table 3. Comparison between different techniques of documents extraction for CACM

dataset	query	Search with inverted file		BSO	
		Solution	Time (s)	Solution	Time (s)
CACM	'computer algorithm'	Doc#= 29 tf*idf =1.00	3	Doc#= 29 tf*idf =1.00	1
CACM	'parallel computing'	Doc#=1262 tf*idf=1.13	5	Doc#=1262 tf*idf=1.13	1
DBLP	'Web information retrieval'	Doc#= 29 tf*idf =1.00	7	Doc#= 29 tf*idf =1.00	2

For the query 'Computer Algorithm', the most relevant documents and their authors are:

2675 Ravindran, A.
2589 Proll, L. G.
2426 Lewart, C. R.
2004 Bouknight, W. J.
2209 Winters, W. K.
1066 Aronofsky, J. S.
2453 Ganapathy, S., Rajaraman, V.

The social network for this query is illustrated in Fig. 2.

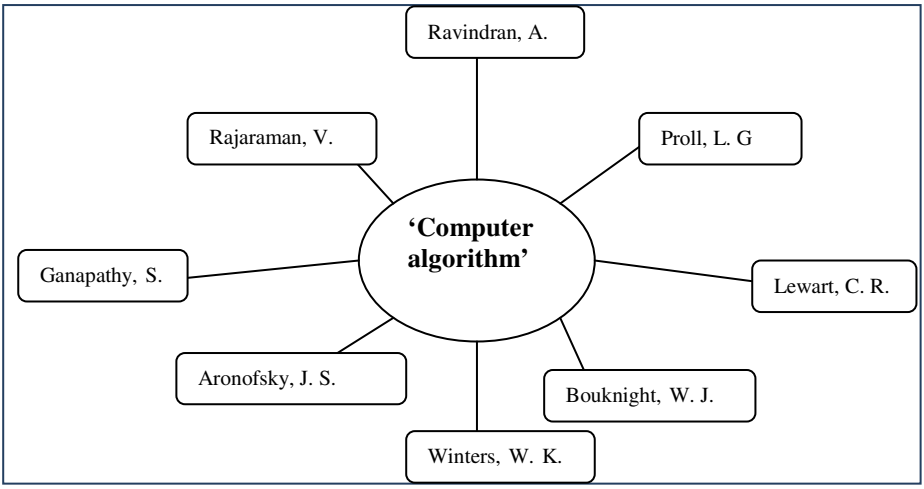


Fig. 2. Social Net of the authors working on '*computer algorithm'*

For the query **'parallel computing', t**he most relevant documents and their authors satisfying this query are:

1262 Burkhardt, W. H.
1525 Singleton, R. C.
1862 Finerman, A.
3156 Hirschberg, D.S.
Chandra A.K.
Sarwate, D.V.
2896 Gries, D.
2700 Lipton, R. J.

The social network for the second query is shown schematically in Fig. 3.

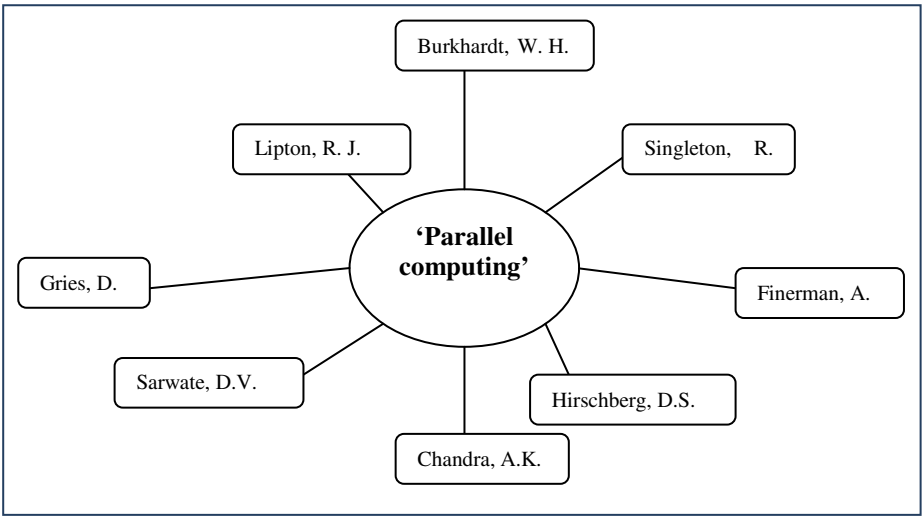


Fig. 3. Social Net of the authors working on '*parallel computing'*

For the query **'web information retrieval',** the authors responding to this topic are:

Taihei Oshino, Yasuhito Asano, Masatoshi Yoshikawa, Gabriella Pasi, Habiba Drias, Hadia Mosteghanemi, Namita Mittal, Richi Nayak, Mr. C. Govil and K. C. Jain.

The social network for Query 3 is illustrated in Fig. 4.

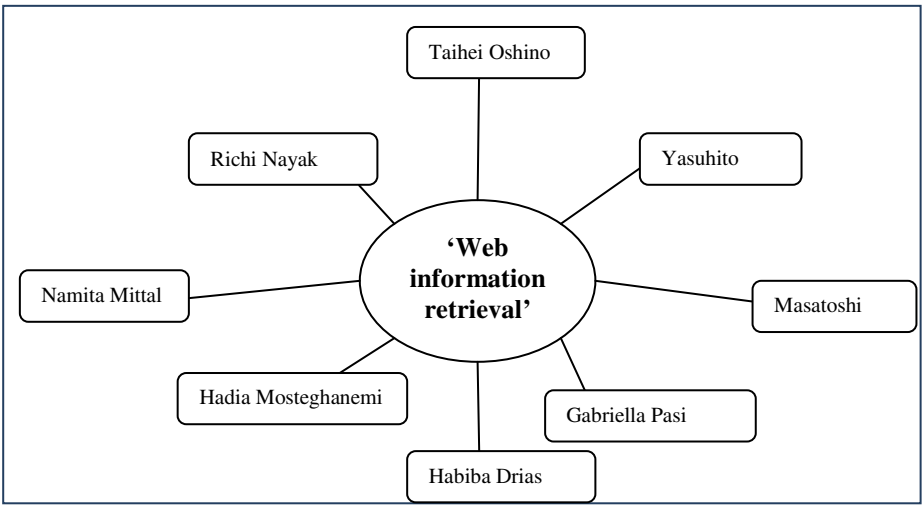


Fig. 4. Social Net of the authors working on '*web information retrieval'*

7 Conclusion

Social networking is a means of communication that is topical at the moment. In this work we defined a new approach based on information retrieval technologies for constructing social networks. In order to speed up the time of documents retrieval and hence to build the social community, we integrated a bio-inspired method namely BSO in the matching process. We also implemented the indexing step with an original way by using a lexical generator, which is Lex.

We developed and implemented the community discovery system and tested it on two datasets that are CACM and DBLP. The results are very encouraging.

For the near future, we plan to tackle the analysis of the obtained social communities and to study scoring function to evaluate how 'good' can be the social interaction between the members of the social networks.

References

1. Christopher, D.M., Prabhakar, R., Hinrich, S.: Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008)
2. Clauset, A., Newman, M., Moore, C.: Finding community structure in very large networks. Physical Review E 70, 066111 (2004)
3. Drias, H., Sadeg, S., Yahi, S.: Cooperative Bees Swarm for Solving the Maximum Weighted Satisfiability Problem. In: Cabestany, J., Prieto, A.G., Sandoval, F. (eds.) IWANN 2005. LNCS, vol. 3512, pp. 318–325. Springer, Heidelberg (2005)
4. Drias, H., Mosteghanemi, H.: Bees Swarm Optimization based Approach for Web Information Retrieval. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 6–13 (2010)

5. Flake, G., Lawrence, S., Giles, C.: Efficient identification of web communities. In: KDD 2000, pp. 150–160 (2000)
6. Fortunato, S.: Community detection in graphs. Physics Reports 486(3-5), 75–174 (2010)
7. Leskovec, J., Lang, K., Dasgupta, A., Mahoney, M.: Statistical properties of community structure in large social and information networks. In: WWW 2008: Proceedings of the 17th International Conference on World Wide Web, pp. 695–704 (2008)
8. Moldovan, D.I., Mihalcea, R.: Using WordNet and lexical operators to improve internet searches. IEEE Internet Computing 4(1), 34–43 (2000)
9. Newman, M., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E 69, 026113 (2004)
10. Domingos, P., Richardson, M.: Mining the network value of customers. In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, pp. 57–66. ACM Press (2001)
11. Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D.: Defining and identifying communities in networks. Proceedings of the National Academy of Sciences of the United States of America 101(9), 2658–2663 (2004)
12. Van Rijsbergen, C.J.: Information Retrieval, 2nd edn. Butterworth, London (1979)
13. Salton, G.: The SMART retrieval system: experiments in automatic document processing. Prentice-Hall, Englewood Cliffs (1976)

Information and Information Systems Project for a Strategic Digital City: A Brazilian Case

Denis Alcides Rezende[1], Frederico de Carvalho Figueiredo[1], Leana Carolina Ferreira Setim[1], Luciane Maria Gonçalves Franco[1], and Gilberto dos Santos Madeira[2]

[1] Programa de Pós-Graduação em Gestão Urbana,
Pontifícia Universidade Católica do Paraná, PUCPR, Brasil
{denis.rezende,frederico.figueiredo,leana.carolina}@pucpr.br,
lucianemgf@yahoo.com.br
www.pucpr.br/ppgtu

[2] Programa de Pós-Graduação em Engenharia Elétrica,
Universidade de Campinas, UNICAMP, Brasil
gil.madeira@gmail.com

Abstract. Making information available is a permanent challenge for cities concerned to their social, urban and local planning and development, focused on life quality of their citizens and on the effectiveness of public management. The objective is to describe the information and information systems project from the planning of a digital city carried out in Vinhedo-SP, Brazil. The research methodology was emphasized by a case study which turned to be a research-action, comprising the municipal administration and its local units. The results achieved describe phases, sub-phases, approval points and resulting products. The contributions have to do with the practical feasibility of the project. The conclusion reiterates the importance of the project, collectively implemented and accepted, as a tool to help the management of cities, in the implementation of strategic digital city projects, in the decisions of public administration managers, and in the quality of life of their citizens.

Keywords: Information, Information Systems, Municipal Information Planning, Strategic Digital City, Municipal Management.

1 Introduction

Making information and information systems available is a permanent challenge for cities concerned to their social, urban and local planning and development, when focused on life quality of their citizens and on the effectiveness of public management. Such a challenge requires the involvement of everyone in the city namely municipal servants and citizens, no matter they are workers, students, retired people, housewives, city councilors, entrepreneurs or others. Also, the endless social, financial and political challenges faced by the cities have demanded from their managers a more competent performance to propitiate suitable life quality to their citizens. On the other hand, the participation of citizens in the conduction of a city is

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 345–355.
DOI: 10.1007/978-3-642-36981-0_33

an inexorable need. One way to contribute, when facing these challenges, is to have an information strategic planning achieved collectively, where systematized pieces of information are pre-requisites for the organization and management activities of the city halls and towns.

The planning of municipal information and information systems, like the municipal strategic planning, is one of the tools for managing city halls and towns of unquestionable relevance. It is like that, mainly because of insufficient financial resources to the cities, obedience to the new public laws, restrictions of the city laws, and pressures exerted by citizens and other people interested in the city (social players or stakeholders). These pressures can be minimized by the preparation and implementation of participative strategic planning in the cities, as it can propitiate the collective engagement of the citizens with their wishes and also can decentralize and share the local managers' decisions.

Another problem comes out: the inexistence or inadequacy of municipal information and information systems resources of the external and internal environment of some administrations, thus corroborating the difficulties to manage cities and count on the respective participation of citizens in the city management and also in the urban and regional development.

The objective of this article is to describe the information and information systems project from the planning of a digital city carried out in Vinhedo-SP, Brazil. It is part of a broad project called Strategic Planning of Information Technology for Strategic Digital City. It has embraced the City Hall, its Municipal Secretaries, a Local Authority and other municipal institutions such as schools, health centers, and social squares for citizen attendance.

The reason for that is mainly related to structuring, storage and availability of all information and the sharing of knowledge by means of information technology resources, which surely do not happen overnight and require short, medium and long term action in the cities.

The information an information systems planning along with other municipal plans (e.g. Strategic Municipal Planning, Municipal Director Plan, Municipal Multi-Annual Plan, Governmental Plan and others) requires exhaustive practical exercises based on deep-seated theoretical fundaments. These exercises are related to dynamic intelligent daily activities of the city halls and to the management actions of the cities. Such activities have to be prepared in an integrated and structured way, where personalized and timely information are essential factors for suitable and participative management of the cities. Evidently, these activities will be more profitable to the cities if they have effective participation of their citizens.

2 Theoretical Fundaments

2.1 Information and Information Model

For initial conceptualization, a piece of information is any datum considered or worked out. It can be understood as a datum with a meaningful value attributed or added to it and having a natural and logical sense for those who use that piece of

information [1, 2, 3]. It can be defined as something useful. Information to be useful for decision-making should have the following characteristics or premises: not to be abstract, not to have verbs, to be different from documents, programs, files and things alike [4].

Information and its respective systems perform fundamental functions in the city halls and towns, presenting themselves as a strategic resource to plan and manage city halls and towns in an intelligent, competitive, participative manner. To ensure this strategic connotation to information the organization usually goes through an evolving cycle in stages, such as: initiation, contact, control, integration, data administration, and maturity [5].

The information acculturation process in the city halls and towns can be ease and effective if public managers and residents (clients or information users) are participative, aware and thoroughly engaged in the utilization of information technological resources. The strategic information supported by information technology may occur in an evolving way, where information systems can be classified in various manners [6]. The information valorization process comprises some logical phases and steps: getting to know, selecting and using information [7].

The information model describes all the necessary pieces of information for management of city halls or towns [4]. Such pieces of information can be structured in levels or types, i.e., strategic, managerial and operational information. They can be distributed according to their respective organizational functions or public issues. In the information model document it is described the strategic pieces of information (in a macro-related way with the external and internal environment), the tactic and managerial pieces of information (grouped and synthesized) and the operational pieces of information (in detail or analytical). Also, the information model may contain integrated information of the kinds: conventional (trivial), personalized and timely.

All and any peculiar or specific information can be named personalized information, either being a natural or legal person or a differentiated product or service. It may also be regarding a single characteristic of a citizen. All and any information of unquestionable quality, however, anticipated can be named timely information [6].

2.2 Information System and Information Technology

A system of information proposes as an output, the result of data processing or of resources to be transformed, for instance, information, being computational resources information or not [8]. The information systems – which have as a major objective the aid to city halls in their decision-making processes and whose focus must be addressed to municipal services – will be able to significantly contribute to solve many municipal problems; in the extent they effectively generate timely and personalized information. They have to be seen as tools to determine differential factors of the city hall [4] and are defined as systems of human and technical components that accept, store, process, produce and transmit information, and can be based on any combination of human engagement, methods and information

technology [9]. They are the ways by which people and organizations put together, process, store, use and disseminate information [10].

An information system can have operational effectiveness, large impact on municipal strategies and on the success of city halls, benefiting them, the citizens or any private individual or group that interact with municipal systems [3]. The information system benefits are widely reported and justified in the pertinent literature and articles. They should present themselves as instruments for generating solution factors of the municipal problems [11, 2]. To facilitate these solutions, the information systems should be integrated and have an interdependent relation with the involved sub-systems, which chiefly leads to information exchange amongst them. Such integration can also be named systemic vision, systemic approach, forest view, municipal gear, information synergy and harmonic set of organizational functions. This way, the information systems can show themselves as information sub-systems or organizational functions, divided into: production or services; commercial or marketing; material or logistics, financial, juridical-legal, human resources and their respective modules [6]. Only through the integration of these sub-systems it will be possible to produce timely and personalized information.

The levels of information and decisions in the city halls can obey the pattern-hierarchy existing in most organizations (organizational pyramid). These levels are known as strategic, tactic and operational. The kind of decision made in each level requires different degree of information and data aggregation. The various decision levels call upon different information in their several types of resulting products, such as screens, reports, etc. Practically, there is not any rigid classification, what allows people and administrations to classify their systems in different ways. All in all, the IS can be ordered in operational, managerial and strategic [3, 2]. The main difference amongst information systems in such a classification lies on the way the information is presented to users.

The operational information systems provide detailed information. The managerial information systems make available grouped or synthesized information. The strategic information systems enable the comparison of information both inside and outside the city hall environments, also between other cities, and eventually generate some indicators and knowledge. Another way to classify information systems is as for the city hall scope. So considered, there are systems at personal, group (or departmental), organizational, and inter-organizational levels in which the global systems of information are framed, involving several city halls of various towns [1].

The information technology (IT) or information or communication technology (ICT) can be conceptualized as computational and technological resources for data storage, generation and use of information. It is based on the following components: hardware and its devices and peripherals; software and its resources; telecommunications systems; data and information management [3, 12, 6]. It can also be defined as telecommunications and computing technologies that supply automatic means and information handling, encompassing tangible hardware and intangible software [9] and telecommunications networks [10].

2.3 Information Planning and Information Technology

The Municipal Information Planning is a project that formalizes the necessary information for city hall and town management and, as a result, it plans the municipal information systems and their respective needs of information technology assets and of human resources [4].

The Information Technology Strategic Planning (ITSP) is a dynamic and interactive process to, strategically, tactically and operationally, structure the organizational information, the information systems, the information technology (and their resources: hardware, software, telecommunications systems and data and information management), the people involved, and the required infra-structure for assisting all decisions, actions and respective processes of the organization [12, 13, 4].

The alignment between the ITSP and the municipal plans and planning can be built from the satisfactory vertical, horizontal, transversal, dynamic and synergetic relations of the municipal functions and of the IT resources. Such alignment can promote the adjustment or strategic and operational suitability of available technologies in the entire city as a municipal management tool put together by quality concepts, productivity, effectiveness, continuity, competitive intelligence and municipal intelligence [14, 4].

2.4 Digital City

The concept of digital cities regards the gathering and organization of digital information to provide some information room such that their residents and visitors can interact among themselves [15]. The cities concern about physical changes to enable technical conditions and to ease the functioning of digital cities with their technological resources and respective conceptual, cultural and urban impacts [16, 17].

Strategic digital city can be understood as an application of information technology resources in the management of the municipality and also in the availability of information and services to residents. It is a far-reaching project if compared to just offering internet to citizens through telecommunications conventional resources. It goes beyond digital inclusion of citizens in the world network of computers, once such ample projects encompass information systems for municipal management and for services to citizens; and also municipal safety systems [4].

3 Research Methodology

The research methodology employed here comprises an approach of applied work nature to a circumstantial reality, emphasizing the inductive method enhanced by professional experience achieved by the authors-researchers, both in advising projects for city halls and in academic projects for degree, master and doctoral classes. It partially takes exploratory research concepts regarding documental and bibliographic raising [18].

The case study, which emphasizes an action-research, had an applied work approach as well. It was applied work because it generated new useful knowledge for science breakthrough and for practical checkup addressed to solution or facilitation of problems concerned to the planning of municipal information with the participation of the society (municipal servers and citizens) and of the management of a city. It has put together other approaches explained by the eclecticism and integration of inseparable methods [18, 19].

The project was accomplished from August 2009 to July 2010 by using the Methodology of Project MIP/IT already carried out by other planning projects of municipalities. The sample of the research-action was applied in the city of Vinhedo, São Paulo, Brazil. Beside Campinas, Vinhedo has 62,240 inhabitants.

The unit of observation involved municipal servants and citizens (workers, students, retired people, housewives, city councilors, and entrepreneurs). The protocol of the research-action is an integrative part of the methodology employed in the project whose variables are its phases and sub-phases. The project has been being implemented since 2011.

4 Planning Project of Municipal Information

It is part of an ample project called Strategic Planning of Information Technology for Strategic Digital City (MIP/IT). It encompassed the City Hall, its Municipal Secretaries, a Local Authority, and other municipal units like schools, health centers, and social squares for citizen assistance.

The realization of the project demanded the definition of a Methodology of Project MIP/IT which formalized parts, phases, sub-phases, products and approval point.

4.1 Phase 0 and Revise the Municipal Strategic Plan

The first activity required to build the MIP/IT Project was its preliminary organization (Phase 0 - Organize, disseminate and train) which consisted of tasks accomplished before and during the effective development of the project.

Vinhedo doesn't have a Municipal Strategic Planning, what hampered the making of the required sub-phases. Municipal analyses, problems and challenges, municipal guidelines, main objectives, strategies, principal actions of Vinhedo are just partially described in a document Public Administration Plan - 2009/2012.

4.2 Municipal Activities and Municipal Information Model

It was made 1,331 Lists of Municipal Activities, containing the main municipal activities elaborated by the 15 Municipal Secretaries and a Local Water Authority. The municipal activities can also be figured out as processes of municipal tasks existing in each Secretary, including all the other municipal units.

It was built the Municipal Information Models containing 18,200 pieces of operational and managerial information throughout the city, City Hall and respective

units. By option of the Management Project Committee, it was formalized all the pieces of operational information (in detail) and all the pieces of managerial information (grouped) and it was not formalized the strategic information (macro-related to the city and City Hall internal and external environment). The modeled information did not excluded redundant pieces of information or the ones with the same content.

4.3 Evaluate and Plan the Information Systems

The evaluation and summarization of the current situation show 407 systems distributed as follow: 20 systems of a service provider company; 18 of various suppliers; 70 of Governments and Public Organizations; 214 of Microsoft™; and 85 of manual systems.

The proposed systems took in consideration the Public Administration Plan 2009/2012 with 4 preferential thematic areas: social; economic; urban; and governmental development.

The Municipal Information Models were revised and altered by the Project Management Committee team and completed with pertinent documents. Considering the Municipal Thematic, the Municipal Activities and the Information Models elaborated, it was named or formalized the Municipal Information Systems, also taking into accounts the debates held in the Vinhedo Society. It was named, diagramed, described and planned 140 proposed municipal information systems, involving the City Hall, its Municipal Secretaries, a Local Authority and other municipal entities. All the referred proposed systems were validated with the servants of their respective Municipal Secretaries. For that, a demonstrative framework of the proposed municipal information systems was elaborated, containing the names of the Municipal Secretaries and the quantity of systems: Administration-23; Agriculture-8; Science and Tecnology-2; Commerce-3; Culture-6; Education-14; Sports-2; Finance-10; Government-5; Housing-5; Industry-1; Juridical-2; Leisure-2; Environment-4; Building-2; Planning-5; Health-15; Safety-4, Municipal Services-8; Social-8; Traffic-1; Transports-5; Tourism-4; and Sanitation-1.

As for the evaluation of acquisition or development of the proposed municipal information system all of them will be bidden and supplied by companies specialized in projects, development and maintenance of systems. No information system will be developed by insiders. At the same time, it was sorted out and assessed 26 supplying firms of municipal information systems which could provide the modeled information and respective proposed systems.

To finalize this planning phase of municipal information systems, a priority was defined by the Project Management Committee: the systems were addressed to public functions and public or municipal thematic such as social; public finance; and health. However, as there is a principle of single database, it was not allowed the recording data redundancy of the city, City Hall, citizens and other standard data.

4.4 Prioritize and Pay for the Execution of the Project MIP/IT

The first criterion of prioritization aimed the document Government Plan 2009/2012. From the criteria of priorities defined by the city, the following other priorities were established by the Project Committee Manager: implementation of a Single Data Base (municipal information systems); information systems of promotion and municipal social assistance; systems of public finance; and information systems of municipal health care.

The necessary resources for the execution of the project are being elaborated in pertinent documents and in accordance with the current Multi-annual Plan. Also, it is being elaborated projects for evaluating and raising resources in the Federal Program PMAT. It will be worked out possibilities of public-private partnerships.3

The proposed municipal information systems and their respective information technology resources should reach the objectives, the strategies and the actions of the city. They should be formalized by means of the Vinhedo's Strategic Planning Project to be worked out.

5 Results Achieved

The results achieved describe, by means of a project methodology, the phases, sub-phases, approval points and resulting products, and formalize their respective challenges and difficulties.

Regarding the subject "information", the main results achieved are described in subchapter *4.2 Municipal Activities and Municipal Information Model*. The highlight was the modeling of 18,200 information (at the operational, managerial and strategic levels) prepared in a collective activity by participatory citizens of Vinhedo.

Regarding the subject "information systems", the main results achieved are described in subchapter *4.3 Evaluate and Plan the Information Systems*. The highlight was the 140 proposed municipal information systems, involving the City Hall, its Municipal Secretaries, a Local Authority and other municipal entities.

The "information" and "information systems" covered 24 public or thematic municipal functions: administration; agriculture; science; technology and innovation; commerce; culture; education; sports; finance; government; housing; industry; juridical; leisure; environment; building; planning; health; sanitation, safety; municipal services; social; traffic; transport; and tourism.

6 Conclusion

All cities' plans require the participation of their citizens and the engagement of their local managers to expand the social, urban and regional development. For that, it is necessary to plan and make available information and information systems in a suitable and transparent way.

From the point of view of the municipal manager, the planned information and information systems can contribute to the management of the cities. On the other

hand, from the point of view of the citizens' participation, this consideration also may propitiate the exercise of citizenship in the extent of their participation in the planning, structuring, preservation and availability of information for their use and benefit regarding the possibility of enhancing life quality. Both the approaches allow that cities to enlarge their activities of communication, education and social development, mainly because countless cities face constant social, financial and political challenges to provide an effective management to the cities and the required quality of life suitable to their citizens.

Undoubtedly, the municipal information and information systems planning, like the municipal strategic planning, can become an inexorable participative instrument of management of city halls and cities. With the right modeling, planning and availability of information and municipal services, it makes possible the breakthroughs in the citizenship building in the public space, in the municipal democratic governability, in the transparency of management and in the urban and regional development of the cities.

The objective of this article was reached, as it treated about the municipal information planning and about information systems from a digital city project carried out in Vinhedo. The referred project involved the City Hall, its Municipal Secretaries, one Local Authority and other municipal units, such as school, health centers, and social squares for attendance to the citizens. The bottom line of this planning was the participation of municipal servants and citizens, namely, workers, students, retired people, housewives, city councilors, entrepreneurs, among others.

The first impact was the elaboration of 1,331 Lists of Municipal Activities containing the main activities or processes of elaborated municipal tasks. It was a challenge because the City Hall had not documented its activities in a structure organized as the science of administration asserts. However, further, the people involved were prepared and motivated to participate of the Project MIP/IT of Vinhedo and to elaborate, mainly, in a collective way, the models of information, the evaluation and the planning of all municipal information systems.

The social players involved propitiated the integration of technological resources with the social participation in a democratic popular experience, and together, local managers and citizens, by means of an organized project, could model 18,200 pieces of managerial and operational information for all the city, City Hall and respective units. Such information formalizes their wishes, interests and, mainly, their personal needs, let us say, it was addressed to the city reality. Beyond these pieces of information, they could plan and prioritize 140 municipal information systems proposed.

Regarding free of charge internet, the project already reaches over 150 residences and over 400 citizens. For these citizens, the following main municipal services are made available: opening, alteration and closing of firms; electronic invoicing; citizen's portal with personal information, property information, etc.; heath card; daycare and school; servant's portal with personal information, about payments, vacations, etc.; transparency portal, municipal processes follow-up. And for the public managers (Mayor, Secretaries, Directors, and servants) countless pieces of information to help them make decisions are made available.

Evidently there were many difficulties throughout the elaboration of the phases, sub-phases and tasks to plan in a participative way the information and information systems of the city. However, such difficulties were day by day and collectively overcome, either by the local administration or by the citizens and organized civil society that took part in the realization of the Project MIP/IT of Vinhedo.

Regarding the contributions of this work, they report the practical viability of the project developed and of the execution details of its methodology. Such contributions are addressed to the academy, to the participant city of the case study, and to other cities. To the academy, it is highlighted the project elaboration methodology in a participative way, putting together interests of local managers and of citizens.

The case study described can facilitate other researches applied to cities to understand and suggest alternatives for their challenges coming from urban and regional planning and development, as well as, from public management that involves citizens, municipal servants, local managers and other players interested in the city researched and other surrounding cities. To the city under consideration this work contributed to the elaboration and further implementation of a vanguard project which will display information for decision-making, either to its local managers or to its citizens, allowing broader understanding on equals or similar projects. This work also contributes to the establishment of analogies among the cities.

There weren't restrictions as for the access of people to the project, nor to information or information systems of the city and of the City Hall. The main restriction to this work had to do with the research methodology itself as it reduces pertinent analyses and considerations to a single city, so not expressing the reality of Brazilian cities.

The conclusion reiterates the importance of a project methodology, collectively implemented by municipal servants, local managers, and citizens; and accepts it as an instrument to, effectively; contribute in the management of cities and in the implementation of strategic digital city projects. With the effective implementation of this kind of project; the democratic public space; transparency; effectiveness and governability of municipal services; the social, urban and regional development can be widened and, as a result, the quality of citizens' life can also be better.

References

1. Kroenke, D.: Management Information Systems. McGraw-Hill, New York (1992)
2. Laudon, K.C., Laudon, J.P.: Management Information Systems: A Contemporary Perspective. MacMillan, New York (1996)
3. Stair, R.M.: Principles of Information Systems: A Managerial Approach, 2nd edn. Thomson, Publishing, Washington (1996)
4. Rezende, D.A.: Planejamento de Estratégias e Informações Municipais para Cidade Digital: Guia para Projetos em Prefeituras e Organizações Públicas (2012)
5. Nolan, R.L.: Note on Information Technology and Strategy. Harvard Business School, Boston (1993)
6. Rezende, D.A., Abreu, A.F.: Tecnologia da Informação aplicada a Sistemas de Informação Empresariais, 8th edn. Atlas, São Paulo (2011)

7. Weitzen, H.S.: O Poder da Informação. Makron Books, São Paulo (1991)
8. O'Brien, J.A.: Management Information Systems, 5th edn. McGraw-Hill, Irwin (2002)
9. Heeks, R.: Reinventing Government in the Information Age: International Practice in IT-Enabled Public Sector Reform. Routledge, London (1999)
10. Ward, J., Peppard, J.: Strategic Planning for Information Systems. John Wiley & Sons, New York (2002)
11. Freitas, H.M.R., Becker, J.L., Kladis, C.M., Hoppen, N.: Informação e Decisão: Sistemas de Apoio e seu Impacto. Ortiz, Porto Alegre (1997)
12. Turban, E., Mclean, E., Wetherbe, J.: Information Technology for Management: Improving Quality and Productivity. John Wiley and Sons, New York (1996)
13. Lederer, A.L., Mahaney, R.C.: Using Case Tools in Strategic Information System Planning. Information Systems Management Journal, USA 13(4), 47–52 (1996)
14. Henderson, J.C., Venkatraman, N.: Strategic Alignment: Leveraging Information Technology for Transforming Organizations. IBM Systems Journal 32(1), 4–16 (1993)
15. Besselaar, P.V.D., Beckers, D.: Demographics and Sociographics of the Digital City. Disponível, http://portal.acm.org/citation.cfm?id=701393 (acesso em: February 02, 2009)
16. Atkinson, R.: The Digital Technology Revolution and the Future of US Cities. Journal of Urban Technology 4(1), 81–98 (1997)
17. Shiode, N.: Urban Planning, Information Technology, and Cyberspace. Journal of Urban Technology 7(2), 105–126 (2000)
18. Nachmias, D., Nachmias, C.: Research Methods in the Social Sciences, 3rd edn. St. Martin's Press, New York (1987)
19. Yin, R.K.: Case Study Research: Design and Methods. Sage Publications Inc., London (1994)

Linking Benefits to Balanced Scorecard Strategy Map

Jorge Gomes[1], Mário Romão[2], and Mário Caldeira[1]

[1] ISEG-UTL, Rua Miguel Lúpi 20 1249 Lisboa, Portugal
jorge.gomes@phd.iseg.utl.pt, caldeira@iseg.utl.pt
[2] ISCTE-IUL, Av.ª das Forças Armadas 1649-026 Lisboa, Portugal
mario.romao@iscte.pt

Abstract. In this paper the authors propose a link between the Benefits Dependency Network, from a Benefits Management approach, and a Strategy Map, from Balanced Scorecard, to improve the management of business benefits and to ensure that actions taken along the investment life-cycle lead to foreseen benefits realization. The goal of this integration is to build a framework that combines useful features of both methods. We sustain that they can be complementary. As a Strategy Map is committed with strategic alignment, communication and monitoring of strategy execution at all levels of an organization, a Benefits Dependency Network is aimed at explaining how benefits are going to be obtained through organizational change. Using the results of a case-study research, we explain how a Strategy Map can cross with the Benefits Dependency Network. The integrated contribution is meant to increase the investments effectiveness, giving to stakeholders the confidence on a clearer delivery path for their expected benefits.

Keywords: Benefits Management, Balanced Scorecard, Strategy Map, IS/IT Investments.

1 Introduction

Organizations are being put under increasing pressure to justify the large amount of financial resources spent on IS/IT assets. As competition increases as a result of globalization and other market factors, it is even more important that an organization performs at its best capabilities (Ashurst & Doherty, 2003)[1]. The decision making process over IS/IT investments is not as objective and transparent as it is claimed to be, creating significant failures on the benefits achievement process (Berghout *et al.,* 2005)[2] It is unlikely that benefits will simply emerge, as if by magic, from the introduction of a new technology. Their realization needs to be carefully planned and managed (Lin & Pervan, 2003)[3], (Markus, 2004)[4]. Benefits are often identified in the early stages to form the business case and to sell the idea to the stakeholder(s). A follow-up procedure with the purpose of evaluating those benefits achievement is often missing, and problems arise after the system delivery, when it's time to show if those previous stated benefits have been realized (Remenyi *et al.*, 2007)[5]. A common characteristic of many unsuccessful programs is the vagueness with which

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 357–369.
DOI: 10.1007/978-3-642-36981-0_34

the expected benefits are defined (Reiss *et al.,* 2006)[6]. Without clearly defined objectives it is difficult to maintain focus when subsequent problems occur.

Bennington and Baccarini (2004)[7] argue that most organizations do not monitor the benefits for the following reasons, namely, lack of experience and/or business awareness, focus on managing deliverables rather than the benefits, lack of focus on the people who will enjoy benefits, emotional commitment to the continuity of the project and, so, not open to change to benefits that threaten project viability and lack of tools to help ensure that benefits will be deliverable. The increased interest in benefits realization has coincided with the increasing use and complexity of IS/IT (Ashurst & Doherty, 2003){1}, (Ward & Elvin, 1999)[8], (Bradley, 2006) [9]. The benefits to an organization from IT-enabled change essentially emerge from three reasons: either stopping doing activities, doing better what have always being done, or even doing completely new things (Peppard & Ward, 2005)[10].

2 Balanced Scorecard and Strategy Maps

The Balanced Scorecard (BSC) is one of the most highly touted management tools today (Atkinson & Epstein, 2000)[11], (Frigo & Krumwiede, 2000)[12]. The BSC not only translates the strategy to operational terms, but it aligns the organization to their strategy by focusing business units and employees on their role in accomplishing the company mission (Frigo & Krumwiede, 2000)[12]. First proposed in 1992, it provides the executives with a comprehensive framework that translates a company's strategic objective into a coherent set of performance measures. Much more than a measurement exercise, the BSC is a management system that can motivate breakthrough improvements in such critical areas as product, process, customer, and market development (Kaplan & Norton, 1993)[13]. A major strength of the BSC approach is the emphasis it places linking performance measures with business unit strategy (Otley, 1999)[14]. The BSC supplemented the traditional financial measures with three additional perspectives, customers, internal processes and learning and growth, and enabled companies to track financial results while simultaneously monitoring progress in building the capabilities and acquiring the intangible assets they would need for future growth.

By using the BSC organizations do not have to rely only on short-term financial measures as the sole performance indicators. The first step in creating management processes for implementation of strategy must be to construct a reliable and consistent framework for describing strategy, known as Strategy Map (SM). The SM outlines all the cause-and-effect linkages between what an organization's strategy is and what everyone does on a day-to-day basis (Kaplan & Norton, 2000)[15]. The SM identify how to get the destinations, the strategic objectives, how tangible and intangible assets are involved, outlines how assets of all types are combined to create customer value propositions, specifies how the desired financial outcomes will be realized and detail the relationships between shareholders, customers, business processes and competencies (Kaplan & Norton, 2000)[15].

3 Benefits Management Approaches and Benefits Dependency Networks

The Benefits Management allows the investor to have the right information for the investments viability studies, pointing out the way to deliver the expected benefits.

Benefits Management (BM) can be described as: *"The process of organizing and managing such that potential benefits arising from the use of IT are actually realized'* (Ward & Daniel, 2006: 384)[16].

The five principles for realizing benefits through IT are: (1) IT has no inherent value - Just having technology does not confer any benefit or create value; (2) Benefits arise when IT enables people to do things differently - Benefits emerge only when individuals or groups within an organization, or its customers or suppliers, perform their roles in more efficient or effective ways; (3) Only business managers and users can release business benefits - benefits result from changes and innovations in ways of working, so only business managers, users, and possibly customers and suppliers, can make these changes; (4) All IT projects have outcomes, but not all outcomes are benefits - Many IT projects produce negative outcomes, sometimes even affecting the very survival of the organization; (5) Benefits must be actively managed to be obtained - Benefits are not outcomes that automatically occur (Peppard *et al.*, 2007)[17].

The key tool of this approach is the Benefits Dependency Network (BDN) that was introduced for the first time by Ward and Elvin (1999) [8], designed to enable the investment objectives and their resulting benefits to be linked in a structured way to the business, organization and IS/IT changes required to realize those benefits.

The following answers are used to develop both a robust business case for the investment and a viable change management plan to deliver the benefits: Why must we improve? What improvements are necessary or possible? What benefits will be realized by each stakeholder if the investment objectives are achieved? How will each benefit be measured? Who owns each benefit and will be accountable for its delivery? What changes are needed to achieve each benefit? Who will be responsible for ensuring that each change is successfully made? How and when can the identified changes be made?

To agree on the objectives and benefits is advisable to organise workshops with all the relevant stakeholders, managers, IT staff and sales personal to meet the alignment needed to develop the BDN. Once the BDN has been constructed, measures for each benefit, responsibilities for achieving all the benefits and implementing all the changes, and timeframes must be established.

The case study used for this paper was identified has a problem-based intervention where the final goal is the target improvements. In the *ends*-driven implementations, the improvement targets form the basis of the business case and financial calculations.

The organization have setting targets, identifying and quantifying the benefits removing known problems through new IT *means* and new *ways* of executing business processes and activities. The main challenge is agreeing on the best combination of *ways* and *means* for accomplishing the improvements.

4 BSC and Benefits Management Limitations

The crucial point of the BSC is to link the measures of the four areas in a casual chain which passes through all four perspectives. Thus, Kaplan and Norton (1996b)[18], emphasize that non-financial strategic objectives should not consist of an arbitrary collection of measures. Instead, they should involve a balanced representation of financial and non-financial measures. Kaplan and Norton (1996a)[19] assume the casual relationship between perspectives.

In short, the measures of the precedent perspective are the drivers for the measures of the following perspective. The assumption that there is a cause-effect relationship is essential because it allows the measurements in non-financial areas to predicted futures financial performance. Nørreklit (2000)[20] argues that there is no cause-and-effect relationship between some of the suggested of measurements in the BSC. Although there is a considerable co-variation between customer loyalty and financial performance, for example, it is not clear that increased customer loyalty is the cause of a long-term financial performance. The lack of a cause-effect relationship is crucial because invalid assumptions in a feed-forward control system will cause individual organizations to anticipate performance indicators which are actually faulty, resulting in dysfunctional organizational behaviour and sub-optimized performance (de Haas & Kleingeld, 1999)[21].

The influence between measures is not unidirectional in the sense that learning and growth are the drivers of internal business processes, which are the drivers of customer satisfaction, which in turn is the driver of financial results. The development process, for example, depends on the financial results; the possibility of providing the capital needed for the necessary investments being limited by unsatisfactory financial results. In order to be able to invest in research and development, organizations need satisfactory financial results, but they likewise need research and development in order to be able to produce satisfactory financial results. The reasoning is circular. So, instead of a cause-effect relationship, the relationship between the areas is more likely to be one of interdependence (Nørreklit, 2000)[20].

Otley (1999)[14] refers to another source of ambiguity. A linear chain suggested that better trained employees will lead to better businesses processes being designed, and these in turn will lead to more satisfied customers and then to happier shareholders. Although this is a plausible chain of events, it is very much a simplification of reality.

BSC aims to solve the problems related to strategy implementation (Kaplan & Norton, 1996a)[19]. However, the control model is a hierarchical top-down model not rooted in the environment or in the organization, which makes it questionable as a strategic management tool (Nørreklit, 2000)[20]. Consequently, a gap must be expected between the strategy expressed in the actions actually undertaken and the strategy planned.

Bontis *et al.,* (1999)[21] believe that the BSC has weaknesses that constrain its potential. To this author the BSC is relatively rigid. This rigidity appears in many aspects of the BSC. First, the perspectives drive the identification of Key Success Factors (KSF). This is limiting, because some KSFs cross perspectives and impact

simultaneously more than one dimension of the intangible resources of the company. Of course, this is not a big problem, if managers can identify all KSFs regardless. The danger, however, is that their attention will be concentrated only on the perspectives, and they will miss some important KSFs just because they do not fall neatly into any of the categories. Another situation identified by Bontis *et al.,* (1999)[21] concerns to personnel mixed together with IT systems into the learning and growth perspective.

Innovation is actually part of the internal business process focus. It feels almost as if innovation is considered a routine, something the organization can do without the people, or at least independently of them. The reason why the scorecard redefines vision and strategy as measures from the four perspectives is an implicit wish to balance company's activities within its stakeholders (Atkinson *et al.,* 1997)[22]. However, not all stakeholders have been included, some of the excluded ones being suppliers and public authorities, which may be important to several firms. Similarly, institutional stakeholders have been left out, as has the importance of business networks (Nørreklit, 2000){20}. Schiemann and Lingle (1999)[23] underlines the increasing partnerships options on the businesses, suggesting new perspectives concerning the partners and suppliers. Kaplan and Norton (1996a)[19] emphasize that it is not decisive whether all stakeholders have been included. It must be pointed out that in the introduction to their 1996 book; Kaplan and Norton explicitly state that the BSC is not a stakeholder approach.

Anand *et al.,* (2005)[24] extends the criticism to BSC, point out that it is difficult to achieve a balance between financial and non-financial measures, due to implementation problems. Mayer (2002)[25] reinforces that the BSC approach doesn't provide guidance on how to combine the dissimilar measures into an overall appraisal of performance.

Daves (2007){26} highlights the danger of establishing "stretch targets" without understanding that to achieve them one needs to first consider the competency levels to implement them.

In addition to such criticisms, a few authors have questioned the novelty of the BSC. For example, the French Tableau de Board it's a similar system used for decades (Lebas, 1994)[27], (Epstein & Manzoni, 1998)[28]. Moreover, the idea of linking measures to strategy is not restricted to the BSC (McNair *et al*., 1990)[29], Beischel & Smith, 1991)[30],(Grady 1991)[31].

The purpose of the benefits management process is to improve the identification of achievable benefits and to ensure that decisions and actions taken over the investment life-cycle can lead to realizing all the expected benefits. The majority of value from IT comes from the business changes that it enables the organization to make. The investment is in 'IT-enabled change', not just technology, to achieve improvements in business and organizational performance through better processes, relationships and ways of working. The achievement of benefits obviously depends on effective implementation of the technology, but evidence from projects success and failure suggests that it is organizations inability to accommodate and exploit the capabilities of the technology that causes the poor return from many IT investments (Ward & Daniel, 2006)[16]. The Benefits Realization Plan and the Benefits Dependency Network are means of ensuring these links are made and are the basis for the business case, since it includes not only what benefits are intended but also how each one can

be achieved. Once the expected benefits from the investments were different concerning their environment, nature and dimension, Shang and Seddon (2002)[32] proposed an enterprise systems benefits framework that includes twenty one benefits classified in five dimensions: Operational, managerial, strategic, IT infrastructure and organizational. Remenyi *et al.,* (2000)[33] points out four causes for difficulties on the benefits identification and management: (1) all benefits should be identified and quantified before the project starts, (2) Due to the complexity of IS/IT investments isn't easy to predict the development processes, (3) The dichotomy tangible/intangible and quantifiable/not-quantifiable were different organizational effects, (4) The benefits aren't stable along the investment.

In essence, the most obvious difference in both approaches, BSC and BM, is in the way benefits management plays its role in 'governing' the investment process. The BSC don't show evidence of having a process to manage the benefits, although the benefits are distributed through the perspectives ruled by the short and long-term objectives. Without clear benefit identification and a structured definition performed on a realization plan that includes the ownership, metrics and targets, there is a low chance of having an assured path towards benefits achievement.

Bennington and Baccarini (2004)[7] suggest that the benefits identification should be a combined approach of interviews and workshops involving key stakeholders. Remenyi and Sherwood-Smith (1998)[34] note that a key aspect of benefits identification process is that the stakeholders learn to better understands what is required from an investment and what is affordable and possible. Best practice is to involve key stakeholders to identify and agree desired benefits maximizing the likelihood of commitment to realize those benefits across a range of levels in the business or the organization (Glynne, 2007)[35]. Benefits at this stage should be structured in order to understand the linkages between technology effects, business changes and overall business effects (Sakar & Widestadh, 2005)[36]. As a process, BM ensures that the business changes achieve the expected results by translating business objectives into identifiable and measurable benefits that can be systematically tracked (OGC, 2003)[37], (Payne, 2007)[38], (Reiss *et al.,* 2006)[6].

Ward *et al., (*1996)[39] note that without a plan it will be difficult to predict how organization might effectively realise business benefits. Assessing the business benefits should not be a one-off task carried out in an initial phase of the implementation. Assessing the benefits is an on-going process because some benefits may not be immediate and will only appear at a later stage when the system has been fully integrated into the running of the organization by all of its users (Caldeira *et al.,* 2012)[40].

Benefits monitoring compares projects results with the benefits realization plan during the project and assesses if any internal or external change have occurred that will affect the delivery of planned benefits (Ward & Daniel, 2006)[16]. Benefits monitoring is a long cycle, according to Bartlett (2006)[41], starting with benefits planning and ending with the benefits realization. Benefits review is the process by which the success of the project in terms of benefits delivery is addressed, opportunities for the realization of further benefits are identified and lessons learned and opportunities of improvement in future projects are identified (Ashurst & Doherty, 2003)[1].

5 Linking Benefits Dependency Network to Strategy Map

Concerning the BDN and VIAPAV specific stream (Gomes, 2011)[42] shown on Figures 1 and 2, we proceed to the full integration of the BDN elements on the Strategy Map, considering O2 as an example.

In this first stage the organization have chosen the investments to perform, identified the business objectives (short-term) and benefits (long-term) to be realized through the business changes enabled by IS/IT investments. The linkage to the strategic objectives (BSC) should be supported on the business case. The business case includes the benefits realization plan with the routes to objectives mapped, the

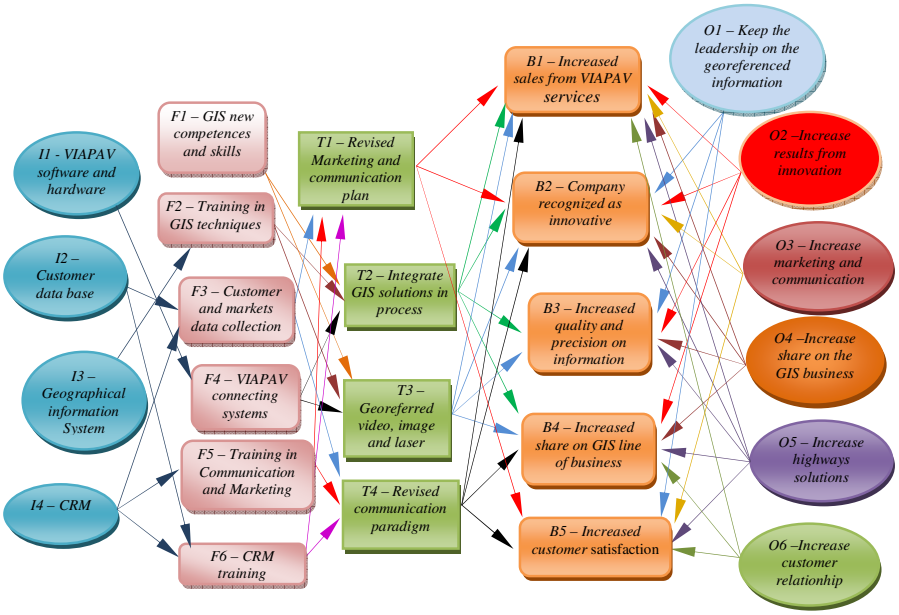


Fig. 1. Benefits Dependency Network (Gomes, 2011)[42]

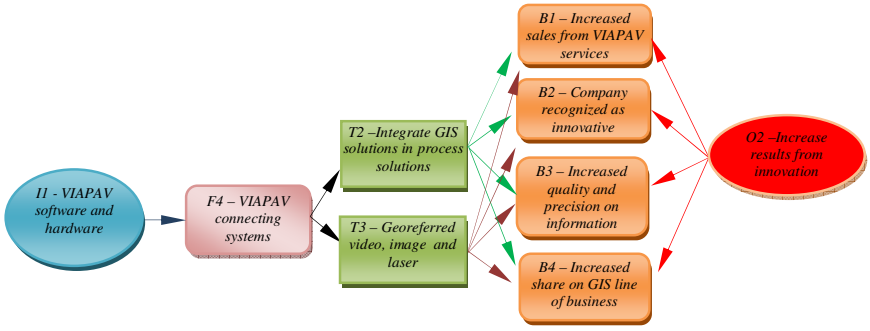


Fig. 2. VIAPAV specific stream (Gomes, 2011){42}

responsible nominations and the established metrics and targets. In the customer perspective of the BSC, managers identify the customers and the market segment in which the business will compete. The customer perspective enclosed several generic outcomes including, customer satisfaction, customer retention, new customer acquisition, customer profitability, market and account share and so on. The core of the business strategy is the customer value proposition which describes the unique set of products and service attributes, customer relations and corporate image. The value proposition defines how the organization differentiates itself from competitors and helps the organization to connect its internal processes to improved outcomes with its customers. The Internal Process and Learning and Growth perspectives are located in an area where the benefits are more intangible. These two perspectives are directly linked to the Enabling Changes and Business Changes from CMS model.

In this approach business changes has been described as new ways of working that will be required by the organization in the near future. The changes include a wide range of different types, for instance: new processes, new roles or responsibilities, new governance, new measures and metrics, new practices for managing and sharing information. There is also a wide range of enabling changes that may be required in order to ensure that the business changes really occur and promote the realization of the identified benefits. These enabling changes are only required to be undertaken once and may be necessary to allow the business changes. Enabling changes required were the following ones: training in how to use the new system and technologies, education in how the new systems can improve the performance, mapping of current processes and the design of new processes, definition of new roles, job descriptions, responsibilities and organizational structures, establishment of rules and practices, definition of new application and information governance structures.

The Internal Process reminds us that the organizational background activity is driven by objectives and goals to ensure that the customer and financial objectives are achieved. Once an organization has a clear picture of its customer and financial perspectives it can then determine the means by which it will achieve the differentiated value proposition for customers and the productivity improvements, the business changes, to reach the objectives and realize de business benefits.

The foundation of any strategy map is the learning and growth perspective, which defines the core competencies and skills, the technologies and the corporate culture. These topics enable the organization to align its human resources and information technology with its strategy.

6 Benefits Achievement

The table 1 shows the project outcomes (2008) and also highlights the benefits results tracked annually for the years 2009-2010. We could confirm that the expected benefits were achieved on different levels of realization.

B1 and B3 exceeded expectations. B1 weighting, in 2010, 7% of the company's total income. B3 reduced the performed hours in almost 40%. B2 reached good inquiry values, but the organization will continue investing on a closer customer relationship. After a good start (2008-2009), B4 dropped dramatically, mainly due to lack of private and government investments following the deep euro zone crisis. Exporting this type of services will be the solution in the near future.

Table 1. Investments outcomes (Gomes, 2011){42}

B1	Financial	Benefit			
Year	Viapav annual revenue (€)	Annual revenue (M€)	Objective (%)	Perform (%)	Target Deviation (%)
2008	95.000	6,750	1,0	1,4	0,4
2009	206.300	6,570	2,0	3,1	1,1
2010	409.644	6,120	3,0	6,7	3,7
B2	Quantifiable	Benefit			
Year	Customer Inquiries	Customer Answers	Answers (%)	Score (%)	Classification
2008	25	15	60	72	Good
2009	28	17	61	80	Good
2010	25	16	64	75	Good
B3	Financial	Benefit			
Year	Revenue (€)	Estimated hours	Perform hours	Time reduction (%)	Deviation (%)
2008	95.000	6.000	4.200	30	0
2009	203.300	10.200	6.280	38	+ 8
2010	409.640	20.120	12.320	39	+ 9
B4	Quantifiable	Benefit			
Year	GIS revenue (€)	Annual revenue (M€)	Objective (%)	Perform (%)	Target Deviation (%)
2008	200.000	6,750	3,0	3,0	0,0
2009	302.000	6,570	3,0	4,5	1,5
2010	20.000	6,120	3,0	0,3	-2,7%

The table 2 shows an integration table with the full conversion of the BDN elements on the company Strategy Map.

Table 2. Integration table

Benefits Management	Balanced Scorecard	Int benefits	Metrics	Target
Objective **O2-** Increase results from innovation	**Strategic Objective** **O1** - Increase results from innovation		Customer Satisfaction Inquiry	Better than 70% 2012
Output benefits **B1-**Increased sales from VIAPAV **B3 -** Increased quality and precision	**Financial** **F1 -**Increased sales from VIAPAV **F3 -** Increased quality and precision	**Int1 (2009)** Increased sales 1% **Int3 (2009)** 30% per year	 Sales Time reduction	Benefit: 3% 2010 Sales Manager Benefit: 30% Proj. Manager 2010
Output benefits **B2-** Recognized as innovative **B4 -**Increased share on GIS line of business	**Customer perspective** **C2-** Recognized as innovative **C4 -**Increased share on GIS line of business	**(2009)** **Int2 -**innovative >70% **Int_3** Increased share on GIS services 3%	Customer Satisfaction Inquiry Sales	Outcome 70% Benefit: > 70% I&D Manager Outcome 1% Benefit: 3% 2010 Sales Manager
Change **Enablers** **Enablers SI/TI** **I1-**VIAPAV soft&hard	**Learning and growth** **N1-** Align strategy **N2-** Team building **N3** Change management **L1-**VIAPAV soft&hard			(2007) Top Manag. H.R. Manager H.R.Manager Proj.Manager

In figure 3 we present the final Strategy Map showing the extended result of the integration process proposed in our work.

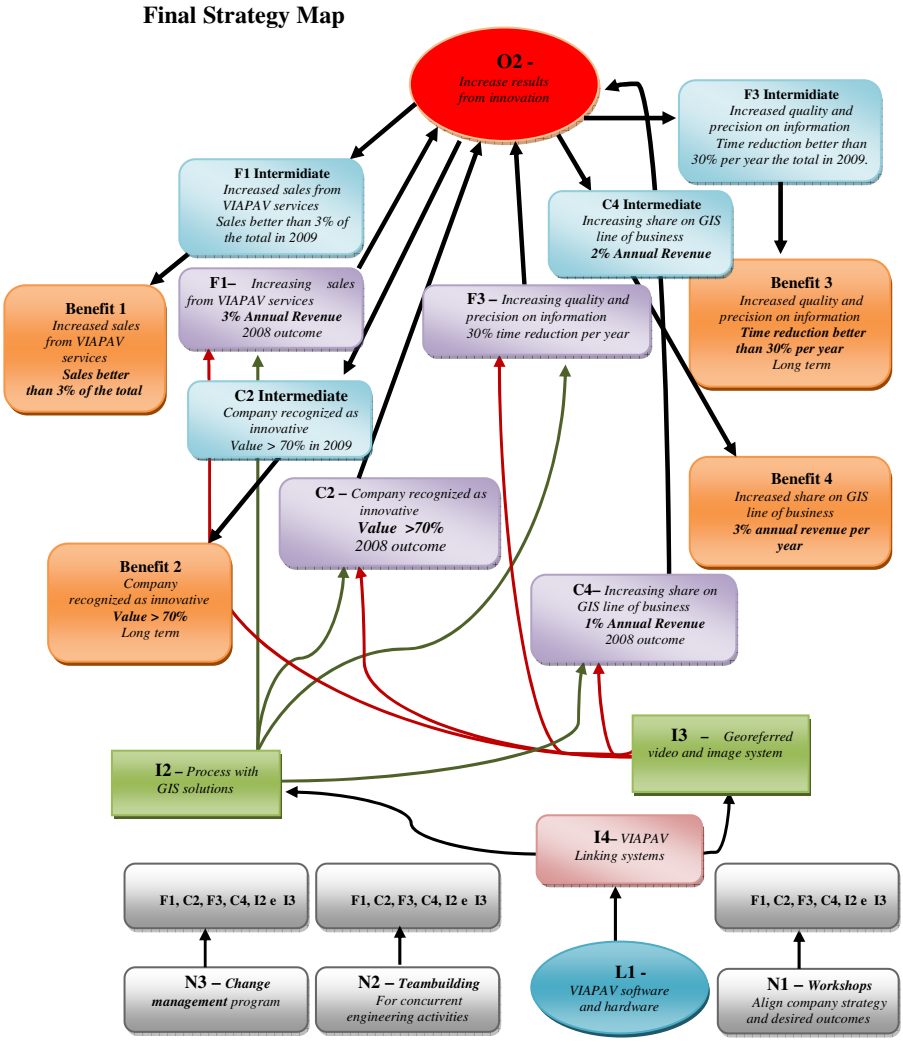


Fig. 3. Integrated benefits on the Strategy Map

7 Conclusions

Organizations are becoming interested in finding ways in which they can ensure their long-term sustainability, whether they are private firms, looking to maximize their shareholder value, or public sector and non-profit organizations, seeking to maximize their quality of service to citizens and/or effectiveness. Using the results of a case-study research, we explained how a 'Strategy Map', used in Balanced Scorecards, can cross with the concept of 'Benefits Dependency Network', from a Benefits Management approach.

The proposed framework highlights the "best of" both approaches. On one side the ability of the Balanced Scorecard to align its management processes and to focus the entire organization on implementing its long-term strategy. On the other side the Benefits Management approach brings to the Strategy Map the process of identifying and structuring benefits, establishing the ownership and determining whether benefits can be managed and measured along their life-cycle.

We have argued in favor of linking the strategic map from Balanced Scorecard with the benefits dependency network from Benefits Management. We claimed that this can enlighten the description how the company will achieve its desired outcomes to satisfying customers, shareholders and employees, including the value proposition (customer perspective), the innovation (internal process perspective), the employee skills and information technology capabilities (learning and growth perspective), and all combined to ensure that the identified benefits will be realized according to the expectations.

This framework results in a clear statement of benefits expected from an investment and the IT capabilities required to achieve those benefits, encourages a range of staff to work together because no single individual or group has all the knowledge necessary to identify all benefits and changes needed, and prevents spending money on projects that cannot deliver the desired benefits. From a practical point of view, the suggested link encourages the reutilization of previously developed Balanced Scorecards. As this common management practice is usually focused on building and monitoring the strategy completion, it is not so useful to drive the businesses changes initiatives. Then, the strategic map can easily integrate the results of other internal initiatives around benefits and projects management. Additionally, this link also promotes cross-fertilization and integrative efforts from top-level managers through change managers, project managers and at least some senior technical staff, somehow involved in projects and business change initiatives.

References

1. Ashurst, C., Doherty, N.F.: Towards the formulation of "a best practice" framework for benefits realization in IT projects. Electronic Journal of Information Systems Evaluation 6, 1–10 (2003)
2. Berghout, E., Nijlan, M., Grant, K.: Seven Ways to get Your Favored IT Project Accepted – Politics in IT Evaluation. Electronic Journal of Information Systems Evaluation 8, 31–40 (2005)
3. Lin, C., Pervan, G.: The practice of IS/IT benefits management in large Australian organizations. Information & Management 41, 31–44 (2003)
4. Markus, M.L.: Techno change management: using IT to drive organizational change. Journal of Information Technology 19, 4–20 (2004)
5. Remenyi, D., Money, A., Bannister, F.: The Effective Measurement and Management of ICT Costs and Benefits. CIMA Publishing, Burlington (2007)
6. Reiss, G., Anthony, M., Chapman, J., Leigh, G., Pyne, A., Rayner, P.: Gower Handbook of programm management. Gower Publishing (2006)
7. Bennington, P., Baccarini, D.: Project benefits management in IT projects – An Australian perspective. Project Management Journal (June 2004)

8. Ward, J., Elvin, R.: A new framework for managing IT-enabled business change. Information Systems Journal 9, 197–221 (1999)
9. Bradley, G.: Benefits Realization Management: A Practical Guide to Achieving Benefits through Change. Gower Publishing Limited (2006)
10. Peppard, J., Ward, J.M.: Unlocking sustained business value from IT investments: Balancing problem-based and innovation-based implementations. California Management Review (2005) (forthcoming)
11. Atkinson, A., Epstein, M.: Measure for Measure: Realizing the power of the balanced scorecard. CMA Management, 23–28 (September 2000)
12. Frigo, M., Krumwiede, K.: The Balanced Scorecard: A Winning Performance Measurement System. Strategic Finance, 50–54 (January 2000)
13. Kaplan, R.S., Norton, D.P.: Putting the Balanced Scorecard to Work. Harvard Business Review (September-October 1993)
14. Otley, D.: Performance management: a framework for management control systems research. Management Accounting Research 10, 363–382 (1999)
15. Kaplan, R.S., Norton, D.P.: The Strategy-Focused Organization: how Balanced Scorecard companies thrive in the new business environment. Harvard Business School Press, Boston (2000)
16. Ward, J., Daniel, E.: Benefits Management, Delivering Value from IS and IT Investments. John Wiley & Sons, Chichester (2006)
17. Peppard, P., Ward, J., Daniel, E.: Managing the Realization of Business Benefits from IT Investments. MIS Quarterly Executive (March 2007)
18. Kaplan, R.S., Norton, D.P.: Using the Balanced Scorecard as a Strategic Management System. Harvard Business Review, 75–85 (January-February 1996b)
19. Kaplan, R.S., Norton, D.P.: The Balanced Scorecard: Translating Strategy into Action. Harvard Business School Press (1996a)
20. Nørreklit, H.: The Balanced Scorecard - a critical analysis of some of its assumptions. Management Accounting Research (March 2000)
21. de Haas, M., Kleingeld, A.: Multilevel design of performance measurement systems: enhancing strategic dialogue throughout the organization. Management Accounting Research 10, 233–261 (1999)
22. Atkinson, A., Epstein, M.: Measure for Measure: Realizing the power of the balanced scorecard. CMA Management, 23–28 (September 2000)
23. Schiemann, W.A., Lingle, J.H.: Hitting your strategic targets through high impact measurement. The Free Press, New York (1999)
24. Anand, M., Sahay, B., Saha, S.: Balanced Scorecard in Indian Companies. Vikalpa 30, 11–25 (2005)
25. Meyer, M.: Rethinking Performance Measurement: beyond the Balanced Scorecard. Cambridge University Press (2002)
26. Davies, R.: The Balanced Scorecard: Panacea or Poisoned Chalice? Executive Briefing (2007), http://www.Drrobertdavies.com
27. Lebas, M.: Managerial accounting in France - Overview of past tradition and current practice. The European Accounting Review, 471–487 (1994)
28. Epstein, M., Manzoni, J.F.: Implementing corporate strategy: from tableaux de bord to balanced scorecards. European Management Journal 2, 190–203 (1998)
29. McNair, C.J., Lynch, R.L., Cross, K.F.: Do financial and nonfinancial performance measures have to agree? Management Accounting, 28–36 (November 1990)
30. Beischel, M.E., Smith, K.R.: Linking the shop floor to the top floor. Management Accounting, 25–29 (October 1991)

31. Grady, M.W.: Performance measurement: implementing strategy. Management Accounting, 49–53 (June 1991)
32. Shang, S., Seddon, P.: Assessing and managing the benefits of enterprise systems: the business manager's perspective, Info. Systems J., 12, 271–299 (2002)
33. Remenyi, D., Money, A., Sherwood-Smith, M.: The Effective Measurement and Management of IT Costs and Benefits. Oxford Butterworth-Heinemann (2000)
34. Remenyi, D., Williams, B., Money, A., Swartz, E.: Doing Research in Business and Management. Sage Publications, London (1998)
35. Glynne, P.: Benefits Management-changing the focus of delivery. Association for Progress Management Yearbook 2006/2007, 45–49 (2007)
36. Sakar, P., Widestach, C.: Benefits Management – How to realize the benefits of IS/IT investments. Master Thesis in Business Technology, IT University Göteborg, Sweden (2005)
37. OGC, Managing Successful Programmes, The Stationery Office, London, Office of Government Commerce (2003)
38. Payne, M.: Benefits Management – Releasing project value into the business. Project Manager Today, Hampshire (2007)
39. Ward, J.M., Taylor, P., Bond, P.: Evaluation and realization of IS/IT benefits: An empirical study of current practice. European Journal of Information Systems 4, 214–225 (1996)
40. Caldeira, M., et al.: Information and communication technology adoption for business benefits: A case analysis of an integrated paperless system. International Journal of Information Management (2012), doi:10.1016/j.ijinfomgt.2011.12.005
41. Bartlett, J.: Managing Programmes of Business Change, Hampshire, UK (2006)
42. Gomes, J.: Gestão de benefícios numa empresa de Geoengenharia. Master thesis in management, ISCTE, Lisboa (2011)

Modeling e-Government for Emergent Countries: Case of S.Tome and Príncipe

Artur Celestino Vera Cruz

Rua Soldado Paulo Ferreira, Edifício Pereira Duarte,
C.P. 86 – S.Tomé
arturveracruz@yahoo.com.br

Abstract. E-government is assumed nowadays as a common and transversal preoccupation to all the modern states. Therefore, in spite of having several methods for planning of the Information Systems, the specificity of Public Administration imposes the need to think in modeling of a method for global planning. This research provides a reflection on how to conduct and implement an e-government initiative in emergent's countries, with special focus on the case of São Tomé and Principe.

Keywords: E-government, method, methodology, Information System, (IS) Information Systems Planning (ISP), Public Administration (PA).

1 Introduction

E-government is assumed nowadays as a common and transversal preoccupation to all the modern States and seen as a way to guarantee the efficiency of public services, to provide surplus value to the citizens and to reduce expenses of the Government.

This concern translates hitself in a process of reengineering of the leadership of the State and of the Public Administration (PA) [1] and [2], supported by the Technologies of Information and of Communication (TIC), that doesn't prove easy task, due to specificity of the organization of PA, much characterized by a multiplicity and complexity of public's services, which differs a lot of the private sector [3].

The project here designed is assumed as a task of building a model for reengineering the Public Administration. With it, we intend to project a global architecture of e-government initiative for "emergent countries", which imposes a necessary reflection concerning the architecture of applications and technologies, its implementation plan, the alignment of the plan with budget and the consequent references of politics.

To assist the need above referred, the first challenge is to find a methodology for planning of IS in PA.

Considering that in this process, the leaders have a primordial role in materializing the vision of a reform for PA, we try to provide a method of ISP that facilitates their intervention, attempting to carry on the best modernization of the State.

Through the inventory of existent Methods, we verified that exists a vast diversity of methods of ISP. However, they are not appropriate for planning complex and

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 371–381.
DOI: 10.1007/978-3-642-36981-0_35

heterogeneous systems like PA, mainly due to the fact that they have been conceived for the private sector, which has as its main intent the creation of advantages over competition. In addition, the environments, the objectives, the "modus operandi", the "modus vivendi" are different of PA reality [4].

The specificity above referred is the reason for which several existent Methods are limited or inadequate to the global planning of the Information Systems (IS) in PA, thus justifying the need to think in modeling a method for global planning of IS with capacity to support the complexity of PA systems. In order to do so I have considered the possibility to consolidate the strongest points of the existing Methods to achieve a global planning of PA. Considering this possibility, the first question is to known which, among several methods, should be selected. The approach was the inventory of methods that present certain continuity and to extract its strong points.

With this work I aim to develop a methodology for a global planning of PA, then to apply it to the case of PA of São Tome and Príncipe, also presentation some references of unified and integrated politics for IS in PA.

2 Methodology for Global Planning of Public Administration

2.1 The Proposal Methodology and Activities of Its Phases

The Method for a Global Planning of Information System in PA referenced along the Thesis under the acronym MPGAP_e-gov is the amalgamation of strong point of the four mentioned methods that identifies respectively with attributes simplicity, flexibility, usability and applicability protected by Carapeto and Fonseca [16], and is supposed to impel and to engage government structure to assume responsibility of larger involvement in the elaboration and maintenance of architecture of information in PA, and of its plan of implementation for medium and long term.

The proposal methodology include 7 (seven) phase. Commentary of activities carried out in each phase:

0. Preparing

Should result an approved working plan, reflecting the approach of methodology.

1. Envisioning primordial objective of State

Should provide the view of organizational structure of PA and how the competences are legally distributed according to primordial objectives of states.

2. Envisioning Integrated Public Administration

Make available the actual organizational architecture and IT of PA grouping entities according to it attribution in prosecution the primordial objectives of states: safety and defence, education and culture, social well-being, well-being economic and justice.

3. Defining Public Administration Processes

The output is a list of processes, description of each one, hierarchy responsibilities of entities and matrixes process/ organization.

4. Defining Public Administration Data

Provide a list of data classes according to description of process identified in preview phase.

5. Definition Information Systems Architecture

Provide a new organizational architecture for PA according to five primordial objectives of state and projecting new architecture for data, selecting applications and platforms of technology.

6. Definition of references of politics

Considering the proposal architecture, defining politics for the success of PA reengineering, according to experience of more advanced as well as the opportunities supplied by news tools.

7. Reviewing and Results

The purposes is providing bases for evaluation the success of initiative and guarantees the connection of strategy with the planning and the budget.

Even so, as the one intent is Modeling of e-government, reconciled to MPGAP_e-gov, IT references of integrated politics for the horizontal and vertical integration of IS and the centralization of the informatics systems.

2.2 Approach to the Methodology

The term Method or methodology is presented along the work to express in generic terms, "the way of do the things" that is to say, the operacionalization of the approach [5].

Considering that the success of the activity in IPS has a great support in the method used, forcing a very careful and thoughtful choice, focusing on the purposes and the circumstances of each situation in particular [6], [7] and [8], as approach, were selected four methods, beeing 3 (three) Methods of Information System Planning (ISP) and 1 (one) of strategic administration.

The 3 (three) Methods of Information System Planning (ISP) selected are those that present a certain continuity in the development and whose comprehension does not imply a very technical knowledge, such as: Business Systems Planning (BSP) [9]; Framework for Information System Architecture, also, well-known as structure Zachman [10] and [11]; Enterprise Architecture Planning (EAP) [12]. The Method of strategic administration selected is denominated of Balanced Scorecard [13] and [14].

The vehicle consisted in melting the four mentioned methods in a single method capable to define a global architecture for e-government, through the conciliation of its strong points.

General information system planning approach is summarized in four steps: Step 1 is about defining the business objectives; step 2 defines business processes, step 3 is about defining business data and step 4 defines the information architecture. Its overlook is top-down analyse and bottom-up implementation, with the adaptation of business objectives to correspond to the primordial objective of the State; the business

organization to the organization of the Public Administration; the business processes to entity processes, the business data to entity data, the database to the metadata, and the applications to meta-applications, according to the outline illustration (Figure 1).

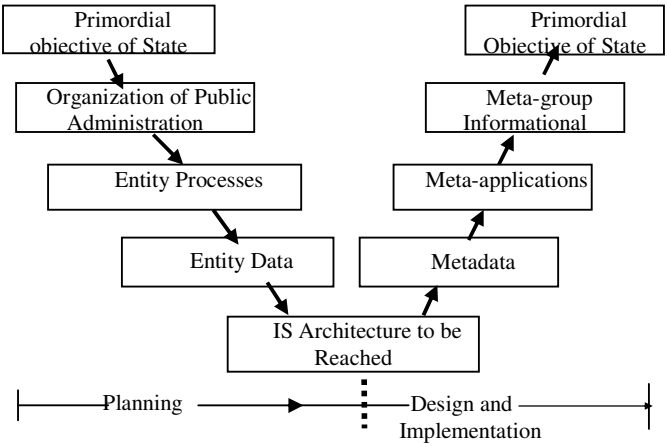


Fig. 1. Top-down analysis with bottom-up implementation. This figure consisting of an information system planning approach adapted from IBM [1984, p.9].

The BSP methodology is reference of many other methodologies of ISP. It was conceived for the private sector, but is used here by adaptating it to support complex systems. The advantages of this method correspond in long term to the objectives of an IS, in supporting the objectives and needs of information of the organization; in corresponding to the needs of all administration levels; in assuring consistency of the information and in surviving the organizational and management changes. From its 13 stages, the conciliation includes them entirely, as shown in figure 2.

Even so, with BSP the planning process is developed until the definition of organization information architecture, that is to say, the information flow. Therefore it does not conceive the architecture of applications and of necessary technologies nor its implementation plan.

With intent to resolve above mentioned problems, we associated the Information System Architecture (ISA), developed by Zachman as a logical structure for connection of the products of BSP, with the most technical products of the implementation. ISA supplies a sequential vision from conception to the final product, helping organizations to migrate from a legacy system for architecture through Data warehouse, bringing as advantage a macro vision of the architecture of an object, product, organization and IS. However it leaves, also, a vacuum due to the fact that it does not describe the form of doing this connection. The 2 (two) first levels are considered for adaptations to the methodology of reference, which are, the proprietor's vision and the vision of the planning personnel considered crucial to support all planning phases.

Methodology of Reference (BSP) \ Adaptations to the Methodology	0. Preparing	1.Visioning primordial objective of State	2. Visioning Integrated Public Administration	3. Defining Public Administration Processes	4. Defining Public Admnistration Data	5. Definition Information Sistems Architecture	6. Definition of references of politics	7. Reviewing and Results
1. Gaining the commitment	■							
2.Preparing for the Study	■							
3. Starting the Study		■						
4. Defining Business Processes			■	■				
5. Defining Business Data					■			
6.Defining Information Architecture						■		
7. Analyzing Current Systems Support						■		
8. Interviewing Executives						■		
9. Defining Findings and Conclusions						■		■
10. Determining Architecture Priorities								■
11. Reviewing Information Resource Management						■		
12. Developing Recommendations						■		■
13.Reporting Resoults						■		■

Fig. 2. Business Systems Planning (BSP) and adaptation. This figure consisting in conciliation of 13 stages of BSP with 7 phases of proposal methodology refereed in point 2.1.

Due to the deficit from BSP and ISA we reconciled to Enterprise Architecture Planning (EAP), for it contains in its structure elements of both, in an attempt trying to describe the road in a more complete way to proceed for definition the architecture of information, applications, technologies and its plan of implementation. However, there are diagrams whose elaboration demands the specialists' involvement in IS, implying risks of the high administration to leave the conclusion of this activity under the personnel's responsibility of IT. Thus, were taken advantages of phase II (to document the organizational structure, the functions and the organization model), IV (architecture of data), V (architecture of applications) and VI (IT existent in the market with future usefulness for PA).

Adaptations to the Methodology / Information System Architecture (ISA)	0. Preparing	1.Visioning primordial objective of State	2. Visioning Integrated Public Administration	3. Defining Public Administration Processes	4. Defining Public Admnistration Data	5. Definition Information Sistems Architecture	6. Definition of references of politics	7. Reviewing and Results
1. Architect's vision	■	■	■	■	■	■	■	■
2. Proprietor's vision	■	■	■	■	■	■	■	■
3. Designer's vision								
4. Constructor's vision								
5. Vision or subcontracted								

Fig. 3. Information Systems Architecture (*ISA*) and adaptation. This figure consisting in conciliation of 5 stages of ISA with 7 phases of proposal methodology refereed in point 2.1.

Adaptations to the Methodology / Enterprise Architecture Planning (EAP)	0. Preparing	1.Visioning primordial objective of State	2. Visioning Integrated Public Administration	3. Defining Public Administration Processes	4. Defining Public Admnistration Data	5. Definition Information Sistems Architecture	6. Definition of references of politics	7. Reviewing and Results
1. Initial Planning			■					
2. Developing organizational model								
3. Relating systems architectures & existent Technologies								
4. Defining Data Architecture						■		
5. Defining application Architecture						■		
6.Defining technological Architecture						■		
7. Developing strategy and implementation plan								■

Fig. 4. Enterprise Architecture Planning (*EAP*) and adaptation. This figure consisting in conciliation of 7 stages of EAP with 7 phases of proposal methodology refereed in point 2.1.

With the fusion of 3 (three) Methods of ISP above referred (BSP, structures Zachman and EAP) the reach of the process of ISP is developed from the architecture of information, applications, technologies, until the plan of implementation.

Even so, it doesn't show the relationship between expenses in Information Technologies and the return caused by these investments, in a perspective of future actions [15] while one of the objectives of this modeling is fact that required consequently reconciling in Balanced Scorecard.

Method Balanced Scorecard (BSC), developed by teachers Kaplan and Norton in 1992, is not a method of ISP, but is here admitted in the present research as a referential to supply the desirable relationship as above referred. Among its strong points, it provides a holistic vision of new paradigms of IT and allow extracting from output, lessons, advices and experience for the success of future action; aligning the indicators of result with indicators of tendency, as well as, aligning of organization with the strategy and also; guarantees the connection of strategy with the planning and the budget.

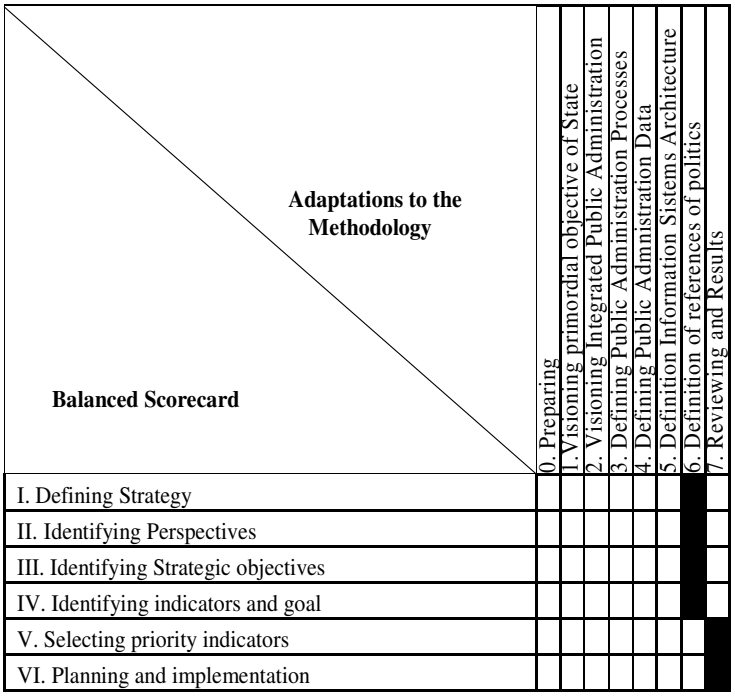


Fig. 5. Balanced Scorecard with adaptation. This figure consisting in conciliation of 6 stages of Balanced Scorecard with 7 phases of proposal methodology refereed in point 2.1.

3 Application the Methodology to the Context of S.Tomé

In order to validate its valences the MPGAP_e-gov was applied to the context of PA in São Tomé and Príncipe, while being a factor of motivation for its popularization and adoption.

The constraint of the insularity and smallness of the country constituted strong points to embrace many subjects of PA in São Tomé and Príncipe, once it created a space for dialogue with the several responsible of the Administration of the State, despite having caused long delays, it provided opportunities to lean on in the experiences of the most advanced countries and to avoid mistakes in planning and implementation of the process. Though, there is the need to implement clear public politics in articulation with others politics, for example, of education, of the society of information, since the development is made with active and integrated politics, in that the institutional factors must show themselves capable to open and to maintain an ideal environment that allows to the citizens, companies and visitors to find the necessary integration adjustments in the society.

Highlighting the State primordial objectives allowed projecting a new strategic architecture for organization of PA as shows figure 5. In this new architecture PA is seen in an integrated way and structured as plurality of interlinked companies, with strong cooperation, relationship, partition of information and knowledge, with inter-department mobility of information, human, material and financial resources. There are identified 5 (five) groups of entities (ministries and organs) with specific attribution and vocations in concerning materialization the five primordial objectives of State, such us, safety and defence, education and culture, social well-being, well-being economic and justice; 1 (one) meta-group informational and; 1 (one) meta-group resources.

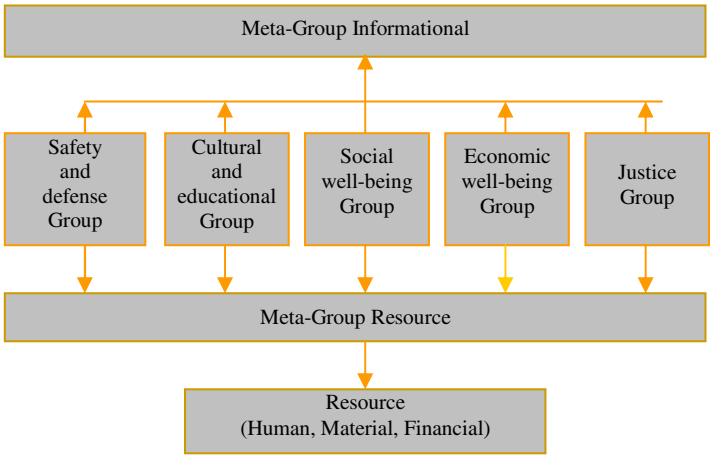


Fig. 6. Strategic organizational architecture for Public Administration (PA). This shows a figure consisting of an integrated vision of PA as an unique system with plurality of services.

In this strategic architecture, emerge two meta-groups that fortify the vision of AP as an integrated system. The meta-group informational should be structuring, integrator and manager of information resource, as well as regulator of the politics of IT in S.Tomé Public Administration, while the meta-group resource will have the basic function in budget programming and execution as well as in production the financial and patrimonial rapport of the State account.

The following figure is the organizational architecture proposed for PA in new era of network services according to result of application the proposal mythology in S.Tomé and Príncipe.

The organization of PA in emergent countries like São Tomé and Principe stands back of the integrated model of the State, with almost null articulation intra and inter government department.

As result of proposal methodology is provided an architecture (figure 6) that approach to the philosophy of the structured net of companies that should work in clustering regime but supported by two met-groups, information and resources. The objective is to provide certain mobility among several department of government in a perspective of autonomous entities and in same time integrated as net system grouping entities through inter-implication in the materialization of the primordial objectives of the State.

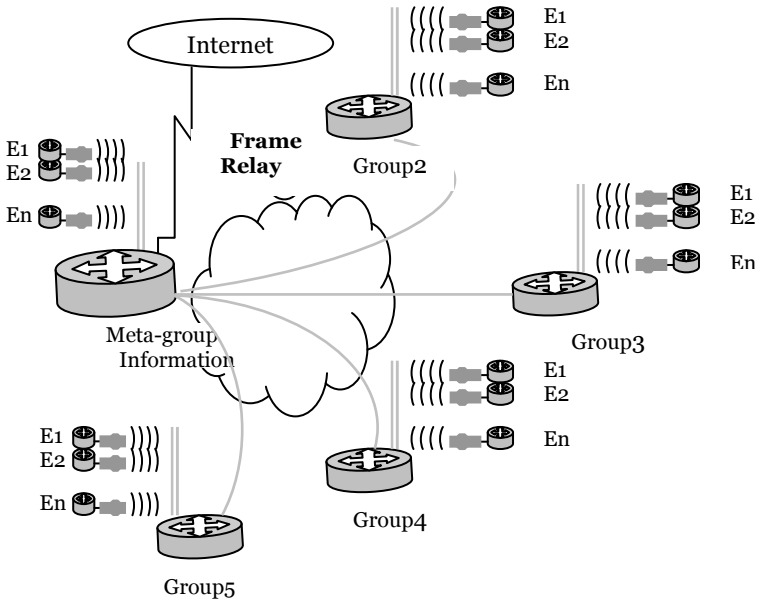


Fig. 7. Wide Area Network (*WAN*) for Public Administration. This shows a figure consisting of a vision of Public Administrations as net structured clustering by services according to primordial objectives of States.

4 Conclusion

Through the amalgamation of strong points of several methods of information systems planning conceived with main intent of creating competitive advantages on the opponent companies, is provided a method for global planning of public administration with capacity to support its complex systems.

With the proposed method, the information systems planning in public administration is developed from the information architecture of the organization, architecture of applications and technologies, until the plan of implementation and provides guidelines for future alignment with new paradigms of IT.

Combining the technological slopes with the five primordial objectives of any State [18] it is possible to build a methodological approach to global information system architecture for public administration, cantered in the needs of the reorganization of services and with flexibility to maximize the interoperability among the several systems of PA answering what Boar highlights [19].

References

1. Clinton, B., Gore, A.: The Team For Change (1992), Free in: http://www.time.com/time/magazine/0,16641,19920720,00.html
2. Vidigal, L.: A Face Oculta da Administração Pública Electrónica - Uma Abordagem Socio-Técnica. In: Amaral, L., Magalhães, R., Morais, C.C., Serrano, A., Zorrinho, C. (eds.) Sistemas de Informação Organizacionais, pp. 525–552. Edições Sí, Lisboa (2005)
3. Bellamy, C., Taylor, J.: Governing in the Information Age. Open University Press (1998)
4. Amaral, L.A.M., Varajão, J.E.: Planeamento de Sistemas de Informação. FCA (2000)
5. Olle, T.W., et al.: Information Systems Methodologies: A Framework for Understanding. Addison-Wesley Publishing Company (1988a)
6. Earl, M.J.: Experiences in Strategic Information Systems Planning. MIS Quarterly (1993)
7. Goodhue, D.L., Quillard, J.A., Rockart, J.F.: Managing the Data Resource: A Contingency Perspective. MIS Quarterly (1988)
8. Swede, V.V., Vliet, J.C.V.: A flexible framework for contingent information systems modelling. Information and Software Technology (1993)
9. IBM. usiness Systems Planning: Information systems planning guide. Fourth Edition. IBM Department 796, P.O. Box, Atlanta, Georgia 30055. Copyright International Business machines Corporation (1984)
10. Zachman, J.A.: A framework for information systems architecture. IBM Systems Journal 26(3) (1997)
11. Zachman, J.A., Sowa, J.F.: Extending and Formalizing the Framework for Information Systems Architecture. IBM Systems Journal 31(3), 590–616 (1992)
12. Spewak, S.H., Hill, S.C.: Enterprise Architecture Planning: Developing a Blueprint for Data, Applications and Technology. QED Publishing Group (1993)
13. Nolan, R.L.: Managing the Computer Resource. A Stage Hypothesis Communications of the ACM 16(7), 399–405 (1973)

14. Nolan, R.L.: Managing the Crisis in Data Processing. Harvard Business Review 57, 115–126 (1979)
15. Strassmann, P.A.: The Business Value of Computers. The Information Economics Press, New Canaan (1990)
16. Carapeto, C., Fonseca, F.: Administração Pública – Modernização, Qualidade e Inovação. Edições Sílabo, Lisboa (2006)
17. Carapeto, C., Fonseca, F.: A Ferramenta CAF. Gestão Pura, nº 3 (2006)
18. Miranda, J.: Manual de Direito Constitucional, Tomo I. Preliminares: O Estado e os Sistemas Constitucionais, (6th Edição). Revista e Actualizada. Coimbra Editora (1997)
19. Boar, B.H.: Practical steps for aligning information technology with business strategy, 2nd edn. John Wiley & Sons, Inc., New York (1995)

An Analysis of the Disclosure of Social Responsibility in Australian Universities

Raquel Garde Sánchez, Manuel Pedro Rodríguez-Bolívar,
Laura Alcaide-Muñoz, and Antonio M. López-Hernández

Departamento Economía Financiera y Contabilidad,
Facultad Ciencias Empresariales y Económicas, Universidad de Granada
Campus de la Cartuja, s/n, 18071, Granda, Spain
{rgarde,manuelp,lauraam,alopezh}@ugr.es

Abstract. Recently, there seems to be some consensus that universities should play a key role in the issues of social responsibility (SR). The higher education system in Australia has undergone of government reforms, providing funding in order to integrate SR issues in their strategic planning and to implement New Technologies of Information (ICT). This paper analyzes the policies and strategies of communication about SR information and wonders if Australian universities are relying on ICT as a means both to disseminate issues of SR and to facilitate interaction with stakeholders. The results show that the disclosure of SR reporting by Australian universities is very low, only with the main aim of legitimacy purposes.

Keywords: Corporate Social Responsibility, Universities, Online disclosure, Australia, New Technologies.

1 Introduction

In recent years, there seems to be some consensus that universities should play a key role in the issue of social responsibility (SR) through the incorporation of SR principles in the design of its educational programs and research, as its mission, vision and strategy [35]. The previous literature highlights the pressures that have undergone the universities of different countries, for example, in New Zealand (period 1985-1989) or the United Kingdom (1992-1994) [8] under a larger and broader accountability of universities, especially in regard to disclosure of university social responsibility issues. Conform to the expectations of various stakeholders of the universities constitute mainly a legitimacy instrument to favor the process of acceptance and approval of their activities in the social environment around them [20, 21].

Up to now, universities are expanding their field by integrating the concept of sustainable development into the daily reality of the campus [46] and are driving the development of social responsibility in two directions clearly defined. On the one hand, in addition to this social responsibility in his research work and university extension, taking into account the essential role that universities develop knowledge

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 383–398.
DOI: 10.1007/978-3-642-36981-0_36

transfer [48], are integrating the concepts of responsibility office within the university educational system [7]. On the other hand, universities are expanding their field of action in social responsibility issues as they are more aware of the negative impact and/or positive feedback from its actions on the environment around them [53] as well as the need to have a key role in this type of action as a model of ethical behavior toward society [53]. In this sense, although some universities already have a long history in the implementation of initiatives related to social responsibility [35], most of them were focused primarily on measures concerning environmental concerns [47, 26].

However, the application of social responsibility to the field of higher education involves, as key issues, identifying the expectations of different stakeholders, establishing mechanisms for dialogue with them, and improving the transparency of information, reinforcing, thus, compliance with the duty of accountability. However, these aspects related to the implementation of accountability in the field of social responsibility seem to have been forgotten. From the viewpoint of transparency of information, universities are increasingly being pressured to improve their disclosure of information in the field of social responsibility [8] because, traditionally, the disclosure has been focused exclusively on financial information and budget, without disclosing the aspects of social responsibility [37].

In the last years, the parallel development of the New Information Technologies (ICTs) could favor the development of these activities in the universities, especially, the Internet could play a key role in the transparency, improving the process of accountability and allowing for improved communication and increased access to information for stakeholders [11]. In this regard, it would be interesting to know whether universities, as models of ethical behavior and society of knowledge and research center on issues of social responsibility, are seizing the opportunities offered by new information technologies, especially internet, as a means for the disclosure of social responsibility and for interaction with stakeholders.

Regardless of the quality of teaching and research universities, which is an essential element of the reputation of these [47], the socially responsible university, and its reflection through the transparency of information, could lead to a competitive advantage in the market for higher education and, at the same time, it becomes an instrument to legitimize the actions of the university towards the stakeholders, since the socially responsible disclosure has become more and more information demanded by stakeholders [3]. In fact, previous research has shown that the provision of information by universities on their websites is an essential element in making students' decision when choosing a college in which to develop future training [45] and has highlighted the need to disclose aspects to assess the performance of the university in the field of social responsibility [25].

In Australia, the higher education system has undergone a major package of government reforms towards improving quality of research and education [4]. In recent years, the Australian government has allocated large sums of money in higher education, with an additional commitment to continue investing in it in the future, because, although the performance of universities is very high, they are currently facing major challenges. On the issue of social responsibility, the Australian government has pledged to make a special investment in higher education, providing funding in areas that relate to such issues, especially sustainability, energy and

climate change, among other issues, from both the academic and research point of view [4], since, as demonstrated by previous research, integrating aspects of RS and sustainability in all the main activities of educational institutions is a great opportunity that allows the universities themselves so place in a position of competitive advantage [24].

Despite the previous comments, no research exists about if Australian universities, as leading universities in the implementation of ICT in the educational system [19], are committed to disclosing social responsibility information as a key point of their corporate SR policies and strategy. In this regard, it would be relevant to wonder if, in the Internet Age, Australian universities are relying on ICT to meet these new reporting requirements and to improve interaction with their stakeholders. In fact, there is a lack of studies which have been focused on the analysis of online communication policies of these universities on issues of social responsibility.

The contribution of our paper, in this sense, is twofold. On the one hand, based on previous research in different fields of knowledge, we have designed and proposed a model to analyze and evaluate the online SR information in universities and this is applied to all universities in Australia to see if the online disclosure of social responsibility is still a fundamental aspect of their communication policies. On the other hand, we observe if Australian universities are using ICT to encourage interaction with their stakeholders in relation to social responsibility issues.

The aim of this article focuses on knowing whether, in the framework of accountability, universities are using ICT in their communication policies as a means to disseminate both issues of social responsibility to facilitate interaction with stakeholders and thus promote participatory management aspects of their social responsibility.

To achieve this aim, the article is organized as follows. Section 2 focuses on describing the system of higher education in Australia and how social responsibility has been highlighted in recent years as an essential aspect of accountability. Section 3 describes the empirical research indicating the sample and research methodology, proposing, ultimately, a model for the evaluation of online information disclosure of social responsibility. Section 4 discusses the main results of empirical study. We conclude our work with the main conclusions and discussions.

2 The Reform of the Australian University System and the Role of Social Responsibility

The higher education sector in Australia is predominantly public [1]. Australian universities operate in an environment characterized by a high degree of control and central government planning that determines the institutional funding, is involved in setting goals, tuition controls, and establishes regulatory controls that affect, directly or indirectly, the educational system [41].

Parallel to other higher education institutions, universities in Australia have undergone profound changes over the past fifteen years, mainly as a result of changes in the policies of the Federal Government. On the one hand, the global financial and economic crisis of recent years has affected education in general, not only in terms of threats of cutting budgets and learning opportunities, but also in terms of educational

aims themselves [54]. Public funding per student has declined dramatically, and has pushed universities to develop alternative sources of revenue to maintain the core activities of teaching and research [15]. Furthermore, questions have been raised as to the type of society and education received by students in the future. Society demands that education is able to respond to the changing reality around us and contribute to a different future [51].

In this context, an initiative whose origin has taken place in Australia, together with the universities of New Zealand, has been the creation of the forum "Australasian Campuses towards Sustainability Association" whose main mission is to promote the integration of ecologically sustainable development principles in curricula and research, and university management [2]. Thus, education professionals have been called to reorient curricula and teacher training programs to integrate sustainability plans and social responsibility into the curriculum in order to help students to understand current social, economic and environmental problems to achieve solutions in the future.

In brief, the objective is to integrate aspects of SR and sustainability in all the major activities of educational institutions. Thus, students with to improved knowledge and skills will have the potential to profoundly affect the social, environmental and economic sustainability and this presents a great opportunity [24]. This is the reason why many Australian universities participate in the creation of international groups, such as the Association of University Leaders for a Sustainable Future [52], and support the approach of the SR and sustainability in teaching, research, and in all operations undertaken by the university [54].

Similarly, Australian universities have joined in various international declarations that promote training and research in the field of university social responsibility. Examples of this have been the signing of the Talloires Declaration towards a sustainable world coordinated by leaders of universities around the world, based in the US or the Swansea Declaration developed in 1993 by the Association of Commonwealth Universities, which states Community colleges should improve their ability to teach and undertake research in sustainable development to increase knowledge of the environment, and to improve understanding of environmental ethics within the university and the general public [51].

Nevertheless, up to date, no Australian university has formally recognized the Declaration of Swansea. Furthermore, despite international appeals for sustainable development and social responsibility in the field of higher education, Australian universities have been slow to implement such policies [9]. Previous studies, such as [50], found that sustainability initiatives in Australia and, in particular, higher education institutions, focus on individual projects to address sustainability, rather than a systemic view of learning and change through the institution.

Australian government, whom in response to the Decade of Education for Sustainable Development and for promote corporate social responsibility, recognized the need to integrate sustainability into the curriculum with a vision of Australian education capable of understanding, knowing, and having the skills and capacity to achieve sustainability. One positive sign was when the Committee of Vice-Chancellors of Australia recognized the important role that universities play in promoting the objectives of Sustainable Education declaring its commitment to education for sustainable development. They expressed the desire that the Australian

university sector become a key player in efforts to achieve sustainable society through their commitment to the communities, leading the ability to change in this regard.

Therefore, it is of vital importance to highlight the social role of universities as institutions that prepare future professionals in ethical evaluation processes of companies or organizations that they lead, providing a world view and understanding the role of business in society [44]. Australian universities have made progress in these areas including, among other actions, social responsibility courses to train future business leaders [12].

However, according to the Australian Learning and Teaching Council [5], in Australia there is no model where successful practices can be undertaken in this field because the actions of university social responsibility currently being developed in Australia are very specific and are widely scattered. That is why one of the first actions taken by the ALTC along with the University of Western Sydney has been the creation of a website for the sustainable higher education sector in Australia, with the aim of grouping in the same sustainability courses that are offered throughout the country.

From the point of view of the research, the Australian Institute for Research in Education for Sustainability, established in 2003 by the Government of Australia, aims to carry out projects with government and business to develop their capacity for improvement in sustainability [49].

In any case, despite such progress, and as stated by [42] Australian universities face a crisis of resources and morale. It is important to highlight the social role of universities based on the belief that the university has the obligation to ensure the development of critical thinking and research and to train professionals who can do the same [39]. Thus, concepts of corporate social responsibility [16], ethics and sustainability have received considerable attention [44].

In fact, as noted previously, the Australian government, knowing of this importance, and being aware of the progressive decline of public funding of Australian universities and the increasing international competition to which they are subject [41, 36] published a Green Paper (1987) to open the debate on the need for fundamental reform of higher education in Australia [41].

In this regard, Australian universities have been active, and often successful in finding alternative sources of income [32] and in the internationalization of universities as centers receiving foreign students. However, the higher education market has become a very competitive environment for international students [36] in which the reputation of the provider country and its educational institutions is an important factor in the selection of a destination for future training of many students.

Such has been the growth of international business in education that many governments have begun to see international studies as a critical factor in educational policy. Governments and institutions have responded by improving access to their fields of education for foreign students and internationalizing the curriculum [31]. This has meant that countries like the United Kingdom, Australia, New Zealand and the USA have been become important destinations for students who consider both a quality educational system and the international recognition of the degree obtained at these universities.

3 Empirical Research in Australian Universities

3.1 Sample Selection

Australian universities have been the objective of our work for several reasons: a) in recent decades, they have suffered reforms to the introduction of social responsibility criteria in both the education and in research, as has mentioned in the previous section of this research, b) as a result, Australian universities, with the rest of universities belonging to Anglo-Saxon countries, are reporting more information on social responsibility issues. To test this fact, we have undertaken an initial search using general search engines of the Internet such as Yahoo, Google and Terra and introduced, as the key concepts, "sustainability report", "Environmental Report", "environment", "ecology", "water", "recycling", "green building", "Biodiversity" or "non for profit" joined with the term "university". In this search, Australian universities were highlighted, at the international, as those that most often disclosed aspects of SR in their website or in their reports and/or SR pronouncements, and finally, in the annual report, traditionally used as a channel of communication and source of accountability [55].

So, we focus on Australian universities and, for our research work, we have been mindful of the full set of Australian universities according with Australian Vice Chancellors Committee (http://www.avcc.edu.au/). This selection resulted in a sample of a total of 39 universities in Australia.

3.2 Research Methodology

In order to analyze the extent to which sample universities are reporting online SR information, and bearing in mind the scarcity of prior research, this paper proposes an evaluation model by creating a series of indices that seek to examine the main aspects to take into account both the content and context in which SR information is disclosed.

In order to propose the evaluation model, this paper considers a number of prior studies relating to the disclosure of information on the Internet [13, 14, 43] and others related to the analysis of visibility [38], accessibility [34], usability [6], privacy and information content [28]. Particular analysis of this previous literature, has made our research work is mainly structured in two parts which are analyzed through the study of six indices, see Table 1.

The first part includes the content of the SR information disclosed by the universities in the sample. In order to study this content, we followed the GRI guidelines on the one hand, and also analyzed the SR Reports from different universities, along with a set of specific qualitative characteristics in the information which although they are of a general nature come from another sphere e.g. financial information. Thus, we established three different sections: General SR Information (GSRI); Specific SR Information (SSRI), in which we analyze a set of specific topics; and the Qualitative Characteristics of SR Information (QCSRI) as questions that must be considered in order to guarantee the quality of the SR information disclosed (See Table 1).

The second part of our empirical study refers to the context in which the SR information disclosure takes place and has been structured into three sections taking into account the characteristics of the usability, interaction, privacy and security of websites. On the one hand, this paper aims to analyze the usability of the website (USRI) i.e. the ease with which users can make use of websites in order to reach a specific goal. On the other hand, in order to obtain information from stakeholders, learn about their expectations and establish mechanisms for interaction with them, this paper analyzes those mechanisms that favor the participation of stakeholders (SKSRI) in the sample universities. Finally, this paper analyzes the level of privacy and security of websites (PSRI) that mainly aim to observe the capacity for personalization in the publication of SR information (See Table 2).

Table 1. General SR Reporting

GENERAL SR INFORMATION		$GSRI=\sum_{i=1}^{m} g_i$
Concept	**Items**	**Score**
G1. Statement of vision and strategy of the university on issues about social responsibility	a) If main SR commitments are disclosed. b) If the webpage or Sustainability Report includes a declaration on SR from the governing body.	0/0.5 based on the absence-presence of each item
G2. Information about profile of stakeholders	a) If the university webpage or the SR/SustainabilityReport identify the stakeholders. b) If there is specific information about the informational needs of each group of stakeholders.	0/0.5 based on the absence-presence of each item
G3. Centralized or decantralized disclosure of SR information by universities	a) If the disclosure of SR information is developed in a centralized way on the university webpage. b) If this disclosure is developed through dependent centres at said university.	0/0.5 based on the absence-presence of each item
G4. Data on performance indicators	a) economic indicators. b) social indicators. c) environmental indicators.	0/0.33 based on the absence- presence of each item
G5. Index of contents or a table to locate different elements of information about SR	Provides the reader with an index or a table to locate different SR elements.	0/1 based on the absence-presence of that item
SPECIFIC SR INFORMATION		$SSRI=\sum_{i=1}^{m} g_i$
Concept	**Items**	**Score**
S1. Energy	Information is disclosed about the installation of systems that save electricity such as movement sensors, incandescent lightbulbs or other alternative sources of energy.	0/1 based on the absence/presence of this item
S2. Buildings and grounds	Information is disclosed about criteria for construction, renovation and rehabilitation of existing buidlings in line with "green criteria".	0/1 based on the absence/presence of this item
S3. Purchasing management	Information is disclosed about the need to prioritize the purchase of reusable, ecological materials that require a minimum of packaging.	0/1 based on the absence/presence of this item
S4. Waste management and recycling	Information is disclosed about questions related to the promotion of the recycling of office material and solid waste providing recipients for articles such as paper, printer cartridges and batteries.	0/1 based on the absence/presence of this item
S5. Transportation	Information is disclosed about the creation of incentives for the university community to use public transport or alternative means of transport such as bicycles and bus.	0/1 based on the absence/presence of this item
S6. Food	Information is disclosed about fair trade and sustainable food through the provision of ecological products in campus cafés and shops.	0/1 based on the absence/presence of this item
S7. Academic	Information is disclosed about courses, seminars and conferences related to SR.	0/1 based on the absence/presence of this item
S8. Research	Information is disclosed about University research centres linked to SR.	0/1 based on the absence/presence of this item
QUALITATIVE CHARACTERISTICS OF SR INFORMATION		$QSRI=\sum_{i=1}^{m} g_i$
Concept	**Items**	**Score**
Q1. Completeness	It is possible to check and/or download the Report online.	0/1 based on the absence-presence of that item
Q2. Timeliness	It offers SR information more frequently than on a yearly basis (monthly, termly, etc).	0/1 based on the absence-presence of that item
Q3. Comparability	a) It is possible to compare information from two or more years. b) It offers comparative summaries on sustainable information provided by the university.	0/0.5 based on the absence-presence of each item
Q4. Understandability	a) It offers ratios and graphics to help to clarify the SR information included in the Reports. b) It incIncludes comments on the SR information provided.	0/0.5 based on the absence-presence of each item
Q5. Relevance	a) It provides technical SR reports made by the University b) It presents SR information in and ordered and classified manner.	0/0.5 based on the absence-presence of each item
Q6. Reliability	The information has been accredited.	0/1 based on the absence-presence of that item

Source: Own elaboration based on the GRI guidelines, adaptations about SR Reports published by universities, on the SR Reports published by universities.

Table 2. Specific SR Reporting

USABILITY		$USRI = \sum_{i=1}^{m} g_i$
Concept	**Items**	**Score**
U1. Reading and scanning	a) A specific section on the universities' websites for disclosing sustainability information exists. b) Electronic formats used to process the sustainability reporting: - htlm - pdf doc - xml o xbrl - xls c) Sustainability reporting is disclosed in different languages	0/0.33 based on the absence- presence of each item. REgarding the type of format (item b), the score of 0.33 is split in the following way: - htlm: 0.066 - pdf or doc: 0.066 - xml or xbrl: 0.099 - xls: 0.099
U2. Search	a) A basic search tool is included in the university website. b) An advanced search tool is included in the university website.	0/0.5 based on the absence-presence of each item
U3. Link characteristics	A system of hyperlinks for the information offered is provided.	0/1 based on the absence-presence of that item
U4.Structure of the web page	A web map showing the contents is available	0/1 based on the absence-presence of that item
U5. Characteristics of accesibility	All information provided on the website is freeware and it can be downloaded	0/1 based on the absence-presence of that item
STAKEHOLDER PARTICIPATION		$SKSRI = \sum_{i=1}^{m} g_i$
Concept	**Items**	**Score**
SK1. Characteristics of interactivity	a) A different e-mail address to the webmaster's is provided to request information or explanations. b) Personal contacts with responsible persons of the university for the information provided are supplied on the website c) The website has a mailing list to update information to those information users that apply this service	0/0.33 based on the absence-presence of each item
SK2. Forums/chats	a) Forums with general contents b) Forums related to SR or sustainability	0.5 if the online forum/chat used allows discussion of general subjects and 1 if there is a specific forum/chat used for SR subjects
SK3. Web 2.0 technology	a) Web 2.0 technology about the University in general b) Web 2.0 technology about aspects of SR or sustainability	0.5 if the use of Web 2.0 technology is aimed at general university subjects and 1 is the Web 2.0 technology is used for
SK4. Online surveys	a) Surveys not specific to SR b) Surveys specific to SR	0.5 of the university uses online surveys of a general nature and 1 if the university uses surveys about SR
SK5. Newsletter	a) General news b) Specific news about SR or sustainability	0.5 if the news disclosed by the university is of a general nature and 1 if it is SR news
PRIVACY AND SECURITY		$PSRI = \sum_{i=1}^{m} g_i$
Concept	**Items**	**Score**
P1. Data collection	The university collects specific data from the user	0/1 based on the absence-presence of that item
P2. Digital signature	A digital signature can be used	0/1 based on the absence-presence of that item
P3. Notificación de privacity policy	If there is notification of a privacy policy	0/1 based on the absence-presence of that item
P4. Access to private info	If there is restricted information	0/1 based on the absence-presence of that item
P5. Use of cookies	Use of techniques such as cookies that collect information on user access or behaviour on the webpage	0/1 based on the absence-presence of that item

Source: Own elaboration based on previous literature [28].

Regarding the score assigned to each one of the questions included in our proposal for the assessment of the disclosure of SR information in these universities and, based on the previous literature [18, 13], we opted for a binary dichotomous scoring system (0/1) in line with the absence or presence of each question on the website or in the Sustainability/SR Report in order to reduce the subjectivity in scoring system when there are no explicit rules for the assignment of the score for each one of the questions analyzed [30], giving the same value to each unit when the aspect analyzed is defined by various items [27] –see Tables 1 and 2 to observe the particular score for each item-.

In order to obtain the necessary data, between the months of April and May 2012, we proceeded to visit the websites of the different universities selected in order to conduct an exhaustive review and obtain the information necessary for our research. All of this process was carried out separately by the three authors of this paper in order to guarantee greater objectivity and the authors subsequently met to discuss the results and reach a consensus. If there were any discrepancies with significant variations the websites were examined again by all two authors.

4 Analysis of Results

According to the data in Table 3, in Section a) content of the SR information disclosed, the low values obtained by the universities in a.1) general SR information reveal the low level of commitment shown by the universities in the sample in terms of the disclosure of information that was analyzed (on average 1.31 out of 5). Of all of the items that make up this section, those which were least disclosed were those that refer to the profile of stakeholders and the indicators of execution in the different dimensions of SR (see Table 3).

Moreover, from the data obtained in section a.2) Specific SR Information, results confirm that the universities are currently disclosing specific SR Information since it is section that has obtained the highest scores according to the model proposed (Table 3).

Regarding the most widely disclosed questions, within the framework of Specific SR Information, we can highlight those items that refer to energy, management of residues and recycling, and in the academic sphere, whilst the least disclosed questions are those that refer to sustainable food/fair trade or the purchase of sustainable food. The results obtained, in line with previous research, see among others [23] and [26], show the weight that still has the environmental dimension in the design of university social responsibility.

Regarding the fulfillment of the qualitative characteristics of the information (section a.3), we observed that low scores were obtained for the content of the SR information disclosed (0.88 out of 6), showing the low priority that the universities in the sample currently give to these qualitative characteristics (Table 3).

Nevertheless, it must be highlighted the fact that very few of the universities analyzed offer annual SR reports on their websites and, therefore, the SR issues that are disclosed by universities mainly appear in the shape of technical summaries (not annual SR reports). These technical summaries are usually accompanied by ratios or graphs with comments that aim to make the information more comprehensible for

Table 3. Descriptive Statistics

a) CONTENT OF THE SR INFORMATION DISCLOSED ON THE WEBPAGES OR ON SUSTAINABILITY REPORTS	MIN	MAX	MEAN	MEDIAN	SD
a.1) General SR content	**0,00**	**3,49**	**1,31**	**1,00**	**1,02**
G1. Expression of the vision and strategy of the university in SR subjects	0,00	1,00	0,37	0,50	0,34
G2. Information on the profile of stakeholders	0,00	0,50	0,01	0,00	0,08
G3. Centralized or decentralized disclosure of SR information by Universities	0,00	1,00	0,41	0,50	0,25
G4. Data on performance indicators	0,00	0,99	0,14	0,00	0,28
G5. Index of contents or a table to locate different elements of SR information	0,00	1,00	0,38	0,00	0,49
a.2) Specific SR content	**0,00**	**7,00**	**2,85**	**3,00**	**2,29**
S1. Energy	0,00	1,00	0,44	0,00	0,50
S2. Buildings and grounds	0,00	1,00	0,28	0,00	0,46
S3. Purchasing management	0,00	1,00	0,21	0,00	0,41
S4. Waste management and recycling	0,00	1,00	0,56	1,00	0,50
S5. Transportation	0,00	1,00	0,38	0,00	0,49
S6. Food	0,00	1,00	0,03	0,00	0,16
S7. Academic	0,00	1,00	0,56	1,00	0,50
S8. Research	0,00	1,00	0,38	0,00	0,49
a.3) Qualitative characteristics of SR information	**0,00**	**4,00**	**0,88**	**0,50**	**1,01**
Q1. Completeness	0,00	1,00	0,03	0,00	0,16
Q2. Timeliness	0,00	1,00	0,05	0,00	0,22
Q3. Comparability	0,00	1,00	0,13	0,00	0,34
Q4. Understandability	0,00	1,00	0,18	0,00	0,35
Q5. Relevance	0,00	1,00	0,49	0,50	0,44
Q6. Reliability	0,00	1,00	0,03	0,00	0,16
b) CONTEXT OF THE SR INFORMATION ON THE WEBPAGES	**MIN**	**MAX**	**MEAN**	**MEDIAN**	**SD**
b.1) Usability	**1,63**	**4,13**	**3,33**	**3,57**	**0,64**
U1. Reading and scanning	0,00	0,46	0,13	0,13	0,09
U2. Search	0,00	1,00	0,67	0,50	0,29
U3. Link characteristics	0,00	1,00	0,97	1,00	0,16
U4.Structure of the web page	0,00	1,00	0,59	1,00	0,50
U5. Characteristics of accesibility	0,00	1,00	0,97	1,00	0,16
b.2) Stakeholders participation	**0,00**	**11,83**	**2,40**	**2,16**	**1,71**
SK1. Characteristics of interactivity	0,00	0,66	0,27	0,33	0,27
SK2. Forums or chats	0,00	0,50	0,36	0,50	0,23
SK3. Uses 2.0 Web technology (facebook, twitter...)	0,00	1,00	0,53	0,50	0,16
SK4. If there are online surveys on university matters	0,00	9,50	0,51	0,50	1,50
SK5. If there is a university newsletter	0,00	1,00	0,73	0,50	0,28
b.3) Privacy and security	**1,00**	**5,00**	**3,38**	**4,00**	**0,91**
P1. Data collection	0,00	1,00	0,62	1,00	0,49
P2. If a digital signature can be used	0,00	1,00	0,10	0,00	0,31
P3. If there is a notification of privacy policy	0,00	1,00	0,77	1,00	0,43
P4. Access to private info	0,00	1,00	0,90	1,00	0,31
P5. Use of cookies	1,00	1,00	1,00	1,00	0,00

Source: Own elaboration.

external users. Furthermore, in the case of publishing annual SR reports, the information contained is normally neither timely nor comparable, since there is no data provided for periods of time of less than one year nor is there any reference to previous years, thus making it impossible to establish comparisons.

Therefore, these results showed that, unlike [35], the universities in the sample are not communicating their socially sustainable actions and policies, through sustainability or social responsibility reports, which makes it difficult to establish a benchmarking process between them. Moreover, there are very few universities in the sample that communicate information that has previously been certified, accredited or audited (Table 3).

Regarding section b), context of the SR information disclosed, the scores related to b.1) usability reaches an average of 3.33 out of 5 with a very low standard deviation (0.64). As can be observed in Table 3, the scores of the different universities in the sample are very similar and the bulk of the score is based on items related to website design, since the universities have included instruments that facilitate web

browsing, such as search engines and hyperlinks that help users to examine the information in greater depth. These search tools partially solve the general absence of a structured website map.

It is also worthwhile noting the high score obtained by all of the universities in section U5 (characteristic of accessibility), showing that the SR information on the website is accessible by being both free and easy to download. Nevertheless, there are very few universities that have a specific section for SR information. Moreover, the SR information is only available in one language, which limits the number of people who can use and understand it, and the formats in which this information is presented are not easy to adapt since universities have opted to use HTLM or PDF formats.

Section b.2, participation of stakeholders, is the one that obtained the lowest average score (2.40). The average scores were practically the same and were mainly based on the score assigned to the items that refer to news on the website and the use of web 2.0 technology. Nevertheless, universities did not promote communication and participation with stakeholders through other channels e.g. e-mail addresses other than that of the webmaster that would help users to find specific SR information or to request additional information not included on the website and even to update this information through the creation of a mailing list. Other tools for participation that were not being promoted by universities were those that used forums or chat and online surveys.

The results in this area confirm those obtained in previous studies such as [40] or [10], in which the authors highlighted the lack of market orientation of universities in their relationship with stakeholders, along with the weak role played by shareholders in the government of these institutions. However, the results are different from those of other studies that stated that universities are strengthening their relationships with their stakeholders by adapting to their expectations and increasing the influence and participation of stakeholders in university management [17, 33].

Regarding section b.3), privacy and security of websites, this is the item that obtained the highest average score (3.38) out of all of the questions analyzed in section b) context of the SR information. The highest score is based on the items that refer to the development of the notification of privacy policies, restriction of access to areas containing personal information and the use of cookies. On the contrary, the least developed item is the use of digital signatures that would allow for greater security and data collection systems. Universities in the sample, as shown in Table 3 show a similar behavior in the different items in this section (standard deviation 0.91).

5 Conclusions

Currently, society is demanding greater involvement of universities in the social environment in which they operate [53]. In fact, in recent years, various stakeholders are pressuring universities to improve their accountability by incorporating non-financial information in their communication policies, trying to see how the universities are fulfilling their teaching and research objectives and how these functions work, avoiding negative impacts to the environment, and being sustainable with the economic, social and environmental context around them [3]. Moreover, they are demanding that, beyond not harming the environment in which they work, can

provide solutions to it and serve as engines of development of the community in which they operate [3; 33].

These demands have increased interest by universities in the social commitment, both in the implementation of criteria for SR in its own governance and in its teaching and research trying, at the same time, increasing the quantity and quality of information disclosed incorporating aspects of SR information about universities [3]. In Australian universities these actions have been strongly supported by the government, conscious of the need to reform the Australian university system to incorporate social responsibility criteria in both education and research [4].

The development of new technologies over the last few decades and, in particular, the Internet, could be used as a means to identify and channel the expectations of different stakeholders in order to establish dialogue as well improving informational transparency. This would also increase the extent to which universities comply with their duty to be accountable to their stakeholders. Nevertheless, the results of the empirical research in this paper show that, far from using new technology to improve their informational transparency and accountability, universities are wasting this opportunity both in terms of disclosing SR information and in terms of involving and interacting with stakeholders regarding the management of the university on SR concerns.

Thus, although results obtained in previous studies argue that universities are becoming key players in promoting socially responsible policies [35], the results of this paper indicate that Australian universities are little committed in terms of online SR information disclosure, do not offer complete SR information on university's websites, or reporting of SR or Sustainability Reports on the Internet. This prevents the effective performance of their duty of accountability to the requirements of this type of information by stakeholders, as well as the knowledge of the activities that universities are undertaking in this field, not allowing the possibility of benchmarking process between universities that could involve aspects of continuous improvement and competitiveness among them.

Furthermore, the SR information that is being disclosed by universities gives most importance to environmental questions as these have the greatest impact at universities [23, 26]. However, this type of information is not usually easy to find on the websites. In general, this type of information is normally dispersed and is mainly included in technical reports and not in annual SR reports despite the emphasis on the need for these reports by some authors [3]. In fact, the issuance of SR reports could serve as a vehicle for communication of universities to be accountable for their commitments with the environment around them. Another option, that it currently is undertaking in the field of private companies [22], could be to integrate all financial and non-financial information in a single report avoiding in this way, the dispersion of information.

Also, although universities could make use of the advantages provided by ICT to improve participation and interaction with stakeholders, our results indicate that, currently, do not seem very concerned about it, at least as far as aspects of SR concerned, in contrast to prior research which has been observed that universities are increasing the influence and participation with stakeholders [3, 17, 33]. The results of our study suggest that universities are not fully committed to the environment around them as they are not showing interest in collecting, with the support of new

technologies, of the opinions and suggestions of stakeholders when designing and implementing their SR policy. In fact, the SR character performances seem to attend more to reasons of legitimacy and image of the university than to the needs and requirements of the community in which they are inserted [20, 21], forgetting, similarly, that the information disclosed by the universities is a key factor in the decision making of students [45].

In conclusion, Australian universities are disclosing little online SR information perhaps only, as noted previously, for purposes of legitimacy and image such as other previous studies have shown [20, 21] but not with the ultimate goal of satisfying the requirements of the stakeholders in the context of fulfilling their duty of accountability. In fact, the universities are not analyzing the needs of stakeholders despite the great influence they have in any organization [29] and the way in how to communicate with them and present the information is still insufficient.

Consequently, Australian universities, as main players, must have an essential role in the SR issues through the incorporation of SR principles in all fields of teaching and research. This way, Australian universities should improve the SR disclosure, which could serve to differentiate from others in a competitive environment. In this regard, it might be interesting, for future research, to put more emphasis on both key factors that could enhance the online disclosure of SR information, and on the analysis and design of management models to establish efficient means to meet stakeholders' needs on SR information, which could make this aspect to be a decisive issue in the prestige of universities.

Acknowledgments. This research was carried out with financial support from the Regional Government of Andalusia (Spain), Department of Innovation, Science and Enterprise (Research project number P11-SEJ-7700) and from the Spanish National R&D Plan through research project number ECO2010-20522-ECON (Ministry of Science and Innovation).

References

1. Abbott, M., Doucouliagos, C.: The efficiency of Australian universities: a data envelopment analysis. Economics of Education Review 22, 89–97 (2003)
2. ACTS, http://www.acts.asn.au
3. Al-Khatar, K., Naser, K.: User's perceptions of corporate social responsibility and accountability: Evidence from an emerging economy. Managerial Auditing Journal 18(6/7), 538 (2003)
4. Australian Government. Department of Education, Employment and Workplace, http://www.deewr.gov.au/HigherEducation/Documents/TransformingAusHigherED.pdf
5. Australian Learning and Teaching Council. Sustainability website, http://www.altc.edu.au/2010May-sustainability-website
6. Badre, A.: Shaping Web Usability: Interaction Design in Context. Addison-Wesley Publishers B.V., Boston (2002)
7. Bampton, R., Maclagan, P.: Why teach ethics to accounting students? A response to the sceptics. Business Ethics 14(3), 290 (2005)

8. Banks, W., Fisher, J., Nelson, M.: University accountability in England, Wales, and Northern Ireland: 1992-1994. Journal of International Accounting Auditing and Taxation 6(2), 211 (1997)
9. Bekessy, S., Burgman, M., Wright, T., Filho, W., Smith, M.: Universities and sustainability. Tela Papers, No. 11 (2003)
10. Benneworth, P., Jongbloed, V.: Who matters to universities? A stakeholder perspective on humanities, arts and social sciences valorization. Higher Education 59(5), 567–588 (2010)
11. Borins, S.: On the frontiers of electronic governance: a report on the United States and Canada. International Review of Administrative Science 68(2), 199–211 (2002)
12. Bridges, C.M., Wilhelm, W.B.: Going beyond green: The "why and how" of integrating sustainability into the marketing curriculum. Journal of Marketing Education 30(1), 33–46 (2008)
13. Caba, P.C., López, H.A.M., Rodríguez, B.M.P.: Citizens' access to on-line governmental financial information: Practices in the European Union countries. Government Information Quarterly 22(2), 258–276 (2005)
14. Caba, P.C., Rodríguez, B.M.P., López, H.A.M.: e-Government process and incentives for online public financial information. Online Information Review 32(3), 379 (2008)
15. Carrington, R., Coelli, T.I.M., Rao, D.S.P.: The performance of Australian universities: Conceptual issues and preliminary results. Economic Papers 24(2), 145–163 (2005)
16. Carroll, A.B.: Corporate social responsibility: Evolution of a definitional construct. Business and Society 38, 268–295 (1999)
17. Castelo, B.M., Lima, R.L.: Factors Influencing Social Responsibility Disclosure by Portuguese Companies. Journal of Business Ethics 83(4), 685–701 (2008)
18. Cooke, T.E.: Voluntary Disclosure by Swedish Companies. Journal of International Financial Management and Accounting 1, 171–195 (1989)
19. Debreceny, R., Ellis, A.: Managing student access to university information networks - the Australian experience. Campus-Wide Information Systems 15(2), 48–53 (1998)
20. Deegan, C.: The Legitimising Effect of Social and Environmental Disclosures. A Theoretical Foundation. Accounting, Auditing and Accountability Journal 15(3), 282–311 (2002)
21. Deephouse, D.L., Carter, S.M.: An Examination of Differences between Organizational Legitimacy and Organizational Reputation. Journal of Management Studies 42(2), 329–360 (2005)
22. Eccles, R.G., Krzus, M.P.: One Report: Integrated Reporting for a Sustainable Strategy, 1st edn., 256 pages. Wiley (2010) ISBN-10: 0470587512
23. Ferrer-Balas, D., Adachi, J., Banas, S., Davidson, C.I., Hoshikoshi, A., Mishra, A., Motodoa, Y., Onga, M., Ostwald, M.: An international comparative analysis of sustainability transformation across seven universities. International Journal of Sustainability in Higher Education 9(3), 295–316 (2008)
24. Glasser, H., Calder, W., Fadeeva, Z.: Definition: research in higher education for sustainability. Halifax Consultation, Halifax (2005)
25. Gordon, T., Fischer, M., Malone, D., Tower, G.: A comparative empirical examination of extent of disclosure by private and public colleges and universities in the United States. Journal of Accounting and Public Policy 21(3), 235 (2002)
26. Hammond, C., Churchman, D.: Sustaining academic life: A case for applying principles of social sustainability to the academic profession. International Journal of Sustainability in Higher Education 9(3), 235–245 (2008)
27. Ho, P., Tower, G., Barako, D.: Improving governance leads to improved corporate communication. Corporate Ownership & Control 5, 26–33 (2008)

28. Holzer, M., Manoharan, A.: Global Trends in Municipal E-Government: An Online Assessment of Worldwide Municipal Web Portals (Marc Holzer in A. Agarwal & V.V. Ramana. Computer Society of India, India) (2007)
29. Johnson, G., Scholes, K.: Whittington: Exploring Corporate Strategy: Enhanced Media Edition: Text and Cases. Enhanced Media Text & Case (2006)
30. Jones, K., Alabaster, T., Walton, J.: Virtual Environments for Environmental Reporting. Greener Management International 21, 121–137 (1998)
31. Kameoka, Y.: The internationalisation of higher education. OECD Observer 202, 34–36 (1996)
32. Karmel, P.: Public Policy and Higher Education. Australian Journal of Management 26(special issue) (2001)
33. Krazimierz, M.: Redefining external stakeholders in Nordic Higher Education. Tertiary Education and Management 16(1), 45–60 (2010)
34. Lawrence, S., Giles, C.L.: Accessibility of information on the web. Nature 8, 107–109 (1999)
35. Lukman, R., Glavič, P.: What are the key elements of a sustainable university? Clean Technologies and Environmental Policy 9(2), 103–114 (2007)
36. Mazzarol, T.W., Soutar, G.N.: Australian educational institutions' international markets: A correspondence analysis. International Journal of Educational Management 22(3), 229–238 (2008)
37. Mellé Hernández, M.: Social Responsibility in the Public Sector. Ekonomiaz 65, 84–107 (2007)
38. Middleton, I., McConnell, M., Davidson, G.: Presenting a model for the structure and content of a university World Wide Web site. Journal of Information Science 25(3), 219–227 (1999)
39. Nagy, J., Robb, A.: Can universities be good corporate citizens? Critical Perspectives on Accounting 19(8), 1414–1430 (2008)
40. Pavičić, J., Alfirević, N., Mihanović, Z.: Market orientation in managing relationships with multiple constituencies of Croatian higher education. Higher Education 57(2), 191–207 (2009)
41. Pratt, G., Poole, D.: Globalisation and Australian universities: policies and impacts. The International Journal of Public Sector Management 12(6), 533–544 (1999)
42. Quiggin, J.: Resolving the university crisis, Submission to Senate Employment, Workplace Relations, Small Business and Education Committee into the capacity of public universities to meet Australia's higher education needs (2001)
43. Rodríguez, B.M.P.: Evaluating Corporate Environmental Reporting on the Internet: The Utility and Resource Industries in Spain. Business and Society 48(2), 179–205 (2009)
44. Rundle-Thiele, S.R., Wymer, W.: Stand-Alone Ethics, Social Responsibility, and Sustainability Course Requirements: A Snapshot From Australia and New Zealand. Journal of Marketing Education 32(1) (2010)
45. Schimmel, K., Motley, D., Racic, S., Marco, G., Eschenfelder, M.: The importance of university web pages in selecting a higher education institution. Research in Higher Education Journal 9, 1–16 (2010)
46. Segalàs, J., Ferrer-Balas, D., Svanström, M., Lundqvist, U., Mulder, K.F.: What has to be learnt for sustainability? A comparison of bachelor engineering education competences at three European universities. Sustainability Science 4(1), 17 (2009)
47. Serap, A.M.G., Eker, T.: Corporate identity of a socially responsible university- a case from the Turkish higher education sector. Journal of Business Ethics 76(1), 55–68 (2007)

48. Stephens, J.C., Hernandez, M.E., Román, M., Graham, A.C., Scholz, R.W.: Higher education as a change agent for sustainability in different cultures and contexts. International Journal of Sustainability in Higher Education 9(3), 317–338 (2008)
49. Tilbury, D., Crawley, C., Berry, F.: Education about and for sustainability in Australian business schools', Report prepared by the Australian Research Institute in Education for Sustainability (ARIES) and Arup Sustainability for the Australian Government Department of the Environment and Heritage (2004), http://www.aries.mq.edu.au/pdf/MBA_Brochure.pdf
50. Tilbury, D., Keogh, A., Leighton, A., Kent, J.: A National Review of Environmental Education and its Contribution to Sustainability in Australia: Further and Higher Education. Australian Government Department of the Environment and Heritage and Australian Research Institute in Education for Sustainability (AIRES), Canberra (2005), http://www.aries.mq.edu.au/pdf/Volume5_brochure.pdf
51. UNESCO. Institute of Statistics, http://stats.uis.unesco.org
52. USLF. History, http://www.ulsf.org/programs_talloires_td.html
53. Vallaeys, F.: Brief theoretical framework of university social responsibility. Pontificia Universidad Católica de Perú, http://blog.pucp.edu.pe/item/4880/breve-marco-teorico-de-rsu
54. Von der Heidt, T., Lamberton, G.: Sustainability in the undergraduate and postgraduate business curriculum of a regional university: A critical perspective. Journal of Management and Organization 17(5), 670–690 (2011)
55. Vuontisjärvi, T.: Corporate Social Reporting in the European Context and Human Resource Disclosures: An Analysis of Finnish Companies. Journal of Business Ethics 69(4), 331 (2006)

Information Architectures Definition – A Case Study in a Portuguese Local Public Administration Organization

Filipe Sá[1] and Álvaro Rocha[2]

[1] Câmara Municipal de Penacova, Largo Alberto Leitão, 5, 3360-341 Penacova, Portugal
filipe@cm-penacova.pt
[2] LIACC, Universidade do Porto, R. Campo Alegre 1021, 4169-007 Porto, Portugal
amrrocha@gmail.com

Abstract. An Information Architecture is the basis for a comprehensive and integrated information system of any organization, capable of providing answers that meet the challenges posed by the contemporary competitive environment. In this paper we identify representative methodologies used in Information Architecture definitions, and we present an Information Architecture Planning case study for a Local Public Administration organization, where we implement an Adapted BSP Methodology that is sufficiently flexible and suited to the goals we propose.

Keywords: Information Systems Planning, Information Architecture, Information Systems Integration, E-Government.

1 Introduction

In a globalized market, with the consequent increase of competitiveness and the growing level of demand concerning products and services, practically every aspect of an organization impacts its competitive positioning, and this is particularly true for the efficiency of its Information System (IS) [9, 12, 13, 22].

Consequently, Local Public Administrations are currently facing a challenge of administrative modernization, where the purpose is to bring services and town residents closer together, together with a constant need process dematerialisation [2, 24, 15].

Penacova Municipality, located in the District of Coimbra, Portugal, is no exception to this need. The available information is increasingly larger and its consultation must be quick and effective, in order to lead to the most efficient decisions and actions.

In this context, the purpose of this work was to develop a study that would define and validate an Information Architecture for the Private Construction Works Department of Penacova Municipality, capable of effectively answering to its information needs.

For that reason, our study focused on existent concepts and methodologies, in order to define a cohesive Information Architecture based on an academic work that would be capable of sustaining a genuinely useful final product.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 399–410.
DOI: 10.1007/978-3-642-36981-0_37

Accordingly, we outlined the following goals in the present work:

- Analyzing methodologies that were representative of the Information Architecture Planning;
- Selecting and implementing a methodology in a real case scenario;
- Proposing and validating an Information Architecture for the Private Construction Works Department of Penacova Municipality;
- Reflecting about the advantages and disadvantages of the implemented methodology.

In the following sections we present the methodology adopted in this study, we identify the methodologies considered representative for the Information Architecture Planning and we present a case study involving the development of an Information Architecture for the Private Construction Works Department of Penacova Municipality. Finally, we will discuss the results, draw a number of conclusions and define the direction for a future work.

2 Methodology of the Study

In a first stage, we developed a review of the available literature, examining in detail a sample of books, essays, thesis and scientific papers published in the field, which allowed us to select and better understand the Information Architecture Planning methodology implemented in our case study.

Maintaining this approach we followed the case study method, frequently used in both quantitative and qualitative investigation works [7, 8], that oriented the implementation of the Adapted BSP Methodology presented by Amaral & Varajão [3], in the Private Construction Works Department of Penacova Municipality.

Based on the collected and examined elements, we proposed and validated an Information Architecture encompassing the complete and integrated development of an Information System that supported the multiple information needs of this Penacova Municipality Department.

Finally, we discussed the results and reflected upon the advantages and disadvantages of the implemented methodology.

3 Information Architecture Planning Methodologies

The Information Architecture of any organization involves the definition of relationships between business processes and data classes. This enables the evaluation of data sharing inside the organization. Additionally, the Information Architecture provides the basis for resource management and tactical planning, allowing for an orderly implementation of the Information Architecture.

The focus of our investigation involved the identification of methodologies that were capable of aiding the Information Architecture Planning of an organization. Consequently, were selected and studied the BSP (Business Systems Planning) Methodology [10], the Adapted BSP Methodology [3], Zachman's Framework

[25, 26, 27], the Federal Enterprise Architecture Framework [21] and the Enterprise Architecture Planning Methodology [5, 18].

These were the methodologies considered representative for the Information Architecture definition, based on the extent that they were mentioned, studied and selected and/or implemented in several academic works [e.g., 1, 4, 6, 9, 11, 12, 14, 16, 17, 20, 23].

In the present case study we selected and decided to implement the Adapted BSP Methodology developed by Amaral & Varajão [3].

Our choice was based on the following factors:

- It was substantially documented;
- It was inspired by the original BSP Methodology but offered new details and a different organization of activities, suited to the present context;
- It met the critical key-factors underlying the successful planning, development and implementation of an Information Architecture;
- Its different steps were easily adapted to fit the reality of Private Construction Works Department in Penacova Municipality;
- The matrices describing the Information Architecture enabled a clear and effective analysis of processes and data classes, as well as an information flow analysis, promoting a greater understanding of the study as a whole;

The Application/Data Classes Matrix, the Application/Processes Matrix and Application/ Organization Matrix, led to a clear and unmistakable identification of the current and prospective IS support to the Information Architecture of the Private Construction Works Department in Penacova Municipality.

4 Case Study: The Private Construction Works Department of Penacova Municipality

In this section of the paper we describe a study developed in the Private Construction Works Department of Penacova Municipality. This study included an Information Architecture Planning based on the Adapted BSP Methodology, developed by Amaral & Varajão [3]. This methodology was adapted to the reality of the context and the size of the study.

4.1 Characterization of the Private Construction Works Department

Penacova Municipality belongs to the District of Coimbra. The Municipality covers an area 220 Km^2 and has 15251 inhabitants – 2011 Census – spread over 11 civil parishes.

Penacova Municipality is, therefore, a Local Public Administration organization. By the end of 2010 it held 149 employees. Its Private Construction Works Department belongs to the Environment, Urban Services and Municipal Construction Works Department, who may delegate tasks to Town Councillors or someone from its office. It is composed by the following organic subunits: Service Planning; Spatial

Planning; Urban and Construction Works Management; Municipal Inspection Services; and Administrative Department. The General Division has 74 employees, 17 of which working full time in the Private Construction Works Department.

Generally speaking, the purpose of this Private Construction Works Department is to licence and regulate all the construction works developed within Penacova Municipality. Some of the processes result in construction work projects, subdivision plans, certificates, correspondence, issuance of permits, fee collection, water-related processes, sanitation, etc.

Despite being an internal department, the treatment of its processes implies an interaction with other internal departments, such as, for instance, the Treasury or the Accounting departments. There are also external interactions, both with town residents and other entities, involving licensing, authorizations or requests for opinions.

This Department is responsible for receiving and running all matters connected with private construction works.

4.2 Implementing the Adapted BSP Methodology

In this section we describe how the Adapted BSP Methodology, developed by Amaral & Varajão [3], was implemented with the purpose of creating an Information Architecture Planning for the Private Construction Works Department of Penacova Municipality. The complete absence of a study or any official ISP document involving this Municipality Department, and also the need to rethink its entire IS, justified this implementation.

Consequently, we implemented the steps of the selected methodology, adjusting them to the reality of the object of study and to the scope of our investigation. The scope and the objective of the study were thus defined, as well as the project team and the logistics. We identified the information to be collected, concerning human resources, financial resources, services and clients. We identified the installed hardware and software, and the information systems. We identified the mission, vision strategic goals and objectives of the Private Construction Works Department of Penacova Municipality. We also defined its entities. Based on the meetings we held with the Head of the Construction Works Department and the President of the City Council, we identified the following entities: President; Head of the Department; Technicians; Administrative Staff; Town Supervisor; Inspection Committee; and Citizens. This group of entities is responsible for all the elements that intervene in the processes of the Department. Finally, we defined the processes and the data classes.

4.2.1 Processes

In the words of Amaral & Varajão [3] the organization processes can be defined as "logically connected decision groups" or "the activities that are necessary to manage the resources of an organization". Still according to the same authors, defining the organizational processes will lead to:

- An IS that is independent of the organizational structure;
- An understanding of how the organization achieves its mission;
- A basis for the Information Architecture Planning, that identifies its scope, renders it modular and determines its development priorities;
- A basis for the identification of key-requirements pertaining to data.

Bearing in mind the rules for the identification of processes, based on the meetings we held, we drafted a first scheme containing all the procedures and existing processes in the Private Construction Works Department. After this survey the processes were organized in a "Dictionary of Processes". Finally, the processes were validated.

We identified and described 37 processes. 'Prepare Plan of Activities' (P1), for instance, was one of the processes we identified, and its definition was: "*Set of actions and decisions concerning the preparation of a Plan of Activities for the Private Construction Works Department, defining goals, responsibilities and identifying the means to achieve it.*"

4.2.2 Data Classes

Amaral & Varajão [3] state: "*A data class is a cluster of data pertaining to aspects (or entities) that are relevant for the information. Data classes should represent data that needs to be available to complete the activities of the organization*". Based on this assumption, and after the dictionary of processes had been completed, we were able to identity and define the existing data classes in the Department and prepare a dictionary of data classes.

We identified and described 42 data classes. "Site Journal" (C30), for instance, was one of the data classes we identified, and its definition was: "*Document containing the entire history of the construction work, including information of the responsible technicians, daily summaries of the construction works progress, amendments to the project, inspections from Town Supervisors and inspection-related files*".

After identifying and describing the dictionary of processes and the dictionary of data classes, we determined which data was created and used by each process. Figure 1 illustrates the process P1 (Prepare Plan of Activities).

P1 – Prepare Plan of Activities	
Create	**Use**
C1 – Activities Plan	C3 - Goals Sheet C5 – Holidays Sheet C4 – HR Assessment Form

Fig. 1. Creation and use of data classes by process

4.2.3 Defining the Information Architecture

After identifying, describing and validating each process and the respective data classes, we needed to establish a connection amongst them in order to define an Information Architecture that allowed us to assess the situation of the organization and build a structured diagnosis of the existing, as well as the future, constraints and difficulties.

Accordingly, we built a matrix displaying the processes on the vertical axis and the data classes on the horizontal axis. In the cells we used the 'C' (create) letter to designate the data classes created by processes, and the 'U' (use) letter to designate de processes that used data classes.

In order to determine the Information Architecture flow chart we organized the processes by grouping those that shared a great deal of data. Afterwards, we reorganized data classes as follows: the data class that was closest to the process axis was created by the first listed process, the next one (the following) by the second process, and so on.

After reorganizing the rows and the columns in the matrix we were able to define groups of processes. When those groups were identified, we were able to identify the data flow amongst them, based on the data classes that each process used.

Figure 2 displays the Processes/Data Classes Matrix, and the groups of processes were named according to their role:

- Administration – Group of processes and data classes concerning the Construction Works Department management and policy;
- Service - Group of processes and data classes concerning Costumer Service and preliminary processing;
- Sanitation – Group of processes and data classes concerning the preparation of a preliminary construction work process analysis by the Administrative Service;
- Licensing - Group of processes and data classes concerning the technical analysis of a construction work process;
- Permit - Group of processes and data classes concerning the issuing of a permit;
- Construction Work Monitoring - Group of processes and data classes concerning the monitoring and the inspection of the construction work carried out by a town resident;
- Utilization Permit - Group of processes and data classes concerning the issuing of an utilization permit;
- Archive - Group of processes and data classes concerning the management of the construction works archive.
- The data flow amongst the groups of processes is displayed in Figure 2. These flows happened each time a process belonging to a group of processes used a data class created by a different process belonging to another group of processes.

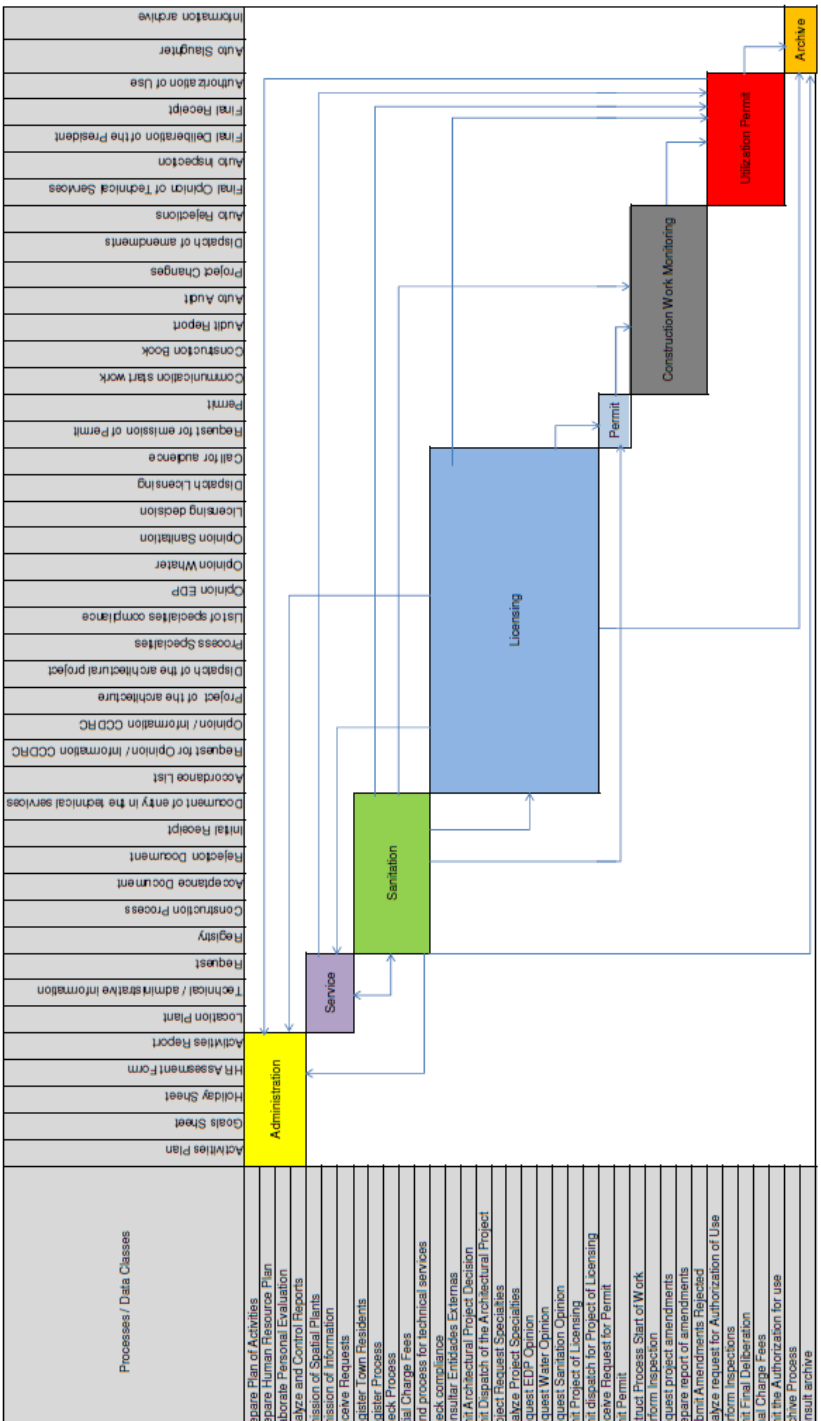


Fig. 2. Creation and use of data classes by process

4.2.4 IT Support to the Information Architecture Analysis

After the new Information Architecture was defined we analyzed the support it currently receives from computer applications/ Information Technologies (IT). To do so we prepared three matrices: 1) Applications/Entities Matrix (Figure 3), displaying the computer applications that support users from the Construction Works Department; 2) Applications/Processes Matrix, displaying the applications that support existing processes; 3) Applications/Data Classes Matrix, displaying the applications that support data classes, in order to understand which applications are currently automated and which applications maintain the data.

Appications/Entities	President	Department Head	Technical Services	Administrative Services	Supervision	Inspections Comission	Town Residents
Operations Processing System	S	S	S	S	S	S	
Fees System				S			
Document Management System	S	S	S	S	S	S	P
Staff Management System	S	S					
Emission Plants in Web		S	S	S	S	S	P
Municipal Portal		S	S	S			
Autocad		S	S	S			
Spreadsheet		S	S	S	S	S	S
Word Processing	S	S	S	S	S	S	S
Email	S	S	S	S	S	S	S

Fig. 3. Applications/Entities Matrix

4.2.5 Problems

After validating the previous steps we needed to identify and define the main problems emerging from the analysis of the computer applications support to the defined Information Architecture, and consequently to the processes and the data classes.

In the course of this analysis, which included a number of meetings with the team members, and during the interviewing we identified and validated the main problems. To summarize them we created a table with the following columns: group of processes, cause, problem, significance, originating process, originating data class and proposed solution.

Despite emerging in different processes, most of the problems revealed a common source. The main problems were:

1) **Lack of available information in the system (paper-based data in some cases)** – The Construction Works Department mainly resorts to an OPS (Operations Processing System) application to deal with the construction work processes and a DMS (Document Management System) application for mail registration. These

two systems should be completely integrated. At the same time, both the OPS and the DMS should enable the electronic scanning of every document, giving way to fully electronic processing and process analysis.

2) **Excess of useless information** – Without the electronic scanning the applications cannot track the information, leading to the storage of information that is irrelevant for the process analysis.
3) **Communication gaps with the town residents** – The communication with the town residents is still paper-based, which interferes with deadlines and an accurate document processing. The OPS and DMS applications should allow the electronic notification and communication with town residents.
4) **Several applications use the same information, without being shared or scanned** – Again, the OPS and the DMS must be integrated to access the same information.
5) **No control over external requests** – The current applications do not support automated requests for information or opinions from external entities. Since these requests are a part of licensing processes the OPS must be altered to allow automated requests.
6) **Access to information from the outside** – Supervising teams and inspection committees, when performing external services, need to have access to updated information from the construction work processes under supervision/inspection, and this calls for the implementation of a remote and real time system that allows access to existing processes in the OPS system.

4.2.6 Priorities and Recommendations

Considering the collected and validated information, the identified problems, and the validation of current needs, defining development priorities and recommendations becomes mandatory. In order to identify the priorities we adopted a method that involves the identification and clustering of criteria into four categories (Figure 4): Potential benefits; Impact on the organization; Probability of success; and Demand. In each category the relevance was rated in a scale of 1 to 10. The sum of the ratings determined the best computer application development sequence.

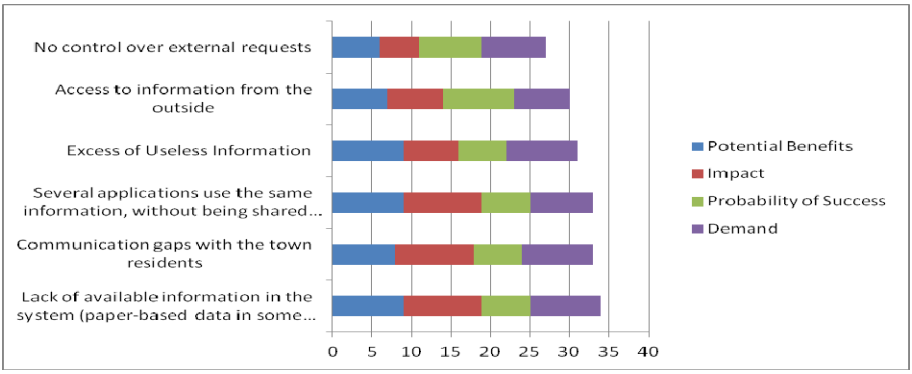


Fig. 4. Defining Development Priorities

Considering the previous analysis, the list of priorities for future development involves solving the following issues:

1) Lack of available information in the system (paper-based data in some cases);
2) Communication gaps with the town residents;
3) Several applications use the same data, without being shared or scanned;
4) Excess of useless information;
5) Access to information from the outside;
6) There is no control over external requests.

Despite the definition of a preliminary list of development priorities, this list should be subjected to a new evaluation after each implementation. For instance, the implementation of the first item (Lack of available information in the system) must be followed by a new identification of development priorities that includes the remaining items.

5 Discussion and Conclusion

5.1 Contributions

The main purpose of this work was to plan an Information Architecture for the Private Construction Works Department of Penacova Municipality. Our goal was motivated by a strategic decision from Penacova Municipality, of conforming the IS to the new demands of town residents - who needed a greater and quicker access to information - and legal demands for increasingly effective and efficient answers.

However, we observed that one of the greatest problems faced by Municipalities is the misuse of the available computer systems, as well as their lack of alignment with implemented and intended policies. Another problem has to do with the vision of the political power when faced with a number of changes. Change is frequently perceived as an obstacle, as opposed to a means for success.

Fortunately, organizations and new human resources departments are starting to perceive IS and IT not only as technology but as a means to achieve their goals and strategies. This change allowed us to involve a small team from Penacova Municipality in the implementation of the Adapted BSP Methodology, developed by Amaral & Varajão [3], in the Private Construction Works Department.

As such, specific goals have reached significant degree of specialization inside information systems, and that is why we performed a literature review involving multiple themes in this field, and several Information Architecture Planning methodologies.

By defining the Information Architecture we were able to characterize the information system and verify the existing computer applications as well as the support they provide to the implemented Information Architecture, and to identify and define eight groups of processes (Administration, Services, Sanitation, Licensing, Permits, Construction Works Monitoring, Utilization Permit and Archive) and forty - three data classes.

In a later stage, we identified the main problems as well as the development priorities and recommendations for the future. The excess of useless information and the lack of an integrated document management system in Penacova Municipality are the main obstacles to the Information Architecture implementation and improvement.

Regarding the used methodology, the Adapted BSP Methodology was sufficiently flexible and suited the context in question. In short, we reached all goals we proposed when we started this investigation.

5.2 Limitations and Possibilities of Future Works

Our work does not end with the results presented in this paper, which is a part of a preliminary information system planning for Penacova Municipality, particularly the Private Construction Works Department. This work can be improved and completed, should the Municipality be interested in exploring the full potential of the current information systems, in a comprehensive and groundbreaking manner.

Despite being a tangible work, time restrains and other limitations made it impossible to explore in greater detail the multiple areas developed, and in a subsequent period we intend to improve and study in greater detail the final steps of the selected methodology.

During the final implementation step we will negotiate a solution and confirm if it meets the goals defined by the Public Construction Works Department of this Municipality.

Should the Municipality be unable to develop a new IS capable of supporting the new information architecture in a comprehensive manner, we suggest that the existing applications are made available for change in a near future, especially the OPS and the DMS applications, according to the new Information Architecture. The Municipality should therefore contact the company that developed the applications to request their integration. The DMS should be improved in order to allow the electronic scanning of documents, the treatment and the control of processes and the communication with external bodies.

References

1. Abbas, A., Mir, S., Fereydon, S.: A Method for Benchmarking Application in the Enterprise Architecture Planning Process Based on Federal Enterprise Architecture Framework, Shahid Beheshti University, Iran. Computer Engineering Department (2010)
2. Almeida, F.B.: e-Government – Análise e Avaliação dos Municípios no Distrito de Viseu. Dissertação de Mestrado, Instituto Politécnico de Viseu, Viseu, Portugal (2011)
3. Amaral, L., Varajão, J.: Planeamento de Sistemas de Informação, 4th edn. FCA - Editora de Informática, Lisboa (2007)
4. Casagrande, N.G.: Metodologia para Modelagem de Arquitetura de Informação estratégica para pequenas empresas: Uma aplicação no setor de turismo rural. Tese de Doutorado em Administração, Universidade Federal de Santa Catarina, Florianópolis, Brasil (2005)
5. CIOC, Federal Enterprise Architecture Framework. Version 1.0. Chief Information Officers Council, USA (2001)
6. Costa, P.: A função Produção de SI/TI – Modelo Informacional. Dissertação de Mestrado, Universidade do Minho, Guimarães, Portugal (2002)
7. Höst, M., Runeson, P.: Checklists for Software Engineering Case Study, Lund University, Sweden. Software Engineering Research Group (2007)

8. Höst, M., Runeson, P.: Guidelines for conducting and reporting case study, Lund University, Sweden. Software Engineering Research Group (2008)
9. Huang, S., Wang, X.: Research on Methods of Integrated Information Systems Based on BSP. In: Proceedings of the 2010 Fourth International Conference on Genetic and Evolutionary Computing, pp. 546–549. IEEE Computer Society (2010)
10. IBM. Business Systems Planning: Information Systems Planning Guide. 4th edn. International Business Machines Corporation, USA (1984)
11. Mesquita, A.: Tecnologias e Sistemas de Informação para a indústria da panificação e pastelaria: proposta de uma arquitectura de informação. Dissertação de Mestrado, Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal (2007)
12. Mohammad, D.: A New Methodology for Developing the MIS Master Plan. Review of Business Information Systems 13(1), 15–24 (2009)
13. Rocha, A.: Evolution of Information Systems and Technologies Maturity in Healthcare. International Journal of Healthcare Information Systems and Informatics 6(2), 28–37 (2011), doi:10.4018/jhisi.2011040103
14. Rocha, A., Santos, P.: Introdução ao Framework de Zachman. Apontamentos de Planeamento de Sistemas de Informação, Universidade Fernando Pessoa, Porto, Portugal (2010)
15. Rocha, A., Silva, C., Lamas, M., Castro, R., Silva, S.: Governo Electrónico nas Juntas de Freguesia: Situação na Região do Minho. In: Actas da 6 Conferência da Associação Portuguesa de Sistemas de Informação, Bragança, Portugal (2005)
16. Rocha, Á.: O Essencial dos Sistemas de Informação. Apontamentos de Sistemas de Informação, Universidade Fernando Pessoa, Porto, Portugal (2002)
17. Sakamoto, J.G., Ball, F.W.: Supporting Business Systems Planning studies with the DWDC Data Dictionary. IBM System Journal 21(1), 54–80 (1982)
18. Sayles, A.: Development of Federal Enterprise Architecture Framework using the IBM Rational Unified Process and the Unified Modeling Language. Software Group, IBM, USA (2003)
19. Song, H., Song, Y.: Enterprise Architecture Institutionalization and Assessment. In: Proceedings of the 2010 IEEE/ACIS 9th International Conference on Computer and Information Science (ICIS 2010), pp. 870–875 (2010)
20. Souza, J.B.: Uma metodologia para planejamento de arquitetura de informações. Dissertação de Mestrado, Universidade Federal de Santa Catarina, Florianópolis, Brasil (2001)
21. Spewak, S.H., Hill, S.C.: Enterprise Architecture Planning: Developing a Blueprint for Data, Applications, and Technology. John Wiley & Sons, New York (1995)
22. Teixeira, P., Brandão, P.L., Rocha, A.: Promoting Success in the Introduction of Health Information Systems. International Journal of Enterprise Information Systems 8(1), 17–27 (2012), doi:10.4018/jeis.2012010102
23. Tomé, P.: Modelo de Desenvolvimento de Arquitecturas de Sistemas de Informação. Tese de Doutoramento, Universidade do Minho, Guimarães, Portugal (2004)
24. Torres, A.: Poder local: Como potenciar a participação dos cidadãos na vida do município de Montijo. Dissertação de Mestrado, ISCTE - Instituto Universitário de Lisboa, Portugal (2009)
25. Zachman, J.: A Framework for Information Systems Architecture. IBM Systems Journal 26(3), 276–292 (1987)
26. Zachman, J.: The Framework for Enterprise Architecture: Backgroound, Descrition and Utility. Zachman International, USA (1996)
27. Zachman, J., Inmon, W., Geiger, J.: Data Stores, Data Warehousing and the Zachman Framework: Managing Enterprise Knowledge, 1st edn. McGraw-Hill, New York (1997)

TSPi to Manage Software Projects in Outsourcing Environments

Jezreel Mejia[1], Andrés Garcia[2], and Mirna A. Muñoz[1]

[1] Centro de Investigación en Matemáticas, Unidad Zacatecas
Avenida Universidad No.222, Zacatecas Zacs., México
{jmejia,mirna.munoz}@cimat.mx
[2] Dpto. de Sistemas y Computación, Instituto Tecnológico de Zacatecas,
C/ Panamericana Entronque Guadalajara, s/n.
Apartado Postal 245 Zacatecas Zacs., México
asalas750@msn.com

Abstract. Software development outsourcing is now a reality in both international and national organizations because it represents a competitive strategy. However, even when organizations recognize in the outsourcing competitiveness and business strategy, the software development projects in outsourcing enviroments fails due to the inadequate project management. Therefore, this paper presents a proposal of how to implement TSPi Methodology to manage software project in outsourcing environment. To achieve this, the main failure factors in software development in outsourcing environments are identified and a traceability to identify the adaptability of TSPi to this environment is showed.

Keywords: Outsourcing, TSPi, Project management, Failure factors.

1 Introduction

The Information Technology (IT) Outsourcing is the organizational decision to turn over part or all the process related to the development, maintenance or exploitation of systems [3]. And nowadays the IT Outsourcing is having a fast growth worldwide [2, 16]. Because the organizations have seen in the IT outsourcing a way to achieve their strategic goals, reduce costs and improve customer satisfaction [1, 2]. Today, Outsourcing is a competitiveness strategy for companies [3]. According to (IDC Nasscom Strategic Review 2010) estimated that by 2013 global spending on IT outsourcing and Business Process Outsourcing would have a value of 292.000 md and 148.000 md respectively.

However, while the IT outsourcing is experienced a considerable growth, the number of reported cases of failure is also increasing [14]. According to a recent study [25], 20 to 25 percent of large information technology (IT) acquisition projects fail within two years and 50 percent fail within five years.

In order to avoid these failures, there are models and standards such as ISO/IEC 12207(*Information Technology/Software Life Cycle)*, CMMI-ACQ (*Capability*

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 411–420.
DOI: 10.1007/978-3-642-36981-0_38

Maturity Model Integration for Acquisition) among others, that provides guidance to extract practices for increasing the product and services quality in outsourcing environments.

However, although there have been developed standards and models focused on software process improvement in outsourcing environments, the implementation of these models and standards has been obstacles for most organizations, especially SMEs due to its organizational structure. Moreover, SMEs face a great challenge: competitiveness problem, because they are organizations that handle limited economics resources and are unable to implement models and standards which ensures the quality and productivity in software product development in outsourcing environments. Besides these models and standards only indicated the "What" without show the "How" to do, to perform the practices or activities that these contains. In response to this, there are methodologies, which are adaptable to this kind of organizations as TSPi Methodology. TSPi achieves better manage in software projects increasing software quality and productivity [4] [5]. Therefore TSPi can be tailoring to project management in outsourcing environments.

This work aims to show a TSPi [5] methodology tailoring proposal for project management in outsourcing environment. To achieve this, the main failure factors in software development in outsourcing environments are identified and a traceability to identify the adaptability of TSPi to this environment is showed.

This paper is structured as follows: section two shows the research context, section three describes the TSPi tailoring in outsourcing environments, and finally section four shows the conclusions and future work.

2 Research Context

Before showing the TSPi methodology tailoring in Outsourcing environment, the main concepts used in this research are briefly described: project management, models and standards for outsourcing and TSPi methodology.

2.1 Project Management

Project management is the application of knowledge, skill, tools, and techniques to project activities to meet project requirements. Moreover, it take into account processes such as: initiating, planning, executing, controlling, and closing in order to create products that meet the goals, requirements and criteria of customers [6]. For many organizations, the lack of good project management is the main cause of projects failure; According to Standish Group report "Chaos Report 2007" [7] reinforce that there are still problems related to project failure because: 1) 35% of the initiated projects are categorized as successful, 2) 19% of the initiated projects failed, and 3) 46% of software projects exceed time and cost and unknown user requirements [7]. A similar situation is found in software project in outsourcing environments. A study performed by Software Engineering Institute (SEI) [16] indicated the Mismanagement, the inability to articulate customer needs, poor requirements

definition, inadequate supplier selection and contracting processes, insufficient technology selection procedures, and uncontrolled requirements changes are factors that contribute to project failure. Therefore, these projects can rarely be completed in time [15] as is show in a study performed by the Garner Group based on quantitative data and analysis of a set of 845 IT projects in outsourcing environments. This study indicated that 42.5% did not deliver all the expected benefits, 44% were delivered over budget, and 42% were not delivered on time.

2.2 Models and Standards for Outsourcing

There are models and standards such as ISO/IEC 12207(Information Technology/Software Life Cycle), CMMI-ACQ (Capability Maturity Model Integration for Acquisition), PMBOK (Project Management Body of Knowledge) and eSCM-CL (eSourcing Capability Model for Client Organizations) among others, that provides guidance to extract practices for increasing the product and services quality in outsourcing environments. A brief description of these is presented:

- **CMMI-ACQ:** is a model that provides guidance to acquisition organizations for initiating and managing the acquisition of software products and related services. The model focuses on acquirer processes and integrates bodies of knowledge that are essential for successful acquisitions [16].
- **PMBOK:** is a collection of processes and knowledge areas generally accepted as best practice within the project management discipline [12].
- **eSCM-CL:** is a best practices model that enables client organizations to appraise and improve their capability to foster the development of more effective relationships and to better manage these relationships. This model allows client organizations to continuously evolve, improve, and innovate their capabilities to develop stronger, longer-term, and more trusting relationships with their service providers [13].
- **ISO/IEC 12207:2008:** the purpose of this International Standard is to provide a defined set of processes to facilitate communication among acquirers, suppliers and other stakeholders in the life cycle of a software product [14].

Although there are models and standards that refer directly or indirectly on outsourcing or procurement process, in all organization are not available its models and standards implementation such as SMEs due to its organizational structure and models and standard implementation cost. Moreover these models and standards indicate only the "what" without indicating the "How" to do the activities or practice mentioned.

2.3 Team Software Process introduction (TSPi)

Most of software is developed by team [9], however teams can do extraordinary work, but only if such teams are properly formed, suitably trained, staffed with skilled members, and effectively led [10]. Because of the situation Watts Humphrey has created a methodology for helping organizations to develop quality software in teams:

the Team Software Process methodology (TSP). TSP can help organizations at all maturity levels to greatly accelerate its CMMI software process improvement.

After TSP, TSPi was released. TSPi is an academic version of the TSP, it helps to introduce a graduate or upper-level undergraduate software engineers within teamwork environment, with defined process and producing high quality products. TSPi is a methodology that provides a defined process to develop software by teams. The team should plan and estimate their work, fulfill their commitments, and improve their productivity and quality.

TSPi objectives are to show defined process components (roles, scripts, forms and standards), assess the process importance and enable how to improve the process by using available historical data. So it was developed to be used by graduate or upper-level undergraduate software engineers in order to learn how to be an effective engineer being able to work on teamwork.

Fig. 1 shows TSPi structure [5]. TSPi uses multiple development cycles to build a final product. Each cycle starts with a launch phase, the team should follow seven process steps: strategy, planning, requirements, design, implementation, test and post-mortem [12].

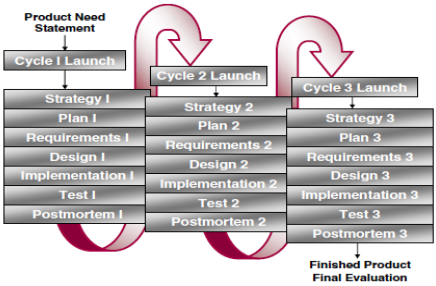


Fig. 1. TSPi Structure [5]

- **The Launch phase:** is used to form and build the teams, determine team member roles, establish team goals, describes client's needs, and help the teams decide on their working practices. The team's goals allow establish the framework for the strategy and plan phases.
- **The Strategy phase:** Team's agreements on strategy criteria are established and then the strategies are produced and updated. The strategy objective is to produce a quality product with the required functions in the available time. A modular design with separable elements that each engineer can develop is established. Moreover a conceptual design is established in order to make size and development time estimates.
- **The plan phase:** defines the artifact size to be developed in each cycle, the task and time estimates for each task and each team member is identified. In this phase a weekly Schedule is developed. Moreover a quality plan is developed.
- **The requirements phase:** the software requirements specification is produced. Its purpose is to describe the functions that team intend to perform. Moreover this phase verifies the customer needs.

- **The design phase:** its main goal is to ensure that engineers produce thorough high quality designs. The overall design structure is established and then it is divided into its principal components. The team members then separately design these components and provide their designs to the development manager, who combines then into the system design specification. It is also important to produce and inspect the integration test plan.
- **The implementation phase:** the principal steps in this phase are implementation planning, detailed design, detailed-design inspection, coding, code inspection, unit testing, component quality review, and component release. Implementation standards are also developed to add to and extend the standards defined in the design phase.
- **The test phase:** the purpose is to assess the product, not to fix it. However, the teams have the data to judge which parts of the system are most likely to be defect-prone. In this phase some members of the team draft and review the user documentation while other members do the testing. A well-design manual is organized around the user's needs and not the produce's structure and functions.
- **The postmortem phase:** this phase provides a structure way to learn and improve. The postmortem starts with the quality/process manager leading the team through a review of the team's data. The postmortem phase is the appropriate point to identify specific improvement opportunities and decide where and how to incorporate theses changes into your personal and team processes.

3 TSPi Tailoring in Outsourcing Environment

Once the research context was presented in the prior section, this section shows the TSPi methodology tailoring proposed to manage software project in outsourcing environments. To achieve this, the main failure factors in software development in outsourcing are identified and then a traceability to identify the tailoring of TSPi to this environment is performed.

3.1 Failures Factors in Software Development Outsourcing

Organizations have been in IT Outsourcing a way to achieve their strategic goals. However, there is evidence to suggest that organizations are not achieving the desired benefits from IT outsourcing.

According to Hanlie Smuts [15] the follow failures factors in Software projects in outsourcing are: 1) Delivery performance: Services must be delivered not only to expectations, specifications and quality, but improved continuously; 2) bad contract management: Both parties must have good contract management skills, processes and people, 3) Communication poor: There must be ongoing and effective communication between parties; 4) Technical expertise inability: The supplier must provide quality technical expertise.

Others failures factors identified by Software Engineering Institute [16] are: Mismanagement, the inability to articulate customer needs, poor requirements

definition, inadequate supplier selection and contracting processes, insufficient technology selection procedures, and uncontrolled requirements changes are factors that contribute to project failure [16].

Eruani and Geneviève [17] have pointed failure factors between acquirer and supplier such as: The lack of managerial competencies that enable the communication and management of the acquirer-supplier relationships is identified as an important obstacle. Moreover, they have indicated the criteria of Project success such as: meeting planning goals, which include: meeting schedules and meeting budgets; providing benefits to management by achieving the expect technical, strategic and economic benefits and Building potential for future revenue for vendors through referrals, testimonials and repeat business (continuous relationship).

Others studies performed by Matt. E. Thatcher [18] identified that project success may be affected by: (1) the level of existing business and technical expertise accumulated by both client and vendor, (2) the extent of bilateral exchanges of business and technical knowledge between client and vendor, and (3) the degree of integration of such transferred knowledge with existing expertise by both client and vendor. In addition, the effectiveness of knowledge exchanges and integration may be influenced by the degree of modular design characterizing the client's processes and outsourced project.

With the goal of improving their processes, companies such Philips, Siemens or Nokia [19] are reviewing their outsourcing projects to find generic problems. As results these companies have identified generic issues related to: level of detail in the contract and the related requirements, insufficient project planning of the supplier, lack of visibility of the project progress, process maturity of the subcontractor, process compatibility between the organization's cultural differences and Risks management.

Finally, J. Yanbing [20] analyzed in the Outsourcing process and found failure factors related to the client, suppliers and technology. This analysis highlights that organizations recognize the outsourcing as a business strategy. However, these organizations have doubts when they outsource: how to select the service supplier, how to ensure the software quality.

According to different authors and studies analyzed above on companies involved in software project in outsourcing environments Table 1 summarizes the most relevant failure factors.

Table 1. Failure factors in software projects in Outsourcing environments

Issues	Issues
(I1) Projects that not satisfy the customers	(I13) Poor comunication between customer and supplier
(I2) Mismanagement of contract	(I14) Requerimients and details undefined
(I3) Poor communication betwwen customer and supplier	(I15) Inadequate Planning supplier
(I4) Suppliers with technical inability	(I16) Project process and monitoring inefficient
(I5) Inadequate customer management	(I17) Lack of provider process measurement
(I6) Poor requirements changes	(I18) Incompatibility betwen customer and supplier
(I7) Uncontrolled requirements changes	(I19) Different culture in organizations
(I8) Improper selection of suppliers	(I20) Managing project risks
(I9) Difeerent culture in organizations	(I21) Deliverables and milestones are not well defined
(I10)Lack of supplier-customer communication management	(I22) Complication project in monitoring
(I11)Projects must be modularized	(I23) Submissions received after the deadline
(I12)Lack of project visibility	(I24) Software quality is not guarantedd

Once, the failure factors were detected, a modified C&C Model for Outsourced Global Software Development [21] was analyzed in order to establish a better classification of prior factors. This modified model covers failure factors related to: organizational, people, process, technical and project factors.

According to analysis performed on Table 1 two groups of failure factors were identified and include in the modified C&C model [21]: Communication and requirements groups. Therefore, a new extended model with seven groups of failure factors was established (Fig. 2).

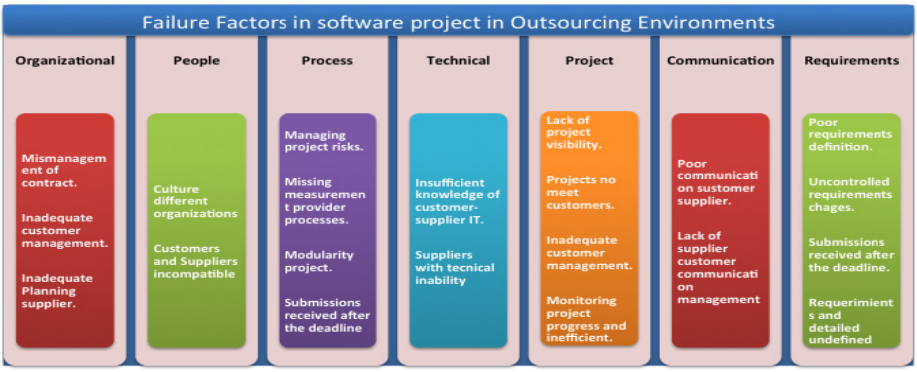


Fig. 2. Failure factors model in outsourcing environment

3.2 TSPi Tailoring

According to the software projects nature in outsourcing environments, these are focused on the following critical factors: organizational, people, process, technical, project, communication and requirements (see Fig. 2.). The base material for doing the TSPi tailoring was:

- The results obtained in Fig.2. This was made to identify the main failure factors in software projects in environment outsourcing.
- All TSPi material. The material includes principles, phases, phases' scripts, forms and forms instructions [5].

The input information was a study about TSPi principles shown in Table 2.

Table 2. TSPi basic principles

# Principle	TSPi Principles
1	Processes and roles well define
2	Team philosophy: commitment and collaboration
3	Quality based on the anticipated reduction defects
4	Introduction of formal inspection of process and quality plan
5	Project monitoring based on earned value
6	Tracking weekly meetings
7	Delimitation realistic and detailed Project scope

These principles allow ensuring the productivity, quality in all software development lifecycle. The TSPi principles were analyzed in order to know:

1) How a TSPi principles are done (what are the involved phases and forms should be performance), what forms are proposed by TSPi as support for record data (what are the forms an how to use it). Table 3 shows how the TSPi principles are covered.

Table 3. Phases and forms involved in TSPi principles

TSPi Principles	Involved Phase	Script- Form
1	• Launch • Strategy	LAU Script, the purpose is to start the team in the development cycle. It defines the purpose of the project team and assigning roles, describes customer needs, individual goals and team goals. STRAT Script, guides the team, describes the strategic decision to include all the features in the product at all levels. Stored lines of code per function, and estimated by function, are registered.
2	• Launch. • Plan.	INFO Form, its purpose is to integrate the teams, each team member define their interests, experiences, work on projects and other related equipment.
3	• Plan. • Test. • Require ments.	SUMS Form, the purpose is to concentrate the estimated sizes and names of components to develop, test plans requirements and specifications. It defines the total of lines of code, text pages. ITL Form, its purpose is to record and monitor the project risks or problems. They define risk or problem, who gives up, settlement date.
4	• Plan.	SUMQ Form, its purpose is to record the data of planned and actual quality. Data are stored as modules, components, performance stages, process, and defect removal rate, defects.
5	• Plan. • Imple mentation.	SUMP Form, its purpose is to store data and actual plan suite. Product defined size; time phased injected defects, defects eliminated.
6	• Require ments. • Postmortem.	SCHEDULE Form, its purpose is to register the estimated and actual hours consumed weekly, and tracking is performed with respect to earn values as planned tasks consumed. Defines number of week, date, hours per week, hours of the team, earned value and accumulated. WEEK Form, is used to make a weekly report, and describes project hours week cycle, by team members, the tasks performed, tracking issues or risks.
7	• Plan	TASK Form, its purpose is to estimate the development time of the project tasks, estimated planned value and estimated completion date. It defines the task name; the team planned hours, planned and actual size.

The analysis identifies how failure factors in Fig. 2., are covering throughout performing the TSPi principles identified in Table 1. Besides, it identifies the involved phase and which information should be register in the provided TSPi Script-forms (Table 3.). Taking into account the factors (see Fig. 2) and TSPi principles (see Table 1),

phases, phases' scripts, forms and forms instructions is presented as the TSPI methodology can be adapted to environments outsourcing. Fig. 3 shows the TSPi tailoring.

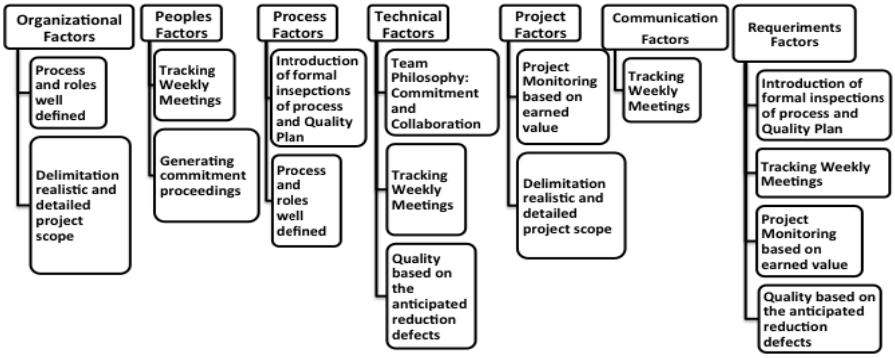


Fig. 3. TSPi tailoring for failure factors in Outsourcing software projects

As shown in Figure 3, the organizational factors are covered by the TSPi principles (2) and (5) which determine the project scope and the roles and process are defined. Personal factors are covered with the principle (7) where weekly meetings and commitments are established. Process factors are covered with the principle (4) through formal inspections and process quality plan. Technical factors are covered with the principles (2), (3) and (7) where team cooperation and commitment are established. The project factors are covered with the principle (6) which project monitoring is established. Communication factors are covered with the principle (7) establishing meetings weekly to analyze the project which enriches the customer communication. Requirements factors are covered through the principles (4), (7), (6) and (3) because the weekly meetings, project monitoring, inspections and process formal quality plan allows to identify problems in project requirements.

4 Conclusions and Future Work

TSPi methodology analysis allows to establish a stage to manage software projects in outsourcing environments, although TSPi is not designed for this kind of environment. This analysis established that according to the principles on which this methodology has been defined, they are tailor to avoid failure factors of such projects. Moreover, the analysis shows the involved phases and forms should be performed.

Therefore, the TSPi methodology through its principles helps ensure success in software project management in outsourcing environments.

As future work is developing a web application for software project management in outsourcing environments based on the TSPi principles for SMEs.

References

[1] Ferguson, I.: El outsourcing, un impulso para México, `http://www.cnnexpansion.com/mi-carrera/2010/01/07/outsourcing-un-impulso-para-mexico` (accessed on August 15, 2012)

[2] Trejo Flores, L., Navarrete, C.J.: Information Technology Outsourcing in Mexico: Adoption Practices and Benefits. In: Americas Conference on Information Systems (AMCIS), pp. 1–19 (2009)
[3] Zainuddin, E., Bassellier, G., Benbasat, I.: Outsourcing projects success: the role of competence and leadership of the vendors and clients project managers. In: SIGMISCPR, pp. 83–86 (2010)
[4] Cuevas, G., Calvo-Manzano, J.A., Mejia, J., Muñoz, M.A., San Feliu, T.: How is CMMI-DEV applying when using TSPi project planning. In: CERMA, pp. 143–148 (2009)
[5] Humphrey Watts, S.: Introduction to the Team Software Process. Addison Wesley (2010)
[6] Pressman, R.: Ingeniería del Software un Enfoque Practico, 7th edn. Mc Graw Hill (2010)
[7] The Standish Group: The report chaos. The Standish Group, http://blog.standishgroup.com/ (accedido el de Agosto de 15, 2012)
[8] Lugo Garcia, J.A., Garcia Perez, A.M., Delgado Martinez, R.: Indicators management in software projects Present and future prospects. In: RCCI, pp. 19–25 (2009)
[9] Lepmets, M., McBride, T.: Process Improvement for the Small and Agile. In: Winkler, D., O'Connor, R.V., Messnarz, R. (eds.) EuroSPI 2012. CCIS, vol. 301, pp. 310–318. Springer, Heidelberg (2012)
[10] Gasca Hurtado, G.P.: Similarity study of risk management process in software outsourcing projects: using a method abstract. In: Redalyc, pp. 119–129 (2010)
[11] Software Engineering Institute: CMMI for Acquisition, version 1.3. Ed. SEI (2010)
[12] Rojas, M., Villamizar, L.A., Orjuela, A.: Model for integration for work management PMBOK guide with engineering activities in software development projects. Revista Avances en Sistemas e Informatica, 98–106 (2011)
[13] Hernández López, A., Colomo Palacios, R., García Crespo, A.: Recommendations for adoption of practices for human capital management in outsourcing environments: integration of eSCM-CL with People-CMM. In: REICIS, pp. 20–37 (2009)
[14] Mejía Miranda, J.: Marco para el Gobierno de la Externalización del Proceso de Desarrollo del Software a través de la Ingeniería del Contrato. Departamento de Lenguajes y Sistemas Informáticos e Ingeniería de Software, Facultad de Informática, Madrid España (2010)
[15] Hanlie Smuts, P.K., Van Der Merwe, A., Marianne, L.: Critical Success Factors for Information Systems Outsourcing Management: A Software Development Lifecycle View. In: SAICSIT ACM, pp. 304–313 (2010)
[16] Mejia, J., Muñoz, M., Calvo-Manzano, J.A., Cuevas, G., San Feliu, T.: Practices Organization before Formalizing an Acquisition Contract according to the CMMI-ACQ Model. In: CISTI, pp. 1–7 (2011)
[17] Eruani, Z., Geneviève, B., Izak, B.: Outsourcing Projects Success: The Role of Competence and Leadership of the Vendors and Clients Project Managers. In: SIGMIS-CPR 2010, pp. 83–87 (2010)
[18] Matt, E., Thatcher, H., Manju, K., Ahuja, D.E.: IT Outsourcing: Assessing the Antecedents and Impacts of Knowledge Integration. In: Proceedings of the 44th Hawaii International Conference on System Sciences, pp. 1–10 (2011)
[19] Calvo-Manzano, J.A., San Feliu, T., Purshe, A.: The Forgotten Practices of Subcontracting. In: Information Systems and Technologies (CISTI), pp. 1–4 (2010)
[20] Yanbing, J., Lei, C., Xin, Z., Ying L.: Process-oriented software outsourcing decision based on genetic algorithm. In: SOLI, pp. 386–391. IEEE (2010)
[21] Thomas, G., John, T.N.: Model of Critical Factors for Outsourcing Agile Development. In: AMCIS, pp. 1–11 (2009)
[22] Caballero, E., Calvo-Manzano, J.A., Cuevas, G., San Feliu, T.: Analysis of quality and productivity with TSPi in a software development Project in a microcompany. In: REICIS, pp. 28–37 (2009)
[23] García, M.C., Garzas J.: Externalización del desarrollo Software. Kybele Consulting, 1–5 (2009)

Aspects That Contribute to the Success of Personalized Web Applications

José Luís Reis[1] and João Álvaro Carvalho[2]

[1] ISMAI, Maia Institute of Higher Education, Research Unit UNICES, Maia, Portugal
[2] Centro Algoritmi and DSI, Universidade do Minho, Guimarães, Portugal

Abstract. Personalized applications exhibit features adjusted to the preferences or needs of users and can provide benefits related to customer relationship or related to the efficiency of organizational work. This article presents a conceptual model that can be viewed as a contribution to the body of knowledge of Information Systems and Technologies professionals with responsibilities in the adoption, and development of personalized Web applications. These contributions include: a conceptual model that combines and articulates aspects that contribute to the success and to the quality of personalized information system (this model combines design features, facets that contribute to the quality and metrics that can be used to evaluate the performance of organizational work supported by personalized information systems). This set of aspects combines a synthesis of information so far scattered in the literature with proposals and recommendations with empirical knowledge based on an empirical study.

Keywords: Personalization, systems, information, quality, metrics.

1 Introduction

With the globalization of information and communication technologies (ICT) (due mainly to wide dissemination of services based on Internet/Web), personalization become a reality in many Web sites and in some organizational information systems. The personalization of applications and computing platforms is a complex process. This work aims to address the following questions: a) what kind of factors contribute to the quality of a personalized information system? b) what aspects should be taken into consideration when implementing a personalized system? what factors influence the performance of a personalized information system?

This article presents the following research results: a) the main quality facets of a personalized information system; b) a conceptual model covering aspects such as the main features that contribute to the personalized system design, the quality facets and the metric for the work execution using personalized applications.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 421–432.
DOI: 10.1007/978-3-642-36981-0_39

2 Personalization Types

There are several areas, both in academia and in organizations, where personalization is considered as an issue. Examples include: management, economics, marketing and sales, political science, sociology, anthropology, medical sciences, clinical information systems and computer science [1]. In what concerns information systems, the main goal to personalize ICT applications is to fulfill the needs and preferences of a person (or group of persons) that use a specific personalized ICT application.

Personalized ICT applications (also named as Personalized Applications – PA) facilitate the interaction among the various actors that cooperate within an organization, and the interaction with information providers and information recipients. It allows people to have easier access to content and services.

The "individuality" of consumer information is the key aspect in the personalization definition, implying that an individual's attributes, such as identity, preferences, constraints and provision (example, location and presence) have a potential impact on the personalization process [2].

According to the personalized features and purpose of a PA it is possible to establish different typologies to achieve personalization based on different aspects, including [1], [3]:

- What is personalized? The system characteristics that will be personalized (for example: content, user interface, information features, and channels for information delivery).
- Who are the personalization targets? Individuals or groups/categories of individuals.
- Who sets the personalization (user or system, or both)? **Implicit or adaptive/proactive**, if the personalization is done automatically by the system when the system, through collaborative filtering techniques, discovers product preferences of a customer. **Explicit or reactive** if individuals participate in providing information about their preferences and needs.
- How is personalization achieved? There are two possible ways to achieve personalization: a) through processes of personalization that not require intervention from users - **observational** [4] **or implicit** personalization; b) personalization processes that somehow require the participation of users. In the last case, when users are interrupted in the work processes they are involved in to provide answers to questions that are placed upon it, the personalization is said to be **intrusive or explicit** personalization [5].

3 Quality Facets of a Personalized Information System

The quality of a system is measured by its desirable characteristics. For example, in Web enable e-commerce support systems, quality involves usability, trust/safety, reliability, adaptability and performance (response time) [6]. The quality of information refers to the comprehension of the system contents [7] and has characteristics like: form (more discursive or more graphical qualitatively or

quantitatively, etc. depending on the technology), age (the time interval between the time the problem/event arises and the time when the user becomes aware of it), frequency (number of times information is available over a period of time), timeliness, relevance, security, profitability, cost, accuracy, completeness and consistency [8]. The individual impact of the system in the user can be measured by the system performance in terms of decision making, effectiveness of quality tasks.

In the case of a Web system, for having quality the system must be personalized, complete, relevant easy to obtain and memorize [6]. Service quality refers to the total support provided by the service provider, particularly in aspects related to the trust/assurance, empathy, responsiveness and accountability [7]. The evaluation of information quality, considered an economic asset, derives from the state of lack of need that the user is (what you want to meet to annul the uncertainty and the imbalances to be corrected) [9]. System personalization can improve the shape, age and frequency of information, tailored to the needs of the users [10].

The quality facets of a personalized system are from various types and have different characteristics. The following paragraphs present the quality facets of a personalized IS, mentioning their characteristics and underlying assumptions.

Usability

The software should be understood, learned, used and appreciated in a positive way by the user according to their personal characteristics, when used under specified conditions. It should be easy to use, must have a help system easy to understand, especially in a language understood by the user and must have good presentation according to user's preferences.

Adaptability

Ability to adapt to different user profiles according to their preferences and needs. Allow user to select their favorite colors, layout options, shortcuts for quick access to features, information and forms to submit, etc. It must be linked to suits of the system provider, as well as presents and expands needs of those who use the system according to the conditions specified by who provides the system.

Flexibility

Different interface forms allow the same system to be used in different circumstances and used by users with different characteristics, thus achieving the same system to achieve the same goal of several ways. A personalized system must recognize if the user is accessing it via a mobile system (smartphone) or via a fixed system (desktop) and allow the options and features are available in different size screens and graphics motion. The system must be prepared to be used according to preferences and context of use, a smartphone, a computer, a tablet, a PC, a multimedia kiosk, a PDA, a plasma media system, either with a touch screen, with or without a mouse, with a pen, in different operating systems, Windows, Android, Linux, Vista, Mac, etc. with different bandwidths, in any type of navigation system, in different languages, etc. always according to users preferences. To verify the consistency and flexibility in a system is guaranteed to make tests on the way with a minimum of clicks and how quickly one can achieve a certain goal, for example finding a product in an e-commerce shop.

Ubiquity
With digital identity, users can change their profile according the contexts and situations where they are, and it becomes even more complex to manage these processes to understand the link between the "real identity" and "digital identity" [11]. The information recorded in databases cannot define the texture of our lives leading to uncertainty about the information. The data, rather than providing a portrait of our personalities, are compiled and stored, as were captured, from what we did, not often have an underlying reason for having done, which means that we are reconstituted on the digital basis data as people made up of data - paradoxically the problems associated with violation of the privacy of personal data, which do not allow their use and dissemination, are distorting what we are in the digital world. As computers become increasingly pervasive (ubiquitous) in our lives, in smart environments that lead to greater collective intelligence, is need to create systems for identifying users who do control, with efficient and easier parts of information belonging to a particular individual which are available to others in the environment, with policies and protocols that require access to this information in an increasingly safe and enabling access to more than one platform [12]. For example, Google allows the same login and password for access to different applications and platforms (Gmail, YouTube, Blogspot, Picasa, Googledocs, etc.). The W3C has created a script (roadmap) which refers to the personalization elements that allow the integration of several features in ubiquitous environment with Web access to all interested [13].

Reliability
Ability to maintain the level of performance when used under specified conditions, allow recovery of failures in time. Ensure robustness for use in extreme conditions, such as a high number of users simultaneously, with greater difficulty of access, in case of breaking system allowing to return to work at the same point of break without information loss.

Opportunity
The information must be displayed when the user needs it, and may be reactive, preferably pro-active, but always going to meet the needs of the user without being too obtrusive. Should take into account user preferences when is using the system and in accordance with the user profile (recommendations or notifications system).

Efficiency
Operate with a level of performance required for the quantity of resources associated with it, when used according to specified conditions. Allow register and remove in time, the information that the user need. Present the information in a legible, accurate and in accordance with user expectations.

Effectiveness
Should be found a way to combine the interest of who provides the information and who receives it, creating value to the system and for who use it. Should be optimized the existing resources for those who produce information (sell, provide a service, etc.) and who receives it (buy, watch, etc.).

Security

The system must be prepared to perform secure transactions without violating the laws that protect organizations and individuals regarding the processing of information in a secure manner. The system must provide accurate and useful information that the user needs in a safe manner, without violating their privacy and in accordance with the user's profile, according to their preferences for the presentation of such information, either in time or in the medium. Should ensure interoperability between different systems and must be secure in the use and in the registration of transactions. A personalized system must identify who uses, ensure privacy, set the preferences to allow use to infer how to use, integrate different components from different systems in the same work environment, possess security certification (encryption, etc.).

Accessibility

The system shall be designed and constructed to be accessible regardless of the user's profile, their physical limitations and types and versions of hardware or software based that will be used. Accessibility is characterized by attributes that let you set up the information system is "accessible" to all citizens in all the fullness users to the content available. The visual and auditory forms of content should be available to users, as well legends to explain that allow describing the more complex visual content, such as charts or diagrams; this content can be converted into voice through adapters [14]. W3C allows, by inserting in the address http://validator.w3.org/ the site pages, to verify that the HTML code for a set of hosted pages, accessible from that address meets the quality standards of HTML for improving the usability and accessibility of Web pages with the subsequent benefits of these quality standards [15], [16]. If this occurs, the W3C returns a connection (link) that lets put a symbol that "certifies" the inherent quality and accessibility of these pages. A personalized system should have features that do not allow the exclusion of any user. Thus, despite the personalization information system to be independent of access is desirable that the information system having these qualities can be used by any person according to personal characteristics.

Performance

The system must be able to increase performance levels of user benefits for the organization and for individuals. Allow the user to have easier access to their favorite menus, to the most used applications and the data it needs to improve its performance. The system must allow the needs of users according to their profile, in a computing environment pleasant easy to use and making the information available useful. According to preferences and the right time, with the right way, the system helps the user to solve their problems by increasing the levels of user comfort. It is often said that a computer application should take no more than 3 seconds to respond. The system must be prepared to respond in a timely manner to requests and queries that are performed by users with information adapted to user's needs and preferences, and without additional waiting time caused by personalization of information or interfaces.

4 Design Features of Personalized Applications

The success of a system can be evaluated through a number of aspects that influence how it is used. The system performance evaluation should be made taking into account that there are external system aspects, for example, the system is installed on a computer that has the minimum requirements to support the system, either for reasons related to low memory or disk available space, either by a problem related to the characteristics of a data network, which may not have enough bandwidth to support the transactions necessary for the proper performance of the system, or because the user is using the system in a wrong device media.

The internal components can influence the performance of a system, because most systems are designed to be used by individuals who require these components to be considered when systems are developed. The internal components of the system, relate to the characteristics of the architecture (design) of computer applications, need attention and should be respected. These are aspects relate to specifications associated with the source code of software that was developed, including techniques used in the construction of the system and in its components in the construction on the following three levels [17]:

- Control of its structure with aspects related with modularity, complexity, cohesion and consistency of the components of the control flow.
- Structure of information with aspects related to the overall structure and types of data, the articulation between data (coupling) and the system, the consistency of the flow and type of data, the integrity and complexity.
- Code related typography for example: format, separating modules or spaces/blank lines.

5 Metrics for Work Execution Using Personalized Applications

Although there are several possible metrics for the different aspects associated with the efficient and effective performance of a system, let us focus on metrics that somehow contribute to the performance on the work using a personalized system [7], [18]:

- Time spent to complete a task in the system. For example, buy an electronic product in a Web store; the system can provide a personalized list of recommendations and prevent the user from losing time on research and transactions subsequent to the acquisition process.
- Number of tasks that can be completed over a period of time if the system is able to identify the sequence of actions that a user makes to complete a given task.
- Number of commands used and completed in a given period of time and access to the commands available to the users in a sequence according their usage preferences.
- Time spent to find and interpret the information in the guide/user manual.
- Time system inactivity (idle time).
- Marking notes of interest. In the case of a Web site, personalization should allow the user 'bookmark' a page or other elements of interest and they remain available for future reference.

- Accessibility of the system. The system shall be designed and constructed to be accessible regardless of the user's profile, their physical limitations, and types and versions of hardware or software that will be used.

6 Integrated Model of Personalized Systems Design

Taking into account the relevant aspects that contribute to the performance, the quality and the success of a PA, it is possible to construct a conceptual model that integrates the quality of a personalized application (facets that contribute to quality), the personalized software application design (design features) and the performance of work execution using the PA (metrics used to express the performance) - see Fig. 1.

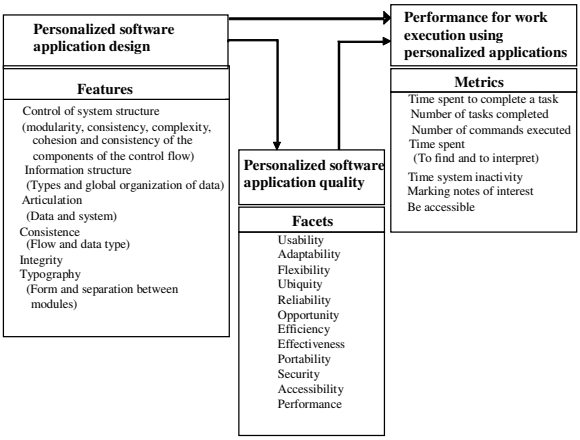


Fig. 1. Conceptual model with aspects that contribute to the design, quality and performance of a personalized system

Fig. 1 depicts the existing relationship between design and quality control of the personalized system (for example, coordination between the control of system structure and the information structure should be organized in a way that ensures accessibility, usability, and adaptability to the users according to their needs). On the other hand, the facets of quality, such as usability and portability, are related to the performance of the work using the personalized system to ensure that the number and commands executed and the time spent on a task, regardless of location and the equipment that is to use, comply with acceptable run times. Furthermore, the design of a personalized system is related to the execution of organizational work system and must have acceptable performance.

7 Case of Practical Application on CVRVV

With the aim of validate the conceptual model with the aspects that contribute for the quality of a personalization, it is used the knowledge inherent in the development of

personalized features on the system of Commission of Viticulture of the Region Vinho Verde (CVRVV), in its Extranet component. The system called INETSIV (SIV - Wine Information System) allows the institution's relationship with its business community and its municipal delegations. In this section of the work it is described CVRVV, the objectives of personalization in CVRVV, what was personalized and the aspects that contribute to the design, quality and performance of the CVRVV personalized system.

7.1 The CVRVV

CVRVV is the entity responsible for the certification of the Portuguese wine products with designation of origin Vinho Verde and Regional Minho. The CVRVV is also responsible for safeguarding the assets of the Region and the institutional promotion in the markets. The Vinhos Verdes Region has about 25,000 producers who have an area of about 22,000 Ha of vines, with 600 entities authorized to bottle on average 55 million liters of wine per year certificate in about 1000 trademarks. CVRVV organic structure sets seven working departments with about 60 employees. The departments ensure the normal functioning of the institution, ensuring certification of wine products, protection of heritage designation and institutional promotion of products and brands from Vinho Verde and Regional Minho. There are various software applications to automate the processes involved in certification, with a set of modules which correspond to different applications required for products certification management. The system contains a set of interfaces, including an Intranet, that enable the interaction between about 50 internal users and the organization and a Extranet accessible to approximately 600 Economic Agents as well as 45 municipal delegations.

7.2 Objectives of Personalization in CVRVV

The project called Web Personalization in CVRVV developed a set of features to improve the way users (Economic Agents and Delegations) interact with CVRVV through the information system. The personalization project allowed to implement a set of personalized features, including:

- A system that records how the Economic Agents (EA) accesses different options via the Extranet (INETSIV), noting in particular the selected options, time of use, as well transactions each done. This system allows users to have access to a range of information related how each user uses the different options available on the Extranet.
- System access options available through facilitating elements, including through quick access buttons to the computer applications system.
- System which allows registration and access to e-mail messages that are personalized according the EA profile.
- Customization of interfaces brands of products with designation of origin certified by CVRVV in its component Extranet, Intranet and Internet. This system component requires an explicit action by EA to change the information associated with their brands.

The personalized features of INETSIV are: 1) personalized interfaces for Economic Agents and Delegations, creating personalized quick access buttons, 2) personalized the messages on the main screen with useful and relevant information to users, 3) customize the mail box; 4) customize interfaces brands. It was also created personalized information elements, which allow access to statistical information from the users. To implement these personalized features where adapted personalization implicit types in most of the functionality's, with the exception of customizing interfaces of the brands and the selection preferences regarding quick access buttons, which require user intervention with explicit personalization.

7.3 Aspects That Contribute to the Design, Quality and Performance of the CVRVV Personalized System

The responsibility on the aspects that contribute to the success of the performance of the information system is from the project coordinator, and even if these aspects are already implemented in the system in operation, should be tailored to the personalized features. Should be implemented, as much as possible, the criteria that influence the facets that contribute to the quality of PA (functionality, reliability, usability, maintainability, adaptability, timeliness, effectiveness, productivity, satisfaction, shape and frequency).

As regards the facets that contribute to the quality of the personalized system in CVRVV the most relevant are contained in the system in operation, namely:

- **Functionality's** appropriated to regulations required by law for certification of products in the wine sector, according to the rules that articulate the specific needs of each user (CVRVV manage a very heterogeneous universe of associates). It is also guaranteed interoperability among information systems, particularly with an external entity (Institute of Vine and Wine) at the level of information that registration is available to members who use the system, as well as some associated with larger (are they are able to make investments in IST) that have information systems that communicate with the CVRVV in an automated way (without human intervention) to make certain transactions (for example participation to affix seals of certification of products in the bottles). In the system of CVRVV there is a guarantee of identifying who uses the system in a safe, secured either by rules imposed by the internal CVRVV system's or by a company that guarantees the safety certification, which ensures the privacy of transactions. The PA in CVRVV also allows users to define their preferences on the options associated with quick access buttons, while also infer, from the use, which the most used options, presenting them in these default buttons.
- **Reliability** that keeps the level of performance under different system access (a user who has access the system through reduced bandwidths has no problem accessing the CVRVV system). When a transaction is broken for any reason, the system allows recovering (in some cases) the transaction that was being made. The PA has no CVRVV user limit and even in extreme conditions with high numbers of users in real time (e.g. from electronic receipt of statements of harvesting and

production, because the legislation requires to be between 1 October and November 15th, rare are the days when you are not at all times working, virtually all delegations related to CVRVV as well as many of the members).

- The **usability** of the information system is appreciated by the user's majority. The available software is easily understood by users, and it is not necessary specific training to promote the system because it is easy to use.
- The **efficiency** is guaranteed by the good performance of the system that allows recording and presenting timely information according the user needs, presenting them in a legible, accurate and in accordance with the legislation (is there detail information about the registration of vineyards, which have some complexity, sometimes it becomes difficult to understand by users and the system help the user).
- The **maintenance** of CVRVV software applications is ensured in accordance with the requirements set by the IST managers, and new applications requested by departments or by CVRVV Economic Agent users (the personalized functionalities developed, together with the brand customization, arose from a system user need). The implementation of new features in the system, does not prevent the system is kept operating in a stable manner.
- The system **portability** is guaranteed, including the possibility of operating in different navigation systems (browsers), on systems with different bandwidths. However, not all features are prepared to function optimally in all touch screens.
- The **adaptability** of the system enables different profiles may be assigned in accordance with the characteristics of each Economic Agent, however it is not possible color selection, nor the provision of options horizontally or vertically (although the system is prepared for accept this preference is not yet implemented), the quick access buttons allow shortcuts for faster access to applications, but not yet implemented shortcuts access the screens applications, although the system provide this functionality.
- The **opportunity** in the information presentation at the time the user needs is ensured in most situations, however, the notification system is the same for all users, only differentiated by the type of Economic Agents rather than their individual profile.
- The system **effectiveness** is guaranteed by the way it can combine the interests of Economic Agents and CVRVV delegations. However, is needed more investments so that they can implement more personalized features.
- Higher the **productivity**, higher is the user satisfaction and higher is the satisfaction of the system provider. The quick access buttons, custom messages and the customized brands contribute to the better CVRVV information system productivity.

As regards the aspects that contribute to the design, quality and performance of a PA in the case of the system of CVRVV, there linkage between data and system structure information which are arranged in a way that ensures accessibility and usability that guaranteed articulation between usability and portability which allows better performance in the execution the system jobs. For example, it is guaranteed the

number of command executed and the time spent on a task is always the same, regardless of where you are and the user equipment that is being used. Moreover, the design of PA is hinged to the performance of organizational job execution. For example, the time spent in execution of a task it's not superior to the established in the specifications because they create personalized information, because they are not hard processing, or the time taken to locate a specific change or information is superior to the times considered reasonable, thus the efficiency and effectiveness of the system of CVRVV are at reasonable levels.

8 Recommendations

The implementation of a personalized system on the Web goes through several stages. It is necessary to ensure the existence of certain conditions for the development, including: evaluating the business model and organizational processes that allow personalization, define personalization types according to the targets, analyze the coordination and combination of the facets that contribute to quality of a PA with the design features of the PA with the metrics used to evaluate the performance of organizational work and define the appropriate technologies to achieve personalization.

The project coordinator must take into account the quality aspects that contribute to the success of the information system. If these aspects have already been taken into account in the system in operation, they must be reviewed and adapted to new personalized functionality's. To minimize the risk of failure it is vital to assure the existence of the resources required to implement the facets that are vital to the quality of the personalized system (functionality, reliability, usability, maintainability, adaptability, portability, efficiency, effectiveness and productivity).

9 Conclusions

The conceptual model presented in this article corresponds to a synthesis of knowledge about the aspects that contribute to the quality of a PA that so far was dispersed in several sources. This dispersion is understandable to the extent that such knowledge results from R&D work in different areas, each focusing on particular aspects of personalization. In this article there is no reference to the technologies that enable personalization, to interface devices and to ways of implementing personalization (personalization levels and dimensions). However the conceptual model provides a basis to sustain the improvement of business management activities. Personal computing and personal Web is a trend that will consolidate in the coming years. Knowledge about the users will increasingly be applied to personalize information content and forms of ICT usage, adjusted to the work environment, allowing users to do the configuration and content reorganization of their work environment to support their professional, personal and even activities related to training according to their preferences. Despite the limitations of this work, particularly relating to validation based on a single study, the conceptual model

constitutes a stable base for PA projects and can be viewed as a contribution to the body of knowledge of managers and IST professionals involved in decisions regarding the adoption of personalized ICT applications.

References

1. Fan, H., Poole, M.S.: What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems. Journal of Organizational Computing and Electronic Commerce 16(3-5), 179–202 (2006)
2. Toth, K., Nagboth, S.R.: Constraint-Based Personalization Model: Multi-Channel Messaging. Dept. of Computer Science, Oregon State University, Oregon, USA (2002)
3. Anand, S.S., Mobasher, B.: Intelligent Techniques for Web Personalization. Springer, USA (2005)
4. Mulvenna, M., Anand, S.: Personalization on the Net using Web Mining. Communications of the ACM 43(8), 122–125 (2000)
5. Perugini, S., Ramakrishnan, N.: Personalizing Interactions with Information Systems. Department of Computer Science, Blacksburg - Virginia (2002)
6. Delone, W., McLean, E.: The DeLone and McLean Model of Information Systems Success: A Ten-Year Update (2003)
7. ISO, I. S.. Ergonomics of human-system interaction - Usability methods supporting human-centred design. ISO/TR 16982 (2002)
8. Oliveira, A.: Anatomia e metabolismo do processo decisório em contexto empresarial * Algumas Questões. Ciências Económicas e Sociais (1999)
9. Oliveira, A.: Análise do investimento em sistemas e tecnologias da informação e da comunicação. Sílabo, Lisboa (2004)
10. Mobasher, B., Cooley, R., Srivastava, J.: Automatic Personalization Based on Web Usage Mining. de DePaul University - College of Computing and Digital Madia (2000), http://maya.cs.depaul.edu/~mobasher/personalization/ (obtained in October 09, 2012)
11. Black, J., Downes, K., Field, F., Mozdzanowska, A., Steel, K.: Technologies: The Metaphors of Emerging Unpacking the disconnects between the "what" and the "how" in the world of "online shopping". M.E. Technologies (Ed.) (2006)
12. Cutting, D.: Identity Management in Context-Aware Intelligent Environments. University of Sydney, School of Information Technology (2006)
13. Heath, A., Schwerdtfeger, R.: W3C Personalization Roadmap: Ubiquitous Web Integration of Access For All 1.0. from W3C (2009), http://www.w3.org/TR/UWA-personalization-roadmap/ (obtained in September 24, 2010)
14. Henry, S. L.: WAI Resources on Introducing Web Accessibility (W3C) from Web Accessibility Initiative, WAI (2005), http://www.w3.org/WAI/gettingstarted/Overview.html (obtained in September 24, 2011)
15. Arch, A.: Ooutdated Incompleted Draft of Auxiliary Benefits of Accessible Web Design (W3C), from W3C - Web Accessibility Iniciative (2002), http://www.w3.org/WAI/bcase/benefits.html#intro (obtained in July 04, 2012)
16. Dubost, K.: My Web site is standard! And yours? from W3C Quality Assurance (2003), http://www.w3.org/QA/2002/04/WebQuality.html#Standard (obtained in June 02, 2012)
17. Oman, P., Hagemeister, J.: Metrics for Assessing a Software Systems Maintainability. In: IEEE (ed.) Conference Publications on Software Maintenance, pp. 337–344 (1992)
18. Pereira, C., Capeto, R.: Indicadores para avaliação de websites, TeleCentros (2002), http://www.telecentros.org/discapacitados/parametro=9861&idioma=br.html (obtained in December 15, 2011)

Implementing eHealth Services for Enhanced Pharmaceutical Care Provision: Opportunities and Challenges

Luís Velez Lapão[1,2,*], João Gregório[1], Afonso Cavaco[3,4], Miguel Mira da Silva[5], and Christian Lovis[6]

[1] International Public Health and Biostatistics, WHO Collaborating Center for Health Workforce Policy and Planning. Centro para a Malária e Doenças Tropicais, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa
Rua da Junqueira, 100. 1349-008 Lisboa Portugal
luis.lapao@ihmt.unl.pt
[2] CINTESIS - Center for Research in Information Systems and Technology in Health, Porto, Portugal
[3] Research Institute for Medicines and Pharmaceutical Sciences, Faculdade de Farmácia da Universidade de Lisboa, Portugal
[4] Department of Social Pharmacy, School of Pharmacy, University of Oslo, Norway
[5] Instituto Superior Técnico, Lisboa, Portugal
[6] Hôpitaux Universitaires de Genève, Switzerland

Abstract. ePharmaCare aims at estimating the health gains from integration of pharmaceutical services; and to assess the potential of ehealth services in the provision of cognitive pharmaceutical services. Collection of data on pharmaceutical services and diagnosis of the use of IT in Pharmacy settings was made with the application of a survey. A set of pharmacies was selected to study the pattern of services' provision supported by observational studies to evaluate the economic and clinical value of services to patients. The results of this research will allow a better understanding of pharmaceutical services' delivery and a new perspective for pharmaceutical services integration with the health system. Due to the shortage of physicians, the development of a new internet based pharmaceutical service could be an important contribution to an effective chronic diseases' management. Furthermore there is the necessity of rethinking the community pharmacy business model to effectively integrate within future health systems.

Keywords: eHealth, Integrated disease management, Enhanced pharmaceutical services, Innovation, eHealth business service models.

1 Introduction

This research aims at estimating and validating the health gains in efficiency and quality from integration of pharmaceutical services; and to assess the potential of eHealth services in the provision of cognitive pharmaceutical services.

* Corresponding author.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 433–443.
DOI: 10.1007/978-3-642-36981-0_40

It is generally recognized that community pharmacies, and their professionals, could play a more active role in healthcare systems. Healthcare reforms on primary-care, due to both the shortage of human resources and the costs associated with medicines, are starting to rethink the role of community pharmacists. The easy accessibility to pharmacies and the perceived affordability positions pharmacists at the first line of contact within the healthcare system [1]. It is of importance for science and society to understand the role of community pharmacy services within an integrated healthcare system and to research how eHealth technologies could be developed and used to address the challenges of economic and human resources crisis.

A clear demographic and epidemiological transition in Europe is leading to a growing burden of complex and chronic diseases. To trail new trends in business innovation and technological development, few pharmacy practices have started to try to shift its focus from a disease oriented to a patient oriented practice [1]. This orientation towards the patient has become the new paradigm of pharmacy practice, and has led to the development of pharmaceutical services concept, information services and the development of a clinical role for the community pharmacists. It implies a change in the focus of practice to ensure the delivery of longitudinal value added services, the raise of Pharmacy's level of responsibility and to establish cooperative relationships with other healthcare professionals. These pharmaceutical services have been considered extremely valuable for professionals and health systems, not only because there are economic savings but also due to a greater efficiency and improvement in health services quality and in patient health related outcomes. Nevertheless, this promise is yet to be accomplished.

Still, one can argue that, for these developments to be beneficial, they should be considered in an integrated and longitudinal perspective of services provision, requiring service's events registering, comprehensive analysis of the data and interactive dialogue with patients.

Innovativeness of pharmacy is positively associated with the number of pharmacy services being offered, according to Doucette et al. [2] and this is supported by having observant personnel in close contact with customers to help identify new opportunities.

Costa et al. [3] have described the status and outline the trends of community pharmacy services in the Portugal. Investing in information technologies (IT) and modernizing the architecture of pharmacies was envisioned as a necessary step. Recent political changes seem to reinforce the need to pursue the strategies already defined: to expand the traditional scope of pharmacy business to a wider range of health services, but this process has been very slow [3, 4].

It seems there is a need for IT support, in developing an extended role for community pharmacies [5]. Calabretto et al. [6] studied the socio-technical implications for information technology related interventions in community, including

medication management, and found out that we are still in the “infancy” of using IT technologies in community pharmacies. Gregório & Lapão [7] explored different scenarios for the future of community pharmacists in Portugal, with the use of IT and eHealth technologies emerging as one innovation that may develop in next years to harvest the full potential of these professionals to enhance the primary care network.

The use of eHealth is defined as “the utilization of IT to support health services’ provision, complying with the needs of citizens, patients, health professionals and other providers” [8]. The development, adoption and implementation of eHealth promises to promote a better access to information by patients and providers, improve the quality, efficacy and safety of healthcare, and to encourage healthier lifestyles. [9, 10]

The main problem addressed by this project is how pharmaceutical services in the community pharmacy could help patient management in order to improve quality of care while integrating with primary health care services. Furthermore, we want to study the actual use of technologies/eHealth for supporting services, including the study patterns of value-added services (in the context of demography and chronic diseases trends) among those delivered by a community service, and to develop a prototype of a "new patient management service model", design to improve quality and level of service, thus establishing its acceptability, feasibility, sustainability, and adaptability to future changes.

In order to address this problem, we selected the Design Science Research Methodology (DSRM). DSRM studies the connection between research and professional practice by constructing and evaluating products or artifacts that address a specific need [11]. In this way, we will also improve our knowledge about the use of DSRM to develop new innovations based on the technical, social and informational resources of a health organization.

Hevner et al. [11] have established the rules for DSRM in the form of guidelines. These guidelines support six activities, each with a specific set of tasks. In this paper, we will focus on the first and second activities necessary to start a DSRM work: diagnose of current situation and identify the problem relevance and opportunities. Thus, the objective of this research paper is to analyze and describe the utilization of IT in the community pharmacy setting, focusing on current state of pharmaceutical services provision.

The hypothesis addressed in this research is that the community pharmacist will only be able to add value and provide pharmaceutical services (as well as to communicate with patients and other health professionals involved in the disease management process) by using information technologies and eHealth.

2 Methods

An online survey was used to collect data on pharmaceutical services and to diagnose the use of IT in community pharmacy setting. An email with instructions and a link to

the survey was sent to 323 pharmacies associated with the Portuguese Association of Pharmacies. A set of four pharmacies was selected to study the patterns of pharmaceutical services provided, supported by observational studies and socio-networking analysis, which will allow the economical and clinical evaluation of pharmaceutical services [12]. To achieve our objectives, a mixed method approach was used. After a literature review, we started with an online survey to collect data on pharmaceutical services and to diagnose the use of IT in community pharmacy setting.

The survey was based on two validated surveys used in previous studies in other contexts. These surveys focused on the characterization of current pharmaceutical services provision and IT use in the community pharmacy setting, as well as exploring perceptions about possible barriers for eHealth pharmaceutical services. After translation, our survey was piloted in two pharmacies by a total of 5 pharmacists, and minor adjustments were made.

To better describe the current state and patterns of pharmaceutical services provision, we choose an exploratory observational time and motion study, using the shadowing method.

This part of the study was made in four pharmacies during a weekday's full 8 hour shift. All the pharmacies were in the metropolitan Lisbon area and were selected by convenience. Informed consent by the observed subjects was formally obtained.

The data collector was free to move and observe all the staff, but was mainly focused on the provision of pharmaceutical services. An Excel database (Microsoft Corporation®) spreadsheet on a laptop PC and a stopwatch were used to record the duration of each activity performed, as well by who, where and how it was performed.

A list of 73 possible single tasks was tested in a previous pilot observation. Due to the large number of multitasking activities it was decided to reduce the number to 51 tasks and allow the possibility of multitasking recording by adding more columns on the spreadsheet. These tasks were then validated by the pilot study's observed pharmacists, and divided in 3 broad categories (Table 1).

To diminish the possibility of affecting behavior by the staff's awareness of participating in a research study, the data collector stood at least two meters from the staff member and was informed not to initiate conversation with him or her, nor with the pharmacy's customers.

Data and Statistical analysis was conducted in Excel® and SPSS® Software Package for Social Sciences; Version 20.0. All data were kept anonymous and confidential. This study was performed in strict accordance with the good research practices and code of ethics of Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa. The study protocol was approved by the Committee on the Ethics of Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa (Permit Number: 7-2012-PN).

Table 1. Examples of tasks within the main categories

Categories	Tasks
Administrative tasks	Ordering and Storage of medicines Checking prescriptions for dispensing errors Controlling expiration dates Billing and factoring of prescriptions Meetings with vendors and salespersons
Dispensing tasks	Dispensing of prescriptions Dispensing of medical devices (bandages, Glucometers, IUD, thermometers, sphygmomanometer). OTC dispensing Home delivery
Pharmaceutical services	Administration of non-injectable medicines Provision of first aid Pharmaceutical consultation Pharmaceutical counseling on medicines and health issues Provision of screening services Production and control of compounded drugs Smoking Cessation Program Disease Management Programs Therapeutic review Veterinary pharmacy service Directly Observed therapy with Methadone

3 Results

3.1 Online Survey

The online survey had a final response rate of 4.76%, comparable with other pharmacy research surveys response rates (showing a somewhat lack of interest in participating). From this survey, we highlight the following results (figure 1):

- The average number of computers per pharmacy is 5 (min: 2; max: 10), in a ratio of 2.5 computers per pharmacist.
- The main use of the installed information systems is for dispensing medicines and administrative tasks (stock management, reimbursement activities).
- 84% of the pharmacies are satisfied with the use of Information Systems for management activities.

- 90% of the pharmacies allow access to health information resources on the internet.
- All pharmacies claim to check their email daily, although only 15% use it to answer to patients' queries, and do so less than 5 times a month.
- 23% of the pharmacies have an internet site and 38% have a site in some social network (ex: Facebook®).

All the pharmacies considered important or very important, the use of information systems for the management of the pharmacies and for the provision of pharmaceutical services. Also, more than 60% of these pharmacies do not provide any disease management, pharmaceutical care or smoking cessation programs. The dispensing of medicines and the collection of used medicines are the only services that all the pharmacies provided. 38.5% of the pharmacies refer that more than 60% of the customers have chronic conditions. From a set of possible new services, the ones related to the provision of information (medicine information kiosk, health promotion campaigns and thematic workshops, sms alerts) were the ones that pharmacies would like to provide the most in the future.

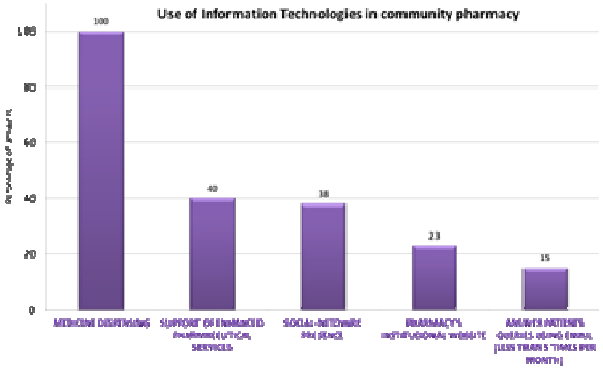


Fig. 1. Use of IT systems in the community pharmacy setting

Also important to highlight are the main barriers to the implementation of eHealth solutions in Community Pharmacy. The respondents claim that no financial incentive for the implementation of eHealth solutions and not enough support and guidance from professional organizations are the most important barriers (figure 2).

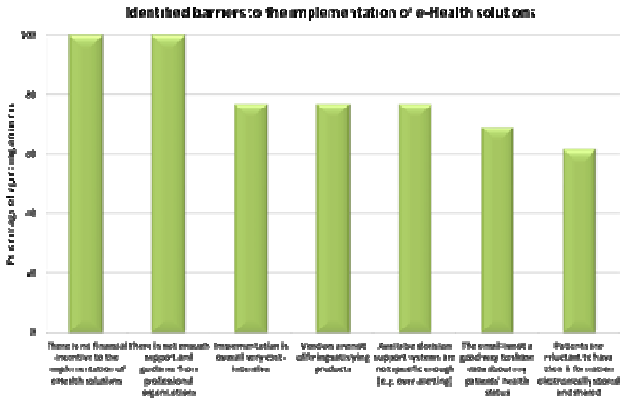


Fig. 2. Most important barriers identified in the online survey

3.2 Observational Study

The observational study took place in four pharmacies. In aggregate, 108h of time were recorded including all the staff's tasks. A total of 894 tasks were recorded (average of 223 per pharmacy) (table 2). Average total aggregated time observed per pharmacy was 27h10min (min: 21h14m; max: 31h 34m). 85% of the tasks were performed by pharmacists, equivalent to 65% of the total recorded time (figure 3). An average of 481 minutes per pharmacy (76 minutes per employee) is not used in any tasks.

Table 2. Professionals and tasks observed

	Number of individuals	Observed time (hh:mm:ss)	Number of Tasks (and %)
Pharmacy technician	4	12:24:17	85 (9,5)
Pharmacy assistant interns	3	17:46:43	21 (2,4)
Other professionals	2	7:29:55	23 (2,6)
Pharmacists	16	70:50:21	765 (85,6)
Total	25	108:31:16	894 (100)

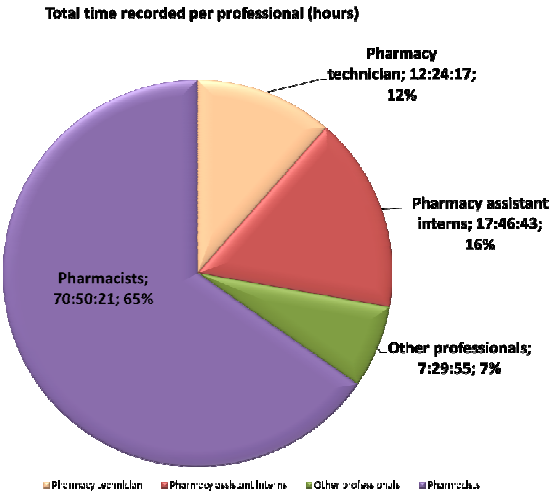


Fig. 3. Distribution of total observed time per professional

Concerning activities performed by pharmacists, almost 50 % of their time is used in interaction with pharmacy customers and 38% is used in administrative tasks, including ordering and storage of medicines, checking for errors in the dispensed prescriptions, preparing prescriptions for reimbursement issues and meetings with vendors and salespersons. Considering that 20% of the unused time is spent in socialization among workers and other minor tasks (ex: "surfing the net", coffee-breaks, etc), we calculated an average of 38,40 minutes of available time per pharmacist (figure 4).

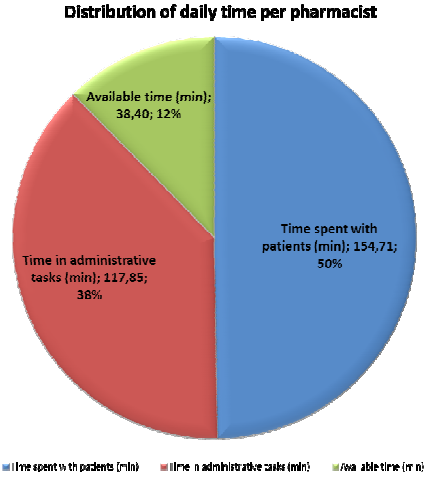


Fig. 4. Distribution of total activity time per pharmacist

When considering the medicine dispensing process only, 35.4% of customers received some form of pharmaceutical counselling, that usually involved provision of therapeutic information: for whom is it, what is it for, daily dosages, possible side-effects that can happen or might have happened in the past, market availability and prices. In 44.6% of the observed dispenses, at least one chronic use medicine was dispensed. When a chronic use medicine was dispensed, 29.3% of patients got pharmaceutical counselling. Instead, when a non-chronic medicine was dispensed, 46.7% of patients got this counselling.

In the recorded observations, the only pharmaceutical service provided with the use of an information system (IS) was the dispensing service. It was used to support the sale, while detecting possible interactions and identifying the main adverse effects of the medicines dispensed. The pharmacy's IS was used in 95.6% of medicine dispensing processes, in 80.4% of patients queries about health issues and medicine availability and in 27.2% of administrative tasks.

4 Discussion

The results thus far allow for a good characterization of pharmaceutical services' provision. Although there was a low response rate, showing a lack of interest in participating, the results from the survey suggest that the potential to develop web-based services is there, since all the pharmacies have an information system installed and the staffs are capacitated to use it. The use of IT system for patient management, the presence in web-based social networks and the use of email to communicate with patients are possible indicators that pharmacies are moving to a more web-based approach to provide information and services.

The main barriers found for the development of IT solutions for the community pharmacy may reflect a biased vision of what a true eHealth solution might be. The perception that "the email is not a good way to share data about my patient's health status" may indicate some renitence in using a web platform to share information that some may consider sensitive. Also, the fact that all the respondents claim that they need more support from the professional organizations may point to the need of a strong leadership in order to implement such services.

As it was expected, the dispensing of medicines was the most significant service provided. Although some pharmacies report to provide a medication management service, it was without surprise that we observed no structured medication management service or any other disease management service. One possibility for this is that neither the medication management is really provided or at best is a very informal service. We can discuss if it is a service or activity, since it is made with no registrations and done verbally in a non-systematic way. The causes for this have long been discussed within the profession and the reasons range from the lack of payment for these services to the professional inertia regarding the adoption of new forms of practice. It is not in the scope of this research to investigate such reasons, but this work may provide new clues in order to better understand the best strategy for incrementing the provision of enhanced pharmaceutical services.

The available time we have found is an indication that there may be an opportunity to develop web-based services. Despite such notice, one of the barriers pointed out to this development has been the "lack of time" reported by pharmacists. It would be important to find the reasons behind this gap of perception. One possibility may be the fact that the pharmacist spends almost 40% of the time in administrative tasks that prevent the professional to take a more clinical role.

Another question that arises from the observational study findings is the necessity to reorganize the internal functioning of community pharmacies, with more clear roles for each of the professionals.

Weather these professionals do have the necessary education regarding the provision of enhanced pharmaceutical services, on a web-platform is an aspect that needs to be further explored. The social abilities that these professionals do have may be used as complementary tools to better interact with patients in a new web-based service. This can be important considering that customer relationships is one of the key aspects of the new services provided through the internet.

Furthermore, since pharmacy services suffer from lack of change and inefficiency, which affects the delivery of sustainable services, the introduction of other channels can create new opportunities to improve services. In this regard, the use of methods based on DEMO to find non value-added transactions can be also used in the pharmacy context to redesign and simplify processes [13]. Modeling methodologies can provide an additional understanding of the dynamics of an organization and to allow a better alignment between the pharmacy services design and operation [14].

5 Conclusion

The evolution of information technologies is developing, and probably will continue to develop, an important opportunity towards the definition of new roles for community pharmacists. The use of "smart" technological solutions in the medicine dispensing process could relief pharmacists' workload, leaving more free time to assume other functions, boosting the possibility to accomplish such enhanced pharmacy care provision. It is likely that the introduction of new information technologies will face some resistance, especially from older professionals, but the potential value of these technologies to improve the health system efficiency (and the interaction with patients), will make pharmaceutical-related expertise and knowledge a very important competence and a precious asset in the healthcare market.

Another possible conclusion of this study is the necessity to rethink the community pharmacy business model in order to effectively and coherently integrate it into the future health system models.

Acknowledgements. The ePharmacare research project is funded by Fundação para a Ciência e Tecnologia (2012–2014) under the number PTDC/CCI-CIN/122690/2010.

References

[1] Chisholm-Burns, M.A., Graff Zivin, J.S., Lee, J.K., Spivey, C.A., Slack, M., Herrier, R.N.: Economic effects of pharmacists on health outcomes in the United States: A systematic review. Am. J. Health Syst. Pharm. 67(19), 1624–1634 (2010)

[2] Doucette, W.R., Kreling, D.H., Schommer, J.C., Gaither, C.A., Mott, D.A., Pedersen, C.A.: Evaluation of Community Pharmacy Service Mix: Evidence From the 2004 National Pharmacist Workforce Study. J. Am. Pharm. Assoc. 46(3), 348–355 (2006)

[3] Costa, S., Santos, C., Silveira, J.: Community pharmacy services in Portugal. Ann. Pharmacother 40, 2228–2234 (2006)

[4] Gallagher, R.M., Gallagher, H.C.: Improving the working relationship between doctors and pharmacists: is inter-professional education the answer? Adv. Health Sci. Educ. Theory Pract. (2010)

[5] George, P.P., Molina, J.A.D., Cheah, J., Chan, S.C., Lim, B.P.: Evolving Roles of the Community Pharmacist – A Literature Review. Annals Academy of Medicine 39(11) (2010)

[6] Calabretto, J.P., Warren, J., Darzanos, K., Fry, B.: Building common ground for communication between patients and community pharmacists with an internet medicine cabinet. In: 35th Hawaii International Conference on System Sciences (2002)

[7] Gregório, J., Lapão, L.V.: Portuguese community pharmacists in 2020: survival of the fittest. Revista Portuguesa de Farmácia 52(5 suppl.), 33–39 (2011)

[8] Callens, S.: The EU legal framework on ehealth. In: Mossialos, E., Permanand, G., Baeten, R., Hervey, T. (eds.) Health Systems Governance in Europe: The Role of EU Law and Policy. Cambridge University Press, Cambridge (2010)

[9] Bates, D.W., Gawande, A.A.: Improving safety with information technology. N. Engl. J. Med. 348(25), 2526–2534 (2003)

[10] Neuhauser, L., Kreps, G.L.: eHealth communication and behavior change: promise and performance. Social Semiotics 20(1), 9–27 (2010)

[11] Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28(1), 75–105 (2004)

[12] Mache, S., Scutaru, C., Vitzthum, K., Quarcoo, D., Schöffel, N., Welte, T.: Does type of hospital ownership influence physicians' daily workschedules? An observational real-time study in German hospital departments. Hum. Resour. Health 7(41) (2009)

[13] Dias, D., Lapão, L.V., Mira da Silva, M.: Using Enterprise Ontology for Improving Emergency Management in Hospitals. Stud. Health Technol. Inform. 180, 58–62 (2012)

[14] Gomes, P., Lapão, L.V.: The Role of a Nation-Wide Information System in Improving the Access to Surgery in Portugal. Stud. Health Technol. Inform. (2009) 10, 978-1

Establishing Multi-model Environments to Improve Organizational Software Processes

Muñoz Mirna and Mejia Jezreel

Centro de Investigación en Matemáticas
Avda. Universidad no. 222, 98068, Zacatecas, México
{mirna.munoz,jmejia}@cimat.mx

Abstract. Organizations worldwide are adopting several international standards and models, in an effort to improve the software processes, however, the difficulty by using multi-model improvement environments in the implementation of successful process improvements is well known. This paper shows the steps in order to establish a multi-model Environment based on the organization's business goals. It includes the results of a case study on project management process improvement in an organization.

Keywords: process improvement, multi-model environment, project management, organizational business goals.

1 Introduction

Nowadays many organizations are under mandatory or market pressure to use more than one improvement model [1]. According to Lisa Marino and John Morley [2], three out of every five large organizations are already facing the challenges of using multiple models to meet the organizational business goals.

As a result, organizations worldwide are adopting several international standards and models, such as Capability Maturity Model Integration (CMMI) or ISO 15504, in an effort to improve the processes used to manage their businesses, increase customer satisfaction and maintain a competitive advantage [3].

However, the difficulty of implementing successfully process improvement by using multi-model improvement environments as reference is well known. A first step in the integration of different models and standards is to recognize that despite the different structures and terminologies and the different levels of abstractions, the standards and models used in the organization share common element types. Therefore, the challenge is to examine the models and standards to identify the common elements and to organize the implementation of these standards and models in the organization's processes based on these common elements [4, 5].

However, it is important to take into account the idea of George Box's: "*all models are wrong, but some are useful"* which is still valid for Mogilensky and Christian [6]. In other words, the models are not mutually exclusive even when each one offers unique features and addresses specific problems. Therefore, the purpose of this paper is showing the steps to establish the multi-model enviroment.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 445–454.
DOI: 10.1007/978-3-642-36981-0_41

This paper is organized as follows, section 2 gives a brief description on multi-model environment; section 3 describes the steps for establishing the multi-model environment; section 4 addresses the case study and finally section 5 presents the conclusions.

2 Multi-model Environment

Taking into account that the multi-model environment emerged as a result of a common effort of the organizations to integrate international standards and models to achieve successful software process improvement [6, 7], such as, Capability Maturity Model and Integration for Development (CMMI-DEV) [8]; Team Software Process (TSP) [9]; Project Management Body of Knowledge (PMBOK)[10]; ISO/IEC15504 Information technology – Process assessment [11]; ISO 9001:2000- Quality Management System [12]; ISO/IEC 12207-2008 [13]. A multi-model environment involves all cultural aspects and the knowledge that makes advisable to use in each process a mix of best practices from more than one model or standard to achieve the organization's business goals.

According to Lawrence y Becker [5], the objectives of a multi-model environment are: reduce redundancy, improve integration, create synergy, leverage best practices and make frameworks transparent.

3 Establish Multi-model Environment

The purpose of this research work is to show the steps that should be performed in order to establish the multi-model environment. These steps use a top-down approach to establish the multi-model environment and consist on the next steps: 1) Identify organizational business goals; 2) Analyze the formal processes documentation, 3) Assess organizational current practices performance, 4) Prioritize process area 5) Select models and standards, 6) Analyze external best practices and 7) Identify dependencies.

Next, each step is described.

3.1 Identify Business Goals

The purpose of this step is to identify, formalize and understand the business goals in order to understand the organization's needs.

The organization's needs should be formalized in order to obtain the commitment of the senior management to support activities and providing all the resources needed for starting the software process improvement. This step includes two activities:

1. *Identify business goals.* Activity that analyzes the business goals and the documentation related to this.
2. *Formalize the business goals.* Activity that conducts meetings to establish and formalize the organizational business goals.

In this step, the Goal Question Metrics (GQM) methodology can be implemented in order to identify and formalize the business goals. The GQM is a top-down approach

that starts by making questions to identify organizational goals and concludes by identifying the metrics that answer to the questions made [14].

One of the reasons for the GQM success is that it is adaptable to many different size organizations and environments, as confirmed by the large number of companies that have used it.

3.2 Analyze the Formal Processes Documentation

The purpose of this step is to collect and analyze the organization's processes documentation. Moreover, the analysis carried out in this activity will help to identify the lack of the organization's processes documentation. This activity includes 3 tasks:

- *Collect the organization's processes documentation:* request and collect the organization's processes documentation.
- *Analyze the organization's processes documentation:* analyze the organization's processes documentation to identify what activities are documented.
- *Document findings:* make a list with findings of the organization's processes documentation. This list will help to improve the organization's processes documentation.

3.3 Assess Organizational Current Practices Performance

Once the business goals and findings in organizational processes are identified, the organizational performance should be assessed in order to know the organizational performance toward the achievement of the current business goals. To achieve this, the follow activities are performed [15]:

3.3.1 Analyze Coverage

Activity focused on analyzing the coverage of the indicators and business goals. The correlation analysis will identify which business goals are really covered by indicators. This activity includes 2 tasks:

- *Analyze coverage of indicators and business goals:* identify the correlation of the indicators related to business goals.
- *Collect current values of the indicators:* collect and analyze data that contains established values (planned) indicators related to organizational business goals.

3.3.2 Analyze the Achievement

This activity is focused on analyzing the achievement of business goals and its related indicators through making a compliance matrix. This activity includes 4 tasks:

- *Make achievement matrix*: perform the compliance matrix where the columns record the business goals and the rows their related indicators. Then, fill it with the established values (planned) and current values assigned for each indicator.
- *Analyze achievement matrix:* calculate the achievement of the indicators through comparing the established values (planned) and current values with the formula: *achievement = current value - planned value*. This result allows establishing the

achievement degree of indicators related to business goals, therefore, the achievement of bussines goals.
- *Establish process performance*: analyze and select the indicators that have the highest and lowest coverage.

3.3.3 Prioritize the Business Goals to Be Achieved

This activity is focused on prioritizing indicators related to business goals through the assignment of weights according to their impact in the achievement of business goals and criteria established by senior management.

3.4 Prioritize Process Area

The purpose of this step is to obtain and establish the priorization of the process areas according to the organization. This step includes one activity:

3.4.1 Preliminary Identification of Priority Processs Areas

This identification is carried out by analyzing the business goals, the organizational performance and the analysis of the models or standards to be applied within the organization. This step will focus the process improvement effort toward the high priority organizational process areas.

3.5 Select Models and Standards

The purpose of this step is focused on selecting the models and the standards. The models and standards are selected focusing on those: 1) which contain the target processes and 2) which have a wide use in the target processes. This activity includes 3 tasks:

- *Collect information about models and standards used in the organization:* identify and collect information of models and standards used or known by the organization's staff.
- *Identify models and standards:* identify the models and standards used by other organizations.
- *Select models and standards to be analyzed:* select the processes to be analyzed based on the prioritization of indicators related to the business goals, how the organization's work and the models and standards used by the organization.

3.6 Analyze External Best Practices

The purpose of this step is to establish a multi-model environment based on the models and standards selected. The multi-model environment will be used as a reference for the implementation of external best practices allowing to choose those practices that best fit the way the organization works and make its processes more efficient.

In order to establish the multi-model environment, the multi-model approach proposes to use the business indicators and the reference model as inputs for selecting and analyzing the external models and standards. Once the business goals and its related business indicators have been identified, models and standards that best fit with the way the organization works are selected.

Then, the selected reference model will enable to know what to do in order to achieve the business goals with high priority, and the selected models and standards will provide the best practices, which will indicate "how to do it" in order to achieve the business goals with high priority.

As a result, a set of external practices that: 1) have a proved performance; 2) are promoted by the most widespread standards and models; and 3) best fit the way the organization works; are obtained and organized in a structured way. This step includes five activities [15]:

3.6.1 Choose the Reference Model

Before studying models and standards it is essential to know where to address the research effort. Therefore, it is necessary to choose the model that best serves as reference to address this effort. To achieve it, models and standards structures are mapping, then, the model that has wider coverage in relation to target is selected. This activity includes 3 tasks: 1) *Identify the process 2) Analyze the processes and 3) Select the processes*.

3.6.2 Establish the Detail Level

Studying models and standards implies handling large amounts of information. Usually the structure of the information used by models is different. Therefore, the difficulty of determining the appropriate level of detail should be kept in mind. It is important to take into account that a high level analysis may not provide enough insight into similarities and differences and a low level analysis could result in an overwhelming number of interrelationships which also fails to properly identify the correspondence among models [16]. In this context, making a glossary based on the analysis of the information contained in the models and standards will be helpful. The glossary will establish how one element is referenced in each model.

3.6.3 Create a Correspondence Template

In order to establish an appropriate comparison among models and standards, a template is created. For this research "template" is the table in which are formalized the process elements such as inputs, subpractices, tools and techniques, work products, and informative components.

3.6.4 Identify the Similarity among Models

The similarity is the correspondence among the information provided about how to carry out an activity with different models. The similarity allows us to establish what information in a model can strengthen the other models.

3.6.5 Show the Obtained Results

An important factor to understand the obtained results is the way they are shown. In order to show this kind of results tables were used as a format because the information is stored in a structured way.

The resulted tables (the multi-model environment) should contain the practices candidates to be included in the processes. So that, the multi-model environment indicates what external best practices, which performance is proved, could be executed within the organization to make its processes more efficient.

3.7 Identify Dependencies

The purpose of this step is to identify the explicit dependencies from the model or standard that was established as reference model. As a result, the implementation sequences, relationships and dependencies from priority process areas are defined. This step includes 3 activities [17]:

3.7.1 Identify Dependencies

A matrix of dependencies among process areas is elaborated. The existing dependencies among process areas are identified through reviewing process areas included on the official bibliography version of the reference model. Then, a matrix of dependencies is elaborated and filled with the dependencies found in reference model.

3.7.2 Analyze Dependencies

In this activity, the strongly connected components (SCC) are verified and the cyclic and SCC groups are selected.

3.7.3 Verify the Strongly Connected Components

The dependencies are evaluated using a mathematical software tool [18] in order to check the SCCs. After executing the mathematical software tool, the SCC groups are obtained. From the SCC groups obtained, the individual groups must be discarded because of its triviality. Then, the SCC group that contains the largest number of processes is selected. This activity includes 2 tasks: 1) *Generate combinations and 2) Cyclic groups*.

3.7.4 Determine the Implementation Sequence

The formal implementation sequence is proposed. According to the cyclical clusters obtained, the implementation sequence cannot be implemented without the implementation of all processes that constitute the cyclical cluster. In order to identify the processes implementation sequence, permutations for each cyclical cluster are generated. A permutation was sorted from higher to lower number of source dependencies.

4 Case Study

This section presents the experimentation of the method carried out in *everis* in order to identify establish a multi-model reference model focuses on its project management best practices.

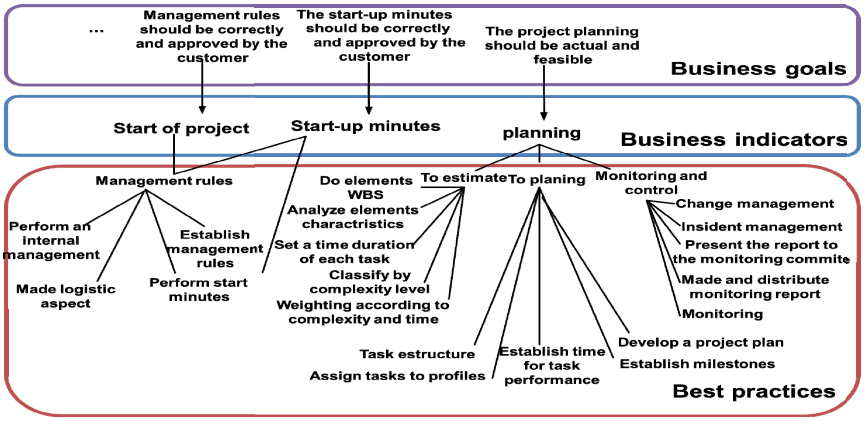
everis is a multinational consulting firm with factories in Europe and Latin America. It offers services which provide solutions to large companies in any sector and it is based on three pillars: innovation, methodologies and efficiency. Since its creation in 1996, it has grown both in revenue and staff in a steady and organic way.

The steps were validated by performing meetings in which all material of the steps were explained to everis' quality and methodology group. Besides, the steps must approve a checklist applied by this group. These meetings allowed to get feedback that was used for improving the task and activities described in each step of the method. Once the steps were validated, the scope of the experimentation was focused in *everis'* project management processes because of everis had the need of improving its

project management processes due to they have a broad impact on the organizational business goals.

Next, the main work products obtained of performing the steps for the establishment of multi-model environment in everis to be used as reference in the implementation of software process improvement are listed.

- Three of the five business goals identified after implementing GQM methodology were: 1) to have a better performance of using project management project management processes; 2) to have a better project performance and 3) to have a better managers satisfaction. Others business goals can be mentioned because of confidentiality.
- The findings in the documentation were grouped on 4 aspects: 1) presentation of process documentation (there is a lack of standardization in the process definition; not all processes have a complete definition, and some process maps do not correspond with the activities description); 2) process content (some process do not correspond with the activities contained on it or have inconsistences); 3) process metrics (too many metrics focused on audits) and 4) others (some activities description are ambiguous).
- The results of the coverage analysis are showed in Figure 1.



- The results of analyzing the achievement are showed in Table 1, to have a better understanding of business goal prioritize, a semaphore was used where red= low coverage; yellow= medium coverage and green= high coverage). Then, the performance can be established by analyzing the color or the semaphore assigned.

Table 1. Achievement analysis

Business goal	Indicator	Coverage
to have a better performance of using project management project management processes	Management rules	10,3%
to have a better project performance	Project planning	4,3%
to have a better managers satisfaction	Start-up minutes	10,3%

- Then, the prioritize of business goals was established by applying criteria established by senior management such as *The management rules are a key document in everis because they define the project framework.* As a result, the indicators related to business goals were: planning tools, management rules, and start-up minutes.
- Interviews with projects managers selected as success case provider were carried out to get information about the models and standards used in the organization. Besides, there were performed interviews with senior manager too. The main models and standards identified in the organization were: CMMI-DEV, PMBOOK, PRINCE2, TSP, COBIT, ISO9001, and ISO/IEC 15504
- The results of applying the activities proposed on the step select models and standards are showed in Table 2. (See section 3.6 and its subsections)

Table 2. Achievement analysis

<table>
<tr><th>Activity</th><th>Work product</th></tr>
<tr><td>3.6.1</td><td>CMMI-DEV v1.2 model was chosen as the reference model because it provides a way to manage an integrated approach to development activities as part of achieving their business objectives.
Besides, because of most of the prioritized business indicators were related to project planning, the project planning process was chosen because it is considered critical to successful project management</td></tr>
<tr><td>3.6.2</td><td>After the information was analyzed and the glossary was prepared, it was decided to establish the level of mapping at specific practice</td></tr>
<tr><td>3.6.3</td><td>A template was designed based on the CMMI-DEV structure. The CMMI-DEV structure includes work products, subpractices and informative components. Other items, such as inputs, tools and techniques, were later added from the other models and standards</td></tr>
<tr><td>3.6.4 and 3.6.5</td><td>The similarities among standards and models were identified. Next an example of a table obtained is showed.
<table>
<tr><th colspan="4">PP SP 2.2 Identify Project risk</th></tr>
<tr><th>Inputs</th><th>Subpractices</th><th>Tools and techniques</th><th>Work products</th></tr>
<tr><td>• Historical risk management plan
• Project planning outputs
• Risk categories (project management, organizational and external risk)
• Historical information
…</td><td>SP2.2.1 Identify risks
…</td><td>T2.2.1 Structured interviews
T2.2.2 Brainstorming
T2.2.3 Documents review
T2.2.4 Delphi
T2.2.5 Strengths, weaknesses, opportunities, and threats (SOWT) analysis
T2.2.6 Risk taxonomies
…</td><td>Identified risks
Risk impacts and probability of occurrence
Risk priorities
Triggers
Overall risk ranking for the project
Trends in qualitative risk analysis result
ITL (Issue Tracking Log) form</td></tr>
<tr><th colspan="4">Informative Components</th></tr>
<tr><td colspan="4">•Participants in risk identification generally include as possible: project team, risk management identification, subject matter experts from others parts of the company, customers, end users, other Project managers, stakeholders, and outside experts
•The organization must plan and develop the needed procedures for product realization. The procedures must be consistent with the quality management plan
•The risk identification process should include qualitative and, where possible, quantitative risk ranking and should obtain input from management brainstorming
•The risk assessment should consider business, regulatory, legal, technology, trading partner and human resources risks
•Allocate to each high risk or critical activity a resource in which management has confidence
•Identify risk to the project both initially within the project strategy and as they develop during the conduct of the project
…</td></tr>
</table>
</td></tr>
</table>

- Dependences between internal and external best practices where identified and analyzed, then, the implementation sequence was established.

The obtained multi-model environment was proved by implementing a software process improvement. The improved processes were established containing a set of external practices from the multi-model environment. The new processes were grouped into a project management method, because everis needed to develop a project management method as a part of its Corporate Methods methodology (COM).

Then, the COM method was launched and results focused on process use were analyzed. The data were collecting at three times of change: before (2007) during (2008); and after (2009). Besides, it was grouped to make a better analysis as follows: before and during the change period on group I and after the change period on group II.

The process use aims to analyze the process acceptance by users. The measure data were obtained by analyzing surveys carried out by managers involved in project planning. Figure 1 shows the control chart with the results obtained by comparing the percentage of managers that use COM in group I and group II.

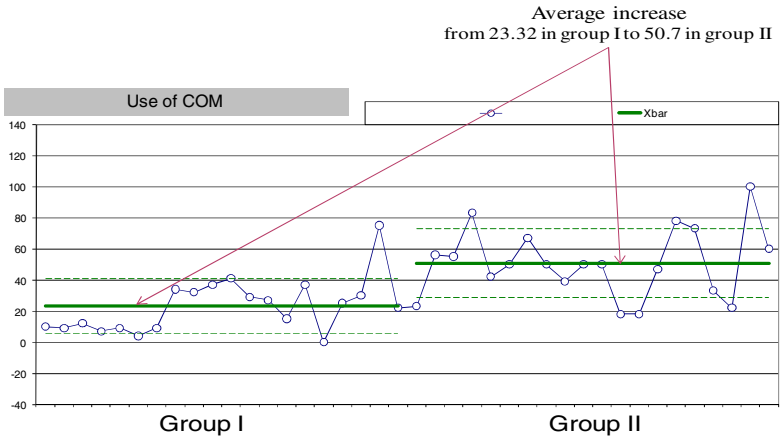


Fig. 1. Improvement of the percentage of managers that use COM project management method

The control chart (See Fig.1) of the use of COM shows an improvement because there is a gradual and continuous increase in the average of managers that use the COM method for managing their projects from 23.32 in group I to 50.7 in group II.

5 Conclusion

The business goal identification is a key issue in the implementation of a software process improvement because it allows understanding the organization's needs. Moreover, the formalization of the organization's needs allows obtain the commitment of the senior management in order to provide all the resources needed for the start of the software process improvement.

Besides, the implementation of a multi-model environment allows getting a better acceptation of new processes by process users because new processes reflect the way

they work and contain external practices with a proved performance that best fit with the way the organization works having more efficient processes.

The experience presented in this paper shows that implementing a software process improvement by establishing and using a multi-model environment as reference model has allowed the adoption of external practices which best fit with the organization and help to achieve the organization' business goals.

References

[1] SuZ, G.: Process Improvement At the Edges. Software Engineering Institute / Carnegie Mellon University (2007)
[2] Lisa, M., John, M.: Process Improvement in a Multi-model Environment Builds Resilient Organizations (2009)
[3] Urs, H.A., Patrick, K.: A Unified Process Improvement Approach for Multi-Model Improvement Environments (2009)
[4] Kirwan, P., Andelfinger, U., Sassenburg, H., Heijstek, A.: A Unified Process Improvement Approach for Multi-model Improvement Environments. Software Engineering Institute (SEI). Carnegie Mellow University (2004), http://www.sei.cmu.edu/news-at-sei/features/2006/04/feature-1-2006-04.htm
[5] Lawrence, J., Becker, N.: Implementing Insights and Lesson Learned Using ISF for Excellence. In: SEPG 2009 North America Conference (2009)
[6] Mogilensky, J.: Pathological Box-Checking: The Dark Side of Process Improvement. In: SEPG 2009 North America Conference (2009)
[7] Kirwan, P., Jeannie, S., Lisa, M., John, M.: Improvement Technology Classification and Composition in Multi-model Environments. In: Software Engineering Institute (SEI). Carnegie Mellon University (2008)
[8] SEI CMMI Production Team, S.: CMMI for Development v1.3. Software Engineering Institute (January 2011)
[9] Humphrey, W.: Introduction to the Team Software Process, p. 459. Addi-son-Wesley, Massachusetts (2006)
[10] IEEE Computer Society: IEEE Guide Adoption of PMI Standard A Guide to the Project Management Body of Knowledge; IEEE Guide Adoption of PMI Standard A Guide to the Project Management Body of Knowledge, IEEE Std 1490-2003 (Revision of IEEE Std 1490-1998), pp. 0_1-216 (2004)
[11] International Organization for Standardization: ISO/IEC 15504: 2004 Information technology – Process assessment (2004)
[12] AENOR: Sistemas de Gestión de Calidad (ISO 9001:2000), UNE-EN ISO 9001, Norma Española (2000)
[13] International Organization for Standardization: Systems and Software Engineering - Software Life Cycle Processes, IEEE STD 12207-2008, Ieeexplorer, pp. cl-138 (2008)
[14] Basili, V.R.: Software modeling and measurement: the Goal/Question/Metric paradigm. University of Maryland at College Park College Park, MD (1992)
[15] Calvo-Manzano, J., Cuevas, G., Gómez, G., Mejia, J., Muñoz, M., San Feliu, T., Sánchez, A.: Methodology for Process Improvement through Basic Components and Focusing on the Resistance to Change. In: Eurospi (2009)
[16] Jose A., C.-M., Gonzalo, C., Jezreel, M., Mirna, M., Tomás, S.F., Ángel, S., Álvaro, R.: Approach to Identify Internal Best Practices in a Software Organization. In: Riel, A., O'Connor, R., Tichkiewitch, S., Messnarz, R. (eds.) EuroSPI 2010. CCIS, vol. 99, pp. 107–118. Springer, Heidelberg (2010)
[17] Calvo-Manzano, J.A., Cuevas, G., Gómez, G., Mejia, J., Muñoz, M., San Feliu, T., Sánchez, Á.: Processes Implementation Sequences to Achieve the Maturity Level 2 in CMMI-ACQ. In: 16th EuroSPI (2009)
[18] Wolfram Mathematica. Documentation Center (2008)

Standardization of Processes Applying CMMI Best Practices

Vítor Serrano[1], Anabela Tereso[1], Pedro Ribeiro[2], and Miguel Brito[2]

[1] Department of Production and Systems Engineering
CGIT Research Centre
[2] Department of Information Systems
Centro ALGORITMI*
University of Minho, 4800-058 Guimarães, Portugal
pg16496@alunos.uminho.pt, anabelat@dps.uminho.pt,
{pmgar,mab}@dsi.uminho.pt

Abstract. Capability Maturity Model Integration (CMMI) is a set of practices that can be applied in companies in order to improve processes. The goal of this work was to understand the barriers in implementing CMMI and improve processes following the model. The study was conducted in a company in the North of Portugal, following three steps: (1) diagnosing the company (2) opinion gathering through questionnaires and (3) reimplementation of CMMI. The analysis of the questionnaires indicated that implementing CMMI is problematical due to bureaucracy and lack of detailed protocols. Based on the difficulties encountered we developed a detailed documentation with standard processes where the employer has a more prominent role in controlling the processes. The adapted CMMI was then re-introduced in the same company. Finally, linking perceptions and results from the reimplementation, we consider fundamental a good use of CMMI to ensure efficient production.

Keywords: CMM (Capability Maturity Model), CMMI (Capability Maturity Model Integration), Reimplementation, Process Improvement.

1 Introduction

The current study aimed to implement CMMI in a technical department of a company in the North of Portugal.

Nowadays, most companies have great care in defining a set of suitable practices for the successful development of their products. It is assumed that a company must have a set of defined and standardized processes that organize the work done by their employees. However, bustle in the industrial world leads sometimes to a lesser concern when developing a product, without deepening all its essential aspects. This may lead to higher costs and to customer dissatisfaction.

* This work is financed with FEDER funds by COMPETE (*Programa Operacional Fatores de Competitividade)* and with National Funds by FCT (*Fundação para a Ciência e Tecnologia*) in the scope of the project: FCOMP-01-0124-FEDER-022674.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 455–467.
DOI: 10.1007/978-3-642-36981-0_42

The study was designed to understand why the CMM project implemented on the past in the company failed and how could the model be improved.

Companies aim to complete the tasks that meet the customers' needs, within time and at a lowest possible cost [1]. Proper project management is central to achieve the aforementioned goals [2]. That is the reason why good practice models focus their study on improving the project management area.

Regarding process management, CMMI seeks that a company achieves high maturity in specific subjects. This model goes beyond a set of mandatory rules and looks for approaches to improve the development and maintenance of the product [3]. The practices are applied throughout the life cycle of the product, from initial design until delivery to the final consumer [2].

According to current literature, the use of CMMI processes has several advantages. Sun and Liu emphasized three benefits: (1) description of processes' requirements, (2) presence of an integration method with prioritization of those requirements and (3) continuous improvement practice [4]. The CMMI is a model that is difficult to apply to organizations because it is a process that requires an involvement of the entire organization [5]. However, as pointed out by Huang and Han, CMMI creates good practices within a company, but the model implementation is not user friendly [6]. When implementing CMMI, managers are faced with the difficulty of defining the priority areas. Applying a formal CMMI model requires a great effort because it can be complex and time-consuming.

A crucial step on the path of business success is achieving maturity. In this sense, there are maturity levels which correspond to process areas[1] [7]. In line with this approach, quality is achieved through continuous improvement of processes.

2 Method

To better understand the action-research method used on this study we will start by explaining its fundamental characteristics: definition, key principles, types, tools and the role of the researcher. We will then describe the stages of the experimental research done: 'Diagnosing the company', 'Opinion gathering through questionnaires' and 'CMMI model reimplementation'.

2.1 Action-Research

The action-research method implies learning throughout its application. After identifying a problem, multidisciplinary teams should be created to analyze the problem. From this analysis, an action plan is designed that will be implemented later. Actions are taken after a team analysis and followed by an evaluation of the outcome. These cycles are repeated until a solution is reached (Fig. 1). These methods entail

[1] "Process area: A cluster of related practices in an area that, when implemented collectively, satisfies a set of goals considered important for making improvement in that area. (CMMi v1.3 glossary)".

several advantages for the companies. A major advantage is the increase on team motivation due to greater involvement in problem solving. Furthermore, solutions are easy to implement since they meet the suggestions provided by all [8].

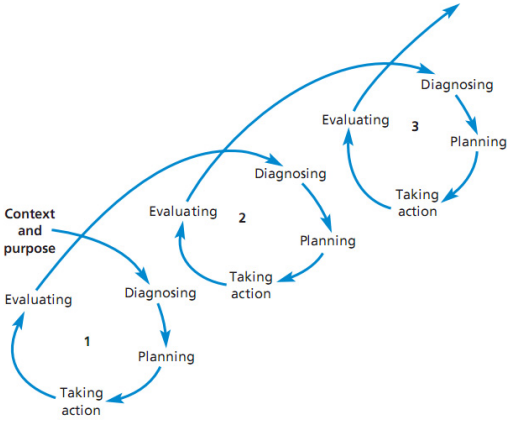


Fig. 1. The action research spiral [8]

2.2 Stages of the Experimental Research

The action-research method was split in three stages. During the first stage we diagnosed the company by verifying the processes description and how the processes are being executed. On the second stage, we administer two questionnaires. The first questionnaire was administered to all employees of the department under study that worked with CMM. The main goal of this questionnaire was to detect and understand the major drawbacks of CMM implementation. With this data we could identify the steps that need to be improved. The second questionnaire was completed by Portuguese companies licensed by SEI (Software Engineering Institute). With this data we wanted to collect the opinion about the model, its implementation and its acceptance.

The third stage consisted in executing an improved version of CMMI that should be a useful tool and an asset to the department under consideration.

First Stage – Diagnosing the Company. To analyze the current situation of the company we examined the existing documents. The entire structure follows the CMM applied in the past. The data regarding the processes was spread and organized differently in various areas. This required different strategies to collect the information. The strategy plan took in consideration the following:

- What is the department process?
- What are the requirements for the process?
- How to manage the software?
- How to manage the resources?
- How to manage a project?

Flowcharts were created for each of the above points, in order to achieve better visualization. So this way, we seek to identify possible improvements and we were able to verify the overall operation of the process in the department (inputs, department functions and outputs).

Second Stage – Administering the Questionnaires. In order to understand why the CMM implementation in the past failed, we considered fundamental to understand the team barriers for the model execution. For that purpose, we administered a questionnaire to nine team workers with past experience in CMM. The questionnaire had four questions (defined after a brainstorm with several team workers) and its main goal was to recognize their insights about the model.

Another questionnaire (with five questions) was sent to ten companies by email after phone contact.

All the answers were analyzed and taken into consideration for CMMI reimplementation.

Third Stage – Reimplementation of CMMI Model

This last stage was used to develop and apply new procedures. All the new procedures are being applied, according to the company plan, with the goal to facilitate the work of the employees and to ensure the best practices of CMMI. A group with three staff members was organized to develop new procedures. Regular meetings were held between this group, with occasional collaboration of the remaining team. These procedures were then introduced and evaluated, and their effectiveness was constantly checked through regular meetings with the employees involved. These regular checkpoints contributed to a sustained improvement of the new procedures.

3 Results

3.1 Questionnaire to Company Employees (Questionnaire 1)

In the company under study, previous CMM implementations were not a success. In some projects, the CMM was not even applied. Therefore it was necessary to identify the problems of those previous implementations in order to search for the reasons behind the failure.

A simple questionnaire with four questions was presented to the nine employees (eight male and one female) involved in previous CMM implementation, with ages around thirty-eight years old.

The answers given allowed concluding that the most appreciated feature of CMM was the use of common terminology (Fig. 2). However the most significant aspect mentioned was the reduction of errors and redundancies.

On the other hand, the worst aspects were the insensitivity about the organizational context and the amount of bureaucracy needed.

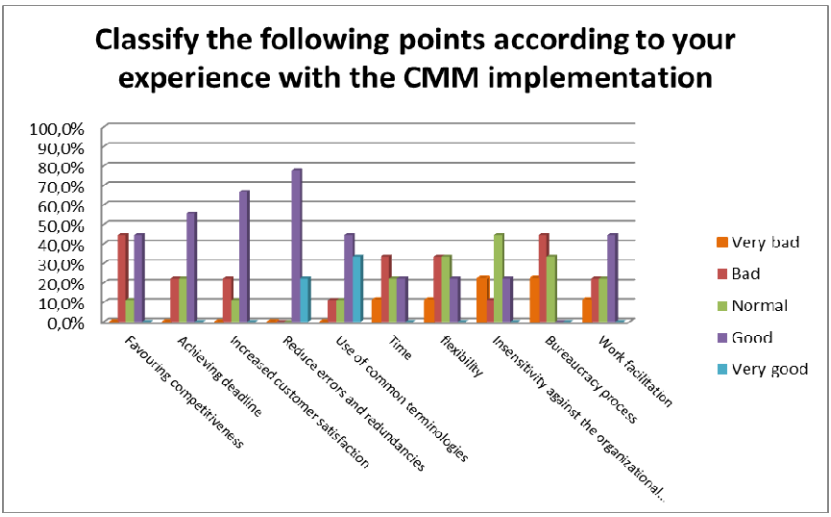


Fig. 2. Answers **to question nr 1** of questionnaire 1

The main conclusion about this first question (Fig. 2) is that employees feel CMM as very bureaucratic and hard to implement. However several positive aspects were also found. So, it is fair to conclude that people are open to the application of CMM provided it does not lead to an increased workload. Therefore solutions should be focusing on creating easy implementation processes.

Considering the second question, "What are the main positive aspects identified in project development before the implementation of CMM?", all answers given by this study participants (table 1) shown that before CMM, there was no existing standard, forcing people to find unrelated and creative ways for controlling their projects.

Table 1. Answers to question nr 2 of questionnaire 1

Respondents	Answers given
1	Greater documental freedom. More time to achieve the task. Better alignment with the project Time To Market (TTM), since over time changes do not require re-documenting all the planning in the department as with CMM.
2	Existence of well-defined systematic procedures.
3	When there are no standards we cannot talk about positive aspects of projects development.
4	Project management with less bureaucracy. Greater flexibility.
5	Individual creativity, lack of systematization promotes unexpected solutions, some of them innovative.
6	Less bureaucracy! Unnecessary standards use!
7	Did not answer.
8	Fast execution. Less time spent on bureaucratic processes.
9	Documents controlling activities and problems follow-up.

According to this question, we could observe that the majority of the answers show that before the CMM implementation, there was less bureaucracy.

The answers to the question of used standards, considering documentation, in projects development before CMM, can be analyzed in Table 2.

Table 2. Answers to question nr 3 of questionnaire 1

Respondents	Answers given
1	TTM time schedule used in project management – internal checklist – department standards checklist.
2	For our internal work processes, we could act without being owners, only in a few existing processes.
3	Non existing standards, each one creating one's supporting documents to project planning.
4	Open point list, cost planning, test matrix.
5	No standards documentation, each one organizing and documenting the activities according to the demands.
6	No memory.
7	No answer.
8	Before CMM we used: test matrix, test coverage and open point list.
9	A few, the ones we created.

According to this question there was no standard documentation for project management. We emphasize answer 9 (Table 2) "A few, the ones we created". This statement shows the existing reality, where each one developed his own procedure for project management.

Finally, Table 3 shows the given answers to the questionnaire's last question, "What changes/suggestions would you like to include regarding reimplementation improvement of CMM processes?"

The answers to this last question show the need of simplifying documents and processes. These suggestions will meet the results of the third question, in which bureaucracy was shown as to be a negative aspect of CMM. Linking these two questions, it seems that these are the key points for success in the reimplementation of CMMI.

3.2 Companies Questionnaire (Questionnaire 2)

A second questionnaire (online) was applied in order to analyze the implementation of CMMI in Portuguese companies certified by SEI. A questionnaire consisting of

five multiple choice questions was prepared and sent to ten companies. However, this activity has not completely successful because only four companies responded within the time needed to complete this project. But this questionnaire should be used for future projects in order to complete the intended analysis.

Table 3. Answers to question nr 4 of questionnaire 1

Respondents	Answers given
1	Simplicity of processes. Whenever possible, gather as much information in the same document serving several requirements. Use a web platform to simplify and have faster documentation.
2	Re-evaluation of all documents and procedures to help simplifying and decreasing documents. Less bureaucratic processes without confirmed utility.
3	Less papers and bureaucracy. Implementation should be supported by computer tools.
4	Fewer documents to fill in and less repeatability of information in different documents.
5	Automatic and standard processes.
6	Greater participation and commitment of all partners in reimplementation. Their participation in developing and reformulating standard tools and procedures to be used in the process is essential to their improvement.
7	Tools improvement to control test system! Using these tools should not be so bureaucratic! Some reviews should be done only once, not constantly (ex: When a Process Alignment and Verification (PAV) comes out, a new review should not be done till a new PAV appears). Creation of a specific TEAM to CMM.
8	Simple documentation process; controlling activities and trusted registration.
9	Evaluation of only strictly necessary documents; possibility to merge documents; create a CMM simplex.

Nevertheless, we have decided to analyze the four companies' answers, followed by presentation and results discussion.

The first question on the implementation of CMMI, 50% of the sample shows that it was considered difficult, 25% found it easy and 25% said that it was a normal process (Fig. 3).

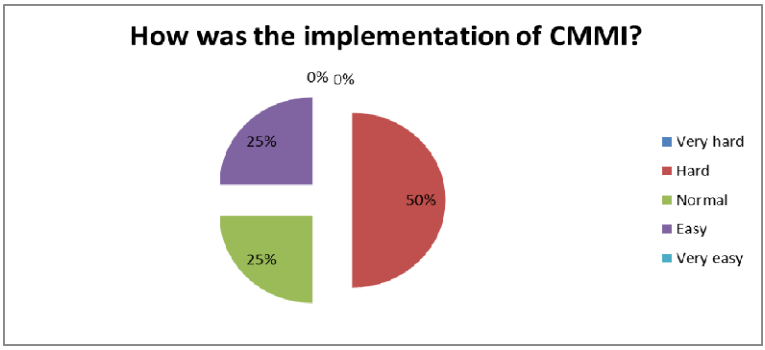


Fig. 3. Answers to question nr 1 of questionnaire 2

These results show that most companies participating in the questionnaire find difficult to implement CMMI. However, these different perceptions might be related to business context, capacity to reduce bureaucracy when applying CMMI or turning it more complex and laborious.

Considering the current state of CMMI in the company, 75% had already implemented this model, but the remaining 25% are still in an implementing phase (Fig. 4).

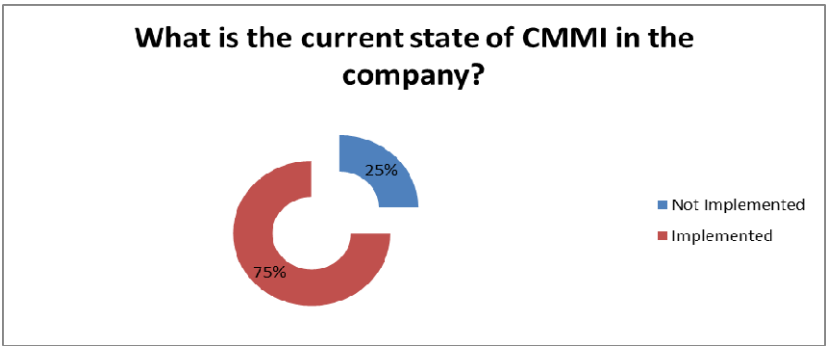


Fig. 4. Answers to question nr 2 of questionnaire 2

Regarding the achieved maturity level, 50% of the companies present level 2 and the remaining 50%, level 3 (Fig. 5).

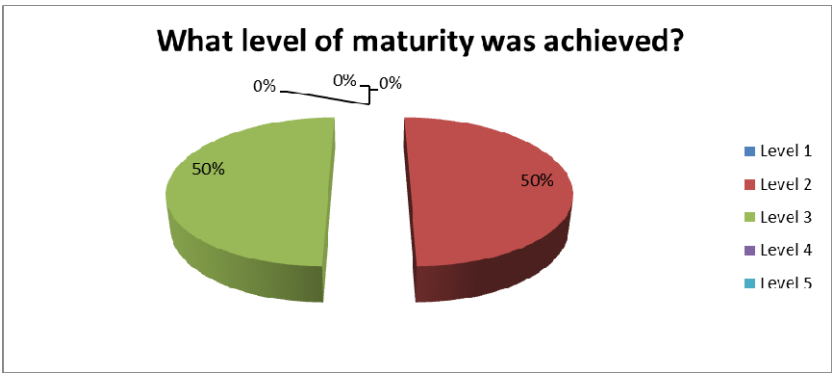


Fig. 5. Answers to question nr 3 of questionnaire 2

These results meet the available information in SEI, confirming that the most certified levels are level 2 and 3. So there is a consistency between information and results [9].

Considering the benefits of CMMI, results show that the highest advantage (75%) is related to processes control and the lowest one (25%) is centered on execution (Fig. 6).

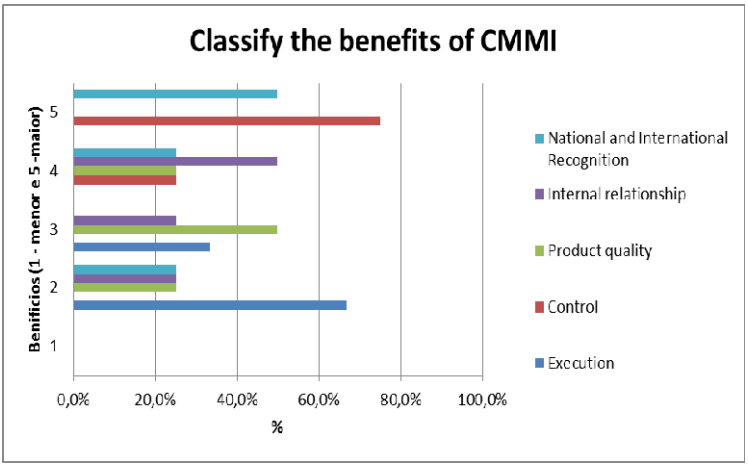


Fig. 6. Answers to question nr 4 of questionnaire 2

Once more, these results meet the answers given to the previous questionnaire and considering the Northern company employees: bureaucracy regarding execution appears as an aspect to improve in this model.

Classifying the disadvantages of CMMI, 75% choose the answer "Lack of involvement of employers and employees" (see Fig. 7).

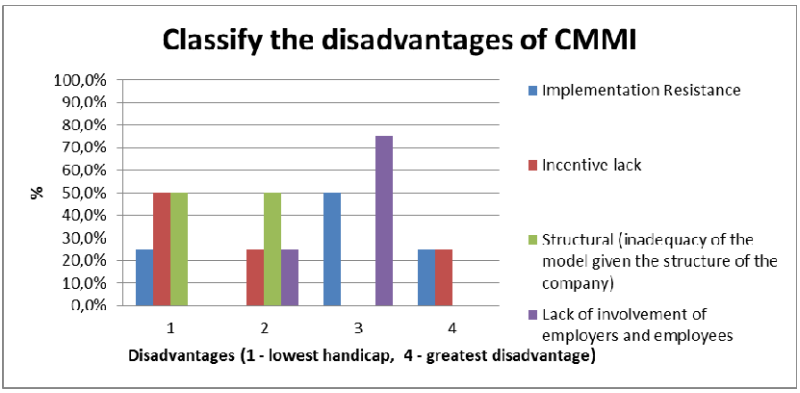


Fig. 7. Answers to question nr 5 of questionnaire 2

These results show a decreasing interest of employers affecting, consequently, the employees. This perception leads to the need on CMMI training to show the advantages of its use, turning the developed work easier.

3.3 Obstacles to CMM Implementation

Analyzing the previous processes and facing the answers to the applied questionnaires, we can see that all the processes involve a great complexity of tasks, leading to the need a higher number of resources to start a new project.

Here are some points that led to an ineffective implementation of the project:

- Process applied in just one department of the factory;
- Only dedicated to software;
- Very bureaucratic procedures (a lot of time spent filling in documents);
- Few new projects;
- Not applied to old projects (two different processes coexisting).

3.4 Reimplementation of CMMI Model

Taking into account the first and the second phases of the method, we started the third phase. Recognizing that the maturity model of the processes leads the company to its application, made us prepare a group of activities (Fig. 8). This plan of activities, accomplished with the responsible for the department under study, aimed to outline the goals to be achieved in this project, as we can see in Fig. 8 (planned tasks designed to achieve CMMI level 2).

Project:Re_Implement Staged Representation CMMI_Level2 "Managed"
CMMI_Level2 managed Kick off - CDQ0302
Define Test System Process Flow
Define Process Roles
Process Policy based CDQ0302 and CMMI level2
CMMI Project TEF7 Tracking
CMMI Senior Manager Tracking Meetings
CM - Configuration Management
PP - Project Planning
PMC - Project Monitoring and Control
SESAM - CM&PP&PMC
MA - Measurement and Analysis
RM- Requirements Management
SAM- Supplier Agreement Management
PPQA- Process and Product Quality Assurance
Roll out phase Pilot Projects "
SESAM - All KPA
Final Review at BrgP\TEF7 "Audit Gap" "guests CMMI experts"
Improvements "Close Gap"
SCAMPI B Decision&date
Final Assement based "SCAMPI B" Supported by SEI auditors members.

Fig. 8. Activities schedule

To implement this phase, we verified the main processes of the company and then, the tasks of each collaborator were defined.

To ensure this project's goals it was very important to include the administrators because it was a way for them to know of the improvements resulting of this implementation.

Finally, all the processes were changed and implemented. The results were evaluated periodically.

4 Conclusions and Future Research

In this study, we questioned about the processes used when implementing CMM. We decided to check this method in a technical department of a company in the North of Portugal. Thus, through questionnaires, we collected perceptions of the company employees, as well as companies certified by SEI in Portugal.

In an attempt to relate affinities between the questionnaires and the work performed by the company, we conducted a general discussion in order to understand the aspects to improve in the implementation of this process.

The results obtained in the questionnaires corroborate the information provided by SEI, already developed in the literature review, because all participants questioned recognized the model presented by SEI as a great benefit to the products developed.

According to the results for the company employees, we found that the lack of standard documentation and processes, constitute an obstacle to the promotion of CMM, forcing the creativity of each employee. Furthermore, accumulation of excess labour and paperwork, were some of the aspects considered disadvantageous. This handicap was also pointed out by the companies certified by SEI.

These companies perceive CMMI as a way to better control processes, but consider this approach difficult to implement. They add the disinterest of the employer as a negative factor to a successful implementation.

Based on these perceptions of CMM and CMMI, we tried to take advantage of the positive aspects highlighted and respond to the negative aspects, in order to facilitate the use of CMM, through the reimplementation of CMMI.

In this sense and taking into account that the business environment should be highlighted, we started our project with a survey of the situation in which the company was.

The analysis of the questionnaires led us to consider the bureaucracy as the main constraint to the success of this model. After reviewing our procedures and templates the reimplementation process continued by simplifying them. For example, we have created cost of equipment and materials tables, which improved the quotation process. Thus, we addressed the needs presented by employees and enterprises, decreasing the time spent in the processes, the accumulation of work, and also making the processes less bureaucratic.

By creating standard processes, we also contribute for greater processes control, which will lead to more successful work.

With these aspects that we think can improve the implementation of CMMI, we reduced the possibility of processes diversity, where each employee acts the way he considers best, preventing the monitoring and execution of the processes.

In an attempt to motivate and engage the employer in this model, we did regular presentations about the changes made, so that they were informed about all the steps done and approve them.

With these results we tried to respond to the research questions of this study. For the existing processes, we found that each employee had their own processes management documents. Despite projects key points were observed, there was no standard, making the interpretation of each project by employees who are not involved difficult.

With regard to flowcharts, this study confirmed that they are good work instruments. With them we can easily describe a process and through analysis we can find problems or improvements.

As for the CMMI, we found that when not implemented with the cooperation of all, this is usually abandoned due to its bureaucracy. Therefore, the great challenge for the reimplementation of CMMI was, and we believe it generally is, creating support tools to facilitate the work of the employees.

The workflow of the department is achieved through the experience of those involved in projects.

Best practices for managing projects include the creation of standards that allow people not to repeat unnecessary work. In addition, creating shared and systematic processes causes people to have the same knowledge of the project, leading to the reduction of possible errors.

As usually, some strengths and weaknesses are pointed out.

The study of the entire process for a department was one of the positive aspects that allowed understanding its operation in a complete and clear way. Moreover, the

finding of solutions to meet the satisfaction of the company and its employees, without reducing the customers' requirements, became an achieved challenge.

However, the search for solutions that pleased the whole organization was one of the limitations of this study. Long time was spent in the approval process for all project phases. Despite not having achieved all objectives yet, given that it is a continuous process of constant improvement, the results achieved so far point out to a successful future.

In the future, we would like to improve other key processes of CMMI model to achieve higher levels of maturity. Therefore, it becomes imperative to do more research on appropriate strategies necessary for the welfare of the employees of a company, when carrying out their work, viewing CMMI as a facilitator. Appropriate documentation, processes and methodologies will make workers more confident, which will be reflected in high performance.

References

1. Santos, R.P., Oliveira, K.M., Silva, W.P.: Evaluating the Service Quality of Software Providers Appraised in CMM/CMMI. Software Quality Journal 17(3), 283–301 (2009)
2. Wangenheim, C.G., Silva, D.A., Buglione, L., Scheidt, R., Prikladnicki, R.: Best Practice Fusion of CMMI-DEV v1.2 (PP, PMC, SAM) and PMBOK 2008. Information and Software Technology 52(7), 749–757 (2010)
3. SEI: Software Engineering Institute, `http://www.sei.cmu.edu`
4. Sun, Y., Liu, X.: Business-oriented Software Process Improvement Based on CMMI Using QFD. Information and Software Technology 52(1), 79–91 (2010)
5. Staples, M., Niazi, M., Jeffery, R., Abrahams, A., Byatt, P., Murphy, R.: An Exploratory Study of Why Organization Do Not Adopt CMMI. The Journal of System and Software 80, 883–895 (2007)
6. Huang, S.J., Han, W.M.: Selection Priority of Process Areas Based on CMMI Continuous Representation. Information & Management 43(3), 297–307 (2006)
7. Nandyal, R.: CMMI - A Framework for Building World-class Software and Systems Enterprises. Tata Mcgraw-Hill Publishing Co. Ltd. (2004)
8. Saunders, M., Lewis, P., Thornhill, A.: Research Methods for Business Students, 5th edn. Pearson Education (2009)
9. SEI: CMMI® for SCAMPISM Class A Appraisal Results 2011 End-Year Update. Pittsburgh, USA (2011)

Developing and Validating a Scale for Perceived Usefulness for the Mobile Wallet

Debby Ho, Milena Head, and Khaled Hassanein

DeGroote School of Business, McMaster University, Hamilton, Ontario, Canada

Abstract. Mobile wallet applications are new and promising mobile payment technologies. While researchers have started to explore adoption models in this domain, they have taken a generic approach to understanding customers' perceived usefulness of this technology. In this research-in-progress paper, we seek to develop a new, goal-oriented construct to understand the perceived usefulness of the mobile wallet from a consumer's perspective, which should provide a richer meaning to both academic and practitioner communities.

Keywords: Mobile Wallet, Perceived Usefulness, Scale Development.

1 Introduction

The mobile wallet (mWallet) is a relatively recent innovation in digital banking, which uses Near Field Communication (NFC) technology to access banking information and complete transactions [1]. As such, there is a need to understand what motivates different consumer segments to adopt this technology. In Information Systems (IS) research, the Technology Acceptance Model (TAM) has become one of the most influential and most widely cited theories in technology adoption [3; 6]. TAM proposes that system use is determined by two key constructs: perceived usefulness (PU) and perceived ease of use (PEOU) [9]. Given that the mWallet application simply requires a consumer to tap his/her mobile device on a designated point of sales unit to communicate with the checkout terminal, the ease of this use is deemed to be straightforward and is not likely to play a key role in consumer adoption. However, it is the perception of usefulness of mWallets and the underlying dimensions of this perception that is unknown. As such, the focus of this proposed research is on understanding the facets that contribute to PU of mWallets.

While the simplicity of the TAM model has been very popular in the research community, the practitioner community has not embraced this as "it does not explain why a system is seen as useful or easy to use" [6, p.1]. Benbasat and Barki [3] placed a call to researchers to advance IS adoption research to the next level. Specifically, they argue that few papers have examined these core constructs such that they have a richer meaning to both academic and practitioner communities. They indicate that these constructs "have largely been treated as black boxes that very few have tried to pry open" [3, p. 212]. In this work-in-progress research, we answer the call of [3], by proposing a methodology to "pry open" the black box of PU in the context of

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 469–476.
DOI: 10.1007/978-3-642-36981-0_43

mWallet adoption. As such, we seek to develop a new, goal-oriented construct to understand the PU of the mWallet.

This paper is organized as follows: In Section 2, we track the development of electronic banking leading to the advent of the mWallet; in Section 3, we propose a new higher-order construct for PU of mWallets; Section 4 outlines our proposed methodology; Section 5 outlines the potential contributions and limitations.

2 The Evolution of Electronic Banking: From Online Banking to Mobile Banking to Mobile Wallets

While online banking was developed during early 1980s, it was not adopted into the mainstream population until mid-1990s [11]. Online banking, otherwise known as Internet banking, is a service that allows clients to securely perform financial transactions via the Internet. Through the online channel, clients are empowered to access their bank accounts to conduct banking activities such as account transfers, balance inquiries, and bill payments, on a real-time basis.

Several studies have revealed consumer perceptions regarding the advantages and disadvantages of online banking. Consumers value factors such as lower fees, better service quality, 24-hour service availability, time savings, location-free access to services, ease of use, speed of service delivery, convenience, compatibility with lifestyle, and security [21]. However some consumers feel uncomfortable with online banking because of the lack of physical access to off-line channels, complexity, lack of trust due to chance of errors during service experience, and security risks [21; 4]. While perception of online security risks remains a concern in 2012, it is interesting to note that Americans, for example, feel more secure visiting a financial institutions' webpages websites than those from the general Internet [8].

Similar to online banking, mobile banking (mBanking) involves performing banking operations through a mobile device. By the end of 2011, there were over 5.9 billion mobile-cellular subscriptions worldwide, with mobile-broadband subscriptions having grown 45% annually since 2007 [17]. As technology progresses and mobile devices become more advanced, consumers are migrating from traditional electronic commerce to mobile commerce. For example, according to White [31], out of the 10 million smartphone users in Canada, 3.5 million Canadians used mBanking applications in 2012. Mobile banking applications ranked sixth in the most popular application categories in Canada, with one of the fastest adoption rates seen in the past decade [25]. However one in five Canadians who downloaded mBanking applications are reluctant to use them on a regular basis due to perceived security, usefulness, cost and trust concerns [25].

The innovation of the mobile wallet (mWallet), also known as a digital wallet, utilizes Near Field Communication (NFC) technology to provide consumers with a convenient alternative for transactional payments. A mWallet is a mobile application that stores a user's debit or credit card information securely in the cloud, which can be later accessed through a NFC enabled mobile device to make purchases at a point of sale. A consumer simply taps his/her mobile device on the designated point of sales

unit to communicate with the checkout terminal to make that purchase. mWallet applications can also be used to simplify online shopping. As a user's payment information is stored on the web under one streamlined account, he/she can sign into his/her account via a computer or web-enabled mobile phone to conveniently pay for the online purchase. Not only do mWallet applications have the potential to replace cash and plastic cards found in traditional wallets, but such applications also provide coupon offers and loyalty programs from participating merchants to help shoppers save and enhance their shopping experience.

Financial institutions, telecommunication companies and technology firms are actively partnering to revolutionize mobile payment methods. For example, in May 2012, the president of Rogers (a major Canadian telecommunication provider) wireless and wireline operations reported that roughly 300,000 customers have phones with NFC technology in Canada, and it expects to have three-quarters of a million devices in use by the end of 2012 [26]. In the United States, a recent report projected that mobile payment transactions will approximately quadruple to exceed $1.3 trillion by 2017 [22]. This report also noted that currently only four percent of global retail transactions are conducted through a mobile phone. Analysts expect that by 2017 with the widespread use of NFC, the mWallet payment method would account for 54 percent of the total value of mobile payment markets.

mWallet applications have been proposed to offer the following benefits: (1) convenience; (2) security; and (3) savings [15; 5]. These benefits will serve as the basis of our proposed higher-order PU construct. mWallet applications not only provide a new method for quick financial transactions through a simple tapping motion, but also offer the convenience of only carrying a mobile phone rather than the various plastic cards in a traditional wallet. When shopping online, users are able to quickly access their payment details under one mWallet account, avoiding the process of re-entering lengthy payment forms for each online transaction. Users can also better organize and manage their online and in-store transaction histories and no longer have to worry about lost receipts and rejected returns [15]. From a security perspective, mWallet applications are secured through entering a PIN to unlock the mWallet before the tapping transaction can be made. All debit and/or credit cards information are encrypted for storage. In the event that the mobile device is lost, users can simply remotely disable the mWallet online to ensure maximum security [5]. Finally, from a savings perspective, mWallet applications allow users to become aware of and redeem offers from retail and online merchants. Additionally, this technology has the potential to reduce the overall cost of transactions, some of which can be passed onto customers in terms of savings [15].

3 Understanding the Perceived Usefulness of Mobile Wallets

TAM-based research typically measures PU via surveys that are based on item response theory [14], where respondents indicate the degree to which they agree or disagree with statements on a Likert scale. The original PU items (as per [9]) are: (1) Using the system in my job would enable me to accomplish tasks more quickly;

(2) Using the system would improve my job performance; (3) Using the system in my job would increase my productivity; (4) Using the system would enhance my effectiveness on the job; (5) Using the system would make it easier to do my job; and (6) I would find the system useful in my job. Clearly some of these items do not apply to the mWallet context. Chesney [6] stressed that these items are independent of user goals and that the inclusion of user goals is crucial to provide usefulness with meaning to both academic and practitioner communities. In the domain of mWallets, some adoption models have been proposed and validated (for example: [29; 20; 1]), but they have all taken a generalized view of PU. As such we seek to develop a meaningful PU construct that is tied to relevant user goals in the context of mWallets.

We propose a second-order model to measure PU for the mWallet as shown in Figure 1. Latent constructs can be measured using either reflective or formative indicators. According to [12], reflective indicators are used to account for observed variances or covariances, while formative indicators do not account for observed variables. As per Jarvis, et al [18], constructs should be modeled as being formative if (1) the direction of causality is from indicators to constructs (i.e. indicators are defining characteristics of the construct); (2) indicators are not interchangeable (i.e. they do not have the same content); (3) the indicators do not covary with each other; and (4) nomological net for the indicators may differ (i.e. they are not required to have the same antecedents and consequences). For the second-order construct we are proposing, these decision rules suggest that the PU construct and its sub-constructs should be modeled as formative. In our formative second-order model each first-order construct represents a unique aspect that contributes a new facet to the higher order PU construct (as per the higher-order construct specification in [7; 13]). The three first-order constructs of convenience, security and savings (based on an initial examination of relevant literature) are not expected to be highly interrelated. Nonetheless, while the proposed three first-order constructs are unrelated to each other, all combine to serve the same overall purpose of perceived usefulness of mWallets. The proposed model is both parsimonious and practically relevant as practitioners can understand and focus on the first-order factors to influence PU of mWallets.

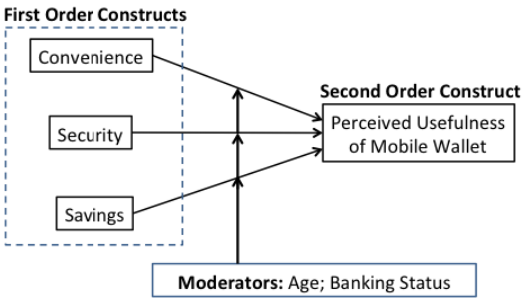


Fig. 1. Proposed Higher-Order Nature of the Perceived Usefulness of Mobile Wallets

3.1 Moderating Effects

Moderators are variables that are proposed to magnify, attenuate or reverse the association between two other variables [2]. In technology adoption research, moderators (such as individual characteristics) are recognized for their potential value, but are currently under-represented [28]. It has been argued that the low explanatory power of TAM can be attributed to the exclusion of important moderating variables that reflect individual differences [30]. From a practitioner perspective, understanding how relationships in a technology adoption model vary by customer differences, can help to guide marketing and promotional strategy development to encourage use of new technologies. In the context of mWallets, we identify two individual differences (age and banking status) that have been suggested as having a potential moderating influence on the relationships between the first order and second order constructs of proposed research model.

***Age*:** Studies have found that within consumer demographics (i.e., age, income, employment and marital status), mobile banking adoption is mostly dependent on age. Younger adults (under the age of 35) have been found to be more likely to adopt mBanking, as they are a technologically savvy group and belong to social norms that focus on and value mobile usage [16; 27]. Schmeiser and Cope [27] suggested that younger populations may have lower emphasis on security concerns (in part because they tend to have fewer assets in their financial accounts), which can explain the higher adoption of mBanking within this demographic group. In 2011, the Royal Bank of Canada (RBC) found that the majority of the bank's mBanking users were in the 18 to 40 age bracket, with two peaks at 22 – 24 and 35 – 40 [31]. RBC suggested that the mBanking usage among 18 – 24 year olds aligned with the technological norms within that cohort; while the 35 – 40 clients used mBanking due to its convenience, which allowed them to be more productive and efficient. A 2011 comScore study conducted in the United States found that 59 percent of mobile bankers are in the 18 to 34 age bracket. Specific to mWallet applications, a 2011 Canadian poll conducted by Solution Research Group found that 66 percent of those aged 18 – 34 expressed interests in adopting mWallet capabilities [31]. Given the above, it is expected that age will moderate the relationships between the first order and second order constructs in our proposed model, where security and convenience will have a stronger influence on PU for adults over the age of 35, whereas savings will have a stronger influence on PU for adults under the age of 35.

***Banking Status*:** Studies conducted by the Federal Reserve System [10] and Javelin Strategy and Research [19] identified two new segments – the "Underbanked" and the "Unbanked" groups – for financial institutions to target for their mobile banking services. The "Underbanked" are defined as consumers who do not have a chequing account, but often have prepaid cards, savings accounts or other financial accounts [10]. The "Unbanked" are categorized as consumers who do not own any banking products [10]. Javelin [19] estimated that 15 percent of the American population is underbanked, whereas 4.3 percent is unbanked [10]. These underbanked and unbanked consumers are disproportionately younger, with three out of five being under the age of 35 [19]. In addition, these consumer segments in the U.S. are less likely to own computers (60 percent versus 72 percent for all consumers) or pay for

broadband services (34 percent versus 59 percent), but are more likely to own a mobile phone (74 percent versus 68 percent) [19]. These defined groups also earn a lower annual income of $52K versus the American average of $73K [19]. However Javelin [19] suggested that collectively the income of this group poses a tremendous opportunity that financial institutions have yet to capture. Given the above, we expect that for the unbanked and underbanked segments, savings will have a stronger influence on perceived usefulness of mWallets, whereas for the banked segment, more emphasis may be placed on convenience and security.

4 Methodology

In order to develop and validate the proposed second-order construct for the PU of the mWallet, we will follow the methodology outlined by Rai, et al [24] in their development of new higher-order constructs. While the model shown in Figure 1 indicates the three first-order constructs of convenience, security and savings are used for conceptualizing PU for mWallets, these constructs need to be carefully considered and assessed. A thorough examination of the extant academic and practitioner literature will be conducted to: (1) determine if the proposed first-order constructs in our model should be modified; (2) determine if additional constructs should be added that are directly relevant to user goals in the context of mWallets or that could play a moderating effect in this regard; and (3) propose appropriate indicators or items for each of these formative first-order constructs. Following this thorough literature review, a panel of four academic experts (as per Moore, et al. [23]) that are familiar with construct formation in information systems research will be asked to independently evaluate the constructs and their indicators. Following their independent assessments, these researchers will meet to discuss and resolve any discrepancies until they come to an agreement on the constructs' content validity. A second set of information systems scholars (2-3) with expertise in survey research and expertise in the subject domain will subsequently evaluate the instrument.

Following these two assessments and incorporating any suggested changes, a two-phase pilot study will be conducted. The first phase of the pilot test will be performed with a broader sample of information systems scholars (faculty and PhD students) who are actively researching the adoption of new information technologies. It is expected that 15-20 scholars would participate in this pilot test, who would be asked to provide their comments on the construct indicators and respond to semi-structured questions on each measure's content validity. Based on their feedback, the instrument will be modified as appropriate. The second phase of the pilot test will use a similar approach with 15-20 banking clients that have some familiarity with mWallets.

Following the two-stage pilot study, the new instrument will be administered within a full-scale study to validate the proposed model and examine the role of the demographic moderators shown in Figure 1. The sample will be recruited through a market research firm to ensure having a sufficient and balanced response rate from the desired demographic sectors (age and banking status). A sample of 200 subjects will be adequate to conduct the proposed analyses. The first and second-order constructs will be assessed using partial least squares (PLS). This approach is recommended for predictive research models where the emphasis is on theory development [24]. PLS also has the ability to model formative constructs and moderating effects.

5 Expected Contributions and Limitations

This research promises to make significant contribution to both theory and practice. From a theoretical perspective, and to the best of our knowledge, this will be the first study to develop a customized PU scale for the mWallet. This second order scale could be used to create more meaningful adoption models for this technology in the future. The methodology for developing this scale could also be adopted for developing customized PU scales for other technologies. On the other hand, practitioners can use the results of this research to customize understand the most salient factors contributing to the benefits perceived by consumers for the mWallet and how the importance of such factors may vary by consumer age and current banking status. This understanding could be leveraged in customizing the mWallet marketing messages to different consumer groups. As the sample for this study is selected from Canada and the US, the generalizability of our findings is limited to those societies. Future research should try to validate our findings for other cultures in the developed, developing and underdeveloped world.

References

1. Amoroso, D., Magnier-Watanabe, R.: Building a Research Model for Mobile Wallet Consumer Adoption: The Case of Mobile Sucia in Japan. Journal of Theoretical and Applied Electronic Commerce Research 7(1), 94–110 (2012)
2. Baron, R.M., Kenny, D.A.: The moderator mediator variable distinction in social psychological research, conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 51(6), 1173–1182 (1986)
3. Benbasat, I., Barki, H.: Quo vadis tam? Journal of the Association for Information Systems 8(4), 211–218 (2007)
4. Benamati, et al.: Are trust and distrust distint constructs? An empirical study of effects of trust and distrust among online banking users. In: Hawaii International Conference on System Sciences, vol. 39 (2006)
5. Ceipidor, U., Medaglia, C., Opromolla, A., Volpi, V., Moroni, A., Sposato, S.: A Survey about User Experience Improvement in Mobile Proximity Payment. In: 4th International Workshop on Near Field Communication (2012)
6. Chesney, T.: Opening the Black Box of PU: An IS Approach to Defining and Measuring Userfulness. Nottingham University Business School Research Paper No. 2010-05 (2010)
7. Chin, W.: Issues and Opinions on Structural Equation Modeling. MIS Quarterly 22(1), 7–16 (1998)
8. comScore. 2011 State of Online and Mobile Banking (Feburary 2012), http://www.comscore.com/Press_Events/Presentations_Whitepapers/2012/2011_State_of_Online_and_Mobile_Banking (retrieved)
9. Davis, F.: Perceived Usefulness, Perceived Ease of Use and User Acceptance of Information Technology. MIS Quarterly 13(3), 319–340 (1989)
10. Financial Brand. Targeting Underbanked Consumers with Mobile Banking Soluations (June 19, 2012), http://thefinancialbrand.com/24356/targeting-underbanked-consumers-with-mobile-banking-solutions/ (retrieved)
11. Fitzmaurice, M.: When did onlilne banking begin? (2012), http://www.ehow.com/facts_5035719_did-online-banking-begin.html (retrieved)
12. Fornell, C., Bookstein, F.: Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory. Journal of Marketing Research 19, 440–452 (1982)

13. Gefen, D.: E-Commerce: The Role of Familiarity and Trust. Omega 28(5), 725–737 (2000)
14. Hambleton, R.: Fundamentals of item Response Theory. Sage (1991)
15. Hoofnagle, C., Urban, J., Li, S.: Mobile Payments: Consumer Benefits & New Privacy Concerns, BCLT Research Paper (2012), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2045580
16. Huynh, S.: The Data Digest: Mobile Banking's Massive Growth is Supplementary to PC Banking (July 6, 2012), http://blogs.forrester.com/category/mobile_banking (retrieved)
17. ITU Telecom World. The World in 2011: ICT Facts and Figures (2011), http://www.itu.int/ITU-D/ict/facts/2011/material/ICTFactsFigures2011.pdf (retrieved)
18. Jarvis, C., Mackenzie, S., Padsakoff, P.: A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. Journal of Consumer Research (30), 199–217 (September 2003)
19. Javelin Strategy & Research. 5th Annual Mobile Security Report: Mitigating Security Risks Transforms the Mobile Banking Channel (December 2011), https://www.javelinstrategy.com/brochure/234 (retrieved)
20. Kim, C., Mirusmonov, M., Lee, I.: Computers in Human Behavior, vol. 26, pp. 310–322 (2010)
21. Laukkanen, T., Lauronen, J.: Consumer value creation in mobile banking services. International Journal of Mobile Communications 3(4), 325–338 (2005)
22. Maisto, M.: Google Wallet, Isis, NFC Will Push Mobile Payments to $1.3 Trillion, eWeek (2012), http://www.eweek.com/c/a/Mobile-and-Wireless/Google-Wallet-Isis-NFC-Will-Push-Mobile-Payments-to-13-Trillion-708700/
23. Moore, G.C., Benbasat, I.: Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. Information Systems Research 2(3), 192–222 (1991)
24. Rai, A., Patnayakuni, R., Seth, N.: Firm Performance Impacts of Digitally Enabled Supply Chain Integration Capabilities. MIS Quarterly 30(2), 225–246 (2006)
25. Robertson, G.: Mobile banking apps take off. Globe and Mail (July 03, 2011), http://www.theglobeandmail.com/technology/mobile-banking-apps-take-off/article588029/ (retrieved)
26. Sturgeon, J.: Rogers and CIBC first to launch 'mobile wallets' in Canada. Financial Post (May 15, 2012), http://business.financialpost.com/2012/05/15/rogers-and-cibc-first-to-launch-mobile-wallets-in-canada/ (retrieved)
27. Schmeiser, M.D., Cope, A.M.: Risk Perception. Risk Tolerance and Consumer Adoption of Mobile Banking Services (April 30, 2012), http://ssrn.com/abstract=2048565 (retrieved)
28. Shin, D.: User acceptance of mobile Internet: Implication for convergence technologies. Interacting with Computers 19(4), 45–59 (2007)
29. Shin, D.: Towards an Understanding of the Consumer Acceptance of Mobile Wallet. Computers in Human Behavior 25, 1343–1354 (2009)
30. Sun, H., Zhang, P.: The role of moderating factors in user technology acceptance. International Journal of Human Computer Studies 64(2), 53–78 (2006)
31. White, S.: Canadians embracing mobile banking. Globe and Mail (April 30, 2012), http://www.theglobeandmail.com/globe-investor/personal-finance/financial-road-map/canadians-embracing-mobile-banking/article4104236/ (retrieved)

Term Proximity and Data Mining Techniques for Information Retrieval Systems

Ilyes Khennak and Habiba Drias

Computer Science Departement, USTHB,
Laboratory for Research in Artificial Intelligence (LRIA),
Algiers, Algeria
heoui@live.com, h_drias@hotmail.fr

Abstract. Term clustering based on proximity measure is a strategy leading to efficiently yield documents relevance. Unlike the recent studies that investigated term proximity for improving matching function between the document and the query, in this work the whole process of information retrieval is thoroughly revised on both indexing and interrogation steps. Consequently, an Extended Inverted file is built by exploiting the term proximity concept and using data mining techniques. Then three interrogation approaches are proposed, the first one uses query expansion, the second one is based on the Extended Inverted file and the last one hybridizes retrieval methods. Experiments carried out on OHSUMED demonstrate the effectiveness and efficiency of our approaches compared to the traditional one.

Keywords: information retrieval, term proximity, word association, Fuzzy Clustering.

1 Introduction

The information technologies progress such as text editors has impacted on daily production of an enormous mass of information. On the other hand, the evolution of electronic media has allowed the storage of this vast amount of information. With the development of electronic communication systems and the increase in the amount of information available online, it becomes important to help users to quickly access information they need.

Traditional Information Retrieval (IR) offers tools for users to find the information they need from sets of data with reasonable size. It consists in selecting from a volume of information, the relevant information to a user query. The traditional IR systems interpret documents as sets of words without sense. Therefore, they are capable to only find documents described by the words considered separately from the query. In other terms, they treat IR only on the basis of morphological aspect of the document text. Intuitively when further considerations such as semantic features are taken into account, the information retrieval system is expected to work more efficiently and faster. Of course these concerns are more complicated to address and constitutes important directions of research.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 477–486.
DOI: 10.1007/978-3-642-36981-0_44

1.1 Related Works

In order to palliate the limits of traditional information retrieval, new approaches taking into account words senses, proximity and association have been proposed in the literature [3], [5], [9], [11] and [12]. They use semantic resources and statistical techniques to improve the performance of information retrieval systems.

The proximity of query terms is intuitive in general for the users. These last years a few works interested in improving document ranking using this heuristic have been published. The majority of these papers focused on designing measures for the term proximity to integrate in the matching process between the document and the query in order to yield better document relevance.

In papers [3], [9] and [11], the authors propose proximity terms measures they combine to the classical term weighting function in order to better translate the importance of the terms in a document. All these studies claim the benefit of integrating term proximity scoring in the term weighting from the success of their experiments performed on known benchmarks.

In paper [5] and [12], the authors present a theoretical study on modeling term proximity and show that the use of term proximity enhances considerably the system efficiency.

1.2 Assigned Objectives

Usually it is meant by term proximity the minimum number of words that separates two terms that appear in the same document. Of course, when this number is small, the proximity is more important. When it is equal to one, that is when the two words are adjacent such as "information retrieval", we talk about word association. In this study we examine the concept of proximity since it is more general than the concept of association. Moreover, one term may have high proximity with several other words and in many documents. For instance the word "information" may be associated with the words "retrieval", "science" and "technology".

In the approach we propose, and relying on [4], all the information retrieval process is revised on both indexing and interrogation phases, which makes its originality relatively to the works reported in the literature. The indexing process is designed according to the new added feature which is term proximity. It is developed in such a way to help the interrogation phase to answer quickly to the user queries. At our knowledge, there exists no study that has proposed so far such approach.

2 Traditional Information Retrieval

Traditional Information Retrieval has been widely investigated because of its important and numerous applications. It is well related in many books such as [1], [2], [8] and [6]. In this section, the major concepts used in the literature are presented.

The set of documents on which the search is performed according to the user request form a *Collection*. The *Document* can be a text, a sound, an image or a video. We call therefore a document any unit that can constitute a response to a need of

information. The latter is introduced to the machine through a *query*. The *Information Retrieval System* aims at finding in the mass of available information, the relevant documents to satisfy the query. A preliminary phase of document analysis is needed and corresponds in practice to the *Indexing process*. The formulation of the query, the search in the collection and the ranking of documents using a matching function define the *Interrogation* phase.

In the traditional IRSs, the document is considered as a set of words represented by *descriptors*. The only information exploited on these words is the frequency of their occurrences in the document. The content of the query is described by a set of words as for documents.

2.1 Traditional Indexing

The indexing process starts by recognizing words from the text and eliminating all superfluous and unnecessary entities such as the punctuation marks and spaces. Once the words are extracted, they are normalized using the Stemming operation [2]. The latter is a morphological process that allows finding more general form of words.

Finally, the terms of the dictionary are coded by identification numbers. These identifiers are stored for use during the search in a file. Then to each term the process assigns a weight that represents its importance in the document where it appears. To weight the terms, the measure *tf*idf* is usually used where *tf* measures the importance of the term in the document and is computed using formula (1), which is the Okapi metric [10]. The variable $occur_{ij}$ is the number of occurrences of term *i* in document *j*. *k* is introduced to take into account the length of the document. In practice it is calculated using formula (2) where $length\ doc_j$ is the length of document *j* and *average length doc* is the average of document lengths. Therefore, formula (3) is used to compute *tf.idf* measures the importance of the term in the collection and is usually computed using formula (4), where *n* is the number of documents and n_i the number of documents containing term *i*.

$$tf = \frac{occur_{ij}}{occur_{ij} + k}. \tag{1}$$

$$k = 0.5 + 1.5 * \frac{length\ \ doc_j}{average\ \ length\ \ doc}. \tag{2}$$

$$tf = \frac{occur_{ij}}{occur_{ij} + 0.5 + 1.5 * \dfrac{length\ \ doc_j}{average\ \ length\ \ doc}}. \tag{3}$$

$$idf = \log(\frac{n}{n_i}). \tag{4}$$

Once the terms are weighted, for each term the process associate a list containing the documents in which it appears followed by the corresponding weight. The achieved index is saved in a file called *Inverted File*.

2.2 Traditional Interrogation

As it was mentioned earlier, it consists in finding documents that are relevant to the query. In the literature, there is a variety of models to represent documents and queries. Among them the *vector model* is the most frequently used. In such context, each document d is described by a vector of terms weights as follows:

$$d=(w_1, w_2, \ldots, w_m).$$

Where w_i is the weight of term i in document d. As with the documents, the query is represented by a vector of weights. If q represents a query, it is then modeled as follows:

$$q=(v_1, v_2, \ldots, v_m).$$

Where v_i is the weight of term i in q for $i=1, \ldots, m$. m being the number of terms in the dictionary. The matching function determines the relevance of a document to a query, and allows ranking the documents in order of assumed relevance. To appreciate the degree of relevance of a document d to a query q, the *scalar product* of formula (5) calculates the retrieval status value $RSV(q, d)$.

$$RSV(q,d) = \sum_{i=1}^{m} v_i * w_i. \tag{5}$$

$RSV(q,d)$ represents the degree of similarity between the document d and the query q.

In order to evaluate the query, all documents containing at least one term of the query are selected and their relevance to the query is calculated. The documents are then ordered according to their degree of relevance.

3 Indexing with Term Proximity

In the indexing phase, the terms are grouped into clusters according to their proximity to each other in the documents. Term clustering is implemented to construct an Extended Inverted file. This file contains instead of basic relations like *[Term, Document]*, more informed relations with the format $[(Term_1, Term_2), Document]$, where $Term_1$ and $Term_2$ belong to the same cluster.

3.1 Term Clustering

Clustering consists in dividing a population of objects into subsets called *Clusters* so that all objects belonging to the same cluster are similar and objects of different

clusters are dissimilar [7]. An *Object* is an elementary datum presented in the input to the clustering algorithm. In general, this unit is represented by a vector of *p* attributes as follows:

$$X=(x_1, x_2, \ldots, x_p).$$

An *Attribute* of an object is a scalar x_i contained in the vector of data. A *Centroid* of a Cluster *C* is a vector *V*, where *V[i]* is the arithmetic means of the attributes *i* of the objects belonging to cluster *C*. *V[i]* is computed using formula (6) where v_{ij} is the value of attribute *i* of the object *j* and |*C*| is the number of objects of cluster *C*.

$$V[i]=\frac{1}{|C|}\sum_{j=1}^{|c|} v_{ij}\,. \tag{6}$$

The *Distance* is a measure that expresses the similarity between two objects. Its value is high when the two objects come together in the same cluster and small when they reside in two different clusters. There are several clustering techniques; the k-means algorithm is the most adequate for this application. The fuzzy k-means is more adapted for term clustering and more effectiveness is expectable. For such algorithm, an object can belong to more than one cluster. The distance measurement corresponds usually to a formula of similarity in the simple k-means algorithm and a membership function in the fuzzy k-means algorithm. It has been shown that the complexity of the k-means algorithm is exponential. For this concern, we propose another algorithm with a simpler complexity.

For the purpose of grouping terms, we proceed on modeling the space of terms according to the concepts of clustering: a dictionary term as the *object*, the weight of the term in a document as the *attribute* and the set of terms often co-occurring in the same documents as a *cluster*. The *cosine similarity measure* is used to calculate the distance between two terms. Formula (7) computes this score where x_i and y_i are respectively the weights of term *x* and term *y* in document *i*.

$$Distance(x,y)=\frac{\sum_{i=1}^{n} x_i * y_i}{\sqrt{\sum_{i=1}^{n}(x_i)^2+\sum_{i=1}^{n}(y_i)^2}}. \tag{7}$$

To cope with the complexity issue of the fuzzy k-means algorithm, we made some changes to the clustering algorithm. We start by determining the number of clusters compared to the number of terms in the dictionary. Then, we assign each term to a class; each term in the dictionary corresponds to a centroid. Once the clusters are defined, we assign to each of them the co-occurring terms with the centroid of the cluster. This assignment is made using a calculation of the *conditional probability of proximity* between the centroid of the cluster and each term in the dictionary. We replaced the distance scheme by the conditional probability of proximity *Proximity*($t_1|t_2$) [12], which is given by formula (8). *Count*(t_1 *V* t_2) counts the number of documents where t_1 and t_2 appear together and *Count*(t_2) counts the number of

documents where t_2 appears. All the counts are undertaken before launching the clustering algorithm.

$$\Pr oximity\ (t_1 | t_2) = \frac{P(t_1 \vee t_2)}{P(t_2)} = \frac{Count\ (t_1 \vee t_2)}{Count\ (t_2)}. \tag{8}$$

The result is a set of clusters, where each one is described by a term centroid in the dictionary and a proximity vector of terms. The modified clustering algorithm is summarized in the following pseudo code:

Algorithm 1. Term Clustering

```
1: Initialize the number k of clusters relative to the number of terms in the dictionary
2: Initialize the clusters: Assign each term in the dictionary to a cluster
3: for each term t_i in the dictionary do
4:     for each term t_j in the dictionary (i<>j) do
5:         Calculate Proximity(t_i|t_j)
6:         Add the value of Proximity to V[i]
7:     end for
8:     Sort V[i]
9: end for
```

V[i] is the vector of proximity measures of the centroid. The result is then stored in a file called *Clusters*. In order to access to the cluster that contains a term, each term of the dictionary is linked to the centroid of the cluster to which it belongs.

3.2 Extended Inverted File

In this step, we recovered for each term different from centroid of a cluster the documents where it appears with the centroid of this cluster. We then save the obtained index in the *ExtendedInvertedFile*.

4 Interrogation Approaches with Term Proximity Scheme

As in traditional IRS, the documents and the query are represented by weight vectors of terms. Thus, the same matching function is used to calculate the similarity between a document and a query. In this section, we present three different methods for extracting relevant documents.

4.1 Query Expansion Based Approach

In this retrieval method, we use the structure *Clusters* to recover the terms co-occurring with certain proximity with those of the query. We then add these terms to the query in order to select the documents that contain at least one of them. One possible version of the retrieval algorithm is as follows:

Algorithm 2. Query Expansion Based Approach

```
1: Select a query Q to satisfy
2: Q'new ← Q
3: for each term t in Q do
4:     Select (from Clusters file) the best co-occurring term (e.g. t') with t
5:     Q'new ← Q'new + t'
6: end for
7: Satisfy the Q' new using the classic search
```

4.2 Extended Inverted File Based Approach

In this method, we select all documents that include at least one pair of query terms.

Algorithm 3. Extended Inverted file Based Approach

```
1: Select a query Q to satisfy
2: for each term pair {ti, tj} in Q do
3:     Select (from Extended Inverted file) all documents containing ti and tj
4: end for
5: for each selected document di do
6:     Calculate the degree of relevance RSV(Q, di) using the Scalar product
7: end for
8: Sort the documents selected according to their relevance
```

4.3 Hybrid Approach

For this search method, we exploit the *InvertedFile* and the *ExtendedInvertedFile* structures together.

Algorithm 4. Hybrid Approach

```
1: Select a query Q to satisfy
2: for each term pair {ti, tj} in Q do
3:     Select (from Extended Inverted file) all documents containing ti and tj
4: end for
5: for each term ti in Q do
6:     Exist ← false
7:     for each term tj in Q (ti<>tj) do
8:         if {ti, tj} exists in Extended Inverted file do
9:             Exist ← true
10:        end if
11:        if not Exist = false then
12:            Select (from Inverted file) all documents that contain ti
13:        end if
14:    end for
15: end for
16: for each selected document di do
17:    Calculate the degree of relevance RSV(Q, di) using the Scalar product
18: end for
19: Sort the documents selected according to their relevance
```

5 Experimental Results

5.1 The OHSUMED Collection

Extensive experiments were performed on OHSUMED test collection (part of RCV1 collection). It is a set of 348 566 references from MEDLINE, the on-line medical information database, consisting of titles and/or abstracts from 270 medical journals over a five-year period (1987-1991). The available fields are title, abstract, MeSH indexing terms, author, source, and publication type. It requires about 381 Mb for storage of the uncompressed files. The evaluations were performed using only the titles of documents.

The designed algorithms are implemented with Python under Intel(R) Core(TM)2 Duo CPU T6570 @ 2.10 GHz 2.10 GHz, 3 Gb RAM hardware configuration.

5.2 Indexing Step

To evaluate our proposed approaches, and especially the search methods, we partitioned the collection of documents into sub-collections. Table 1 presents the characteristics of each sub-collection after the preprocessing step.

Table 1. Characteristics of the sub-collections

Size of the collection:	**(# documents)** **(Mb)**	**50 000** 3.34	**100 000** 6.66	**150 000** 10.1	**200 000** 13.5
Number of terms in the dictionary		35443	50232	62126	71699

Size of the collection:	**(# documents)** **(Mb)**	**250 000** 17	**300 000** 20.6	**350 000** 24
Number of terms in the dictionary		80717	88764	96331

5.3 The Proposed IRS versus the Traditional One

In this section, we compare the results of our retrieval methods with those of the traditional IRS. The experiments have been performed on the basis of 106 queries.

Fig. 1 shows the behavior curves corresponding to the classic (traditional) and the three proposed algorithms in terms of execution time. We observe that the runtime corresponding to the Extended Inverted file is not only the shortest for all the sub-collections but is almost constant and close to 0, which allows real time processing.

From the performance, Fig. 2 illustrates the superiority of the Query expansion strategy on the other methods.

Finally, when comparing the algorithms in terms of the number of extracted documents (Fig. 3), we see clearly the gap separating the result produced by the Extended Inverted file based algorithm and those of the other approaches in favor of the former and it is more important for the classic retrieval.

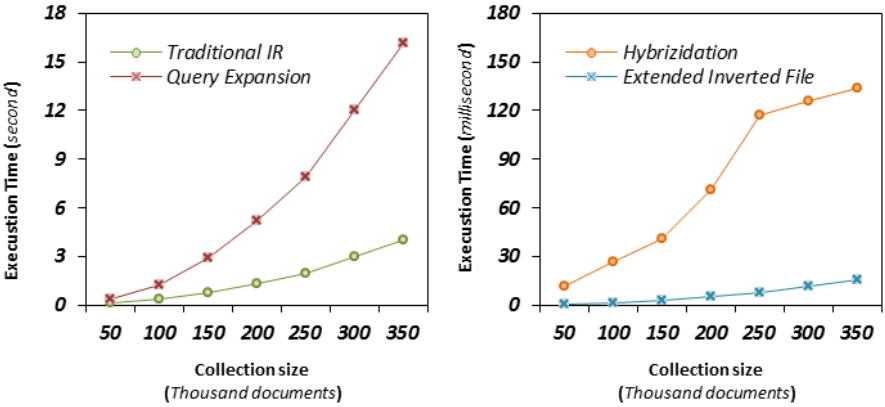


Fig. 1. Comparison of the four retrieval approaches in terms of execution time

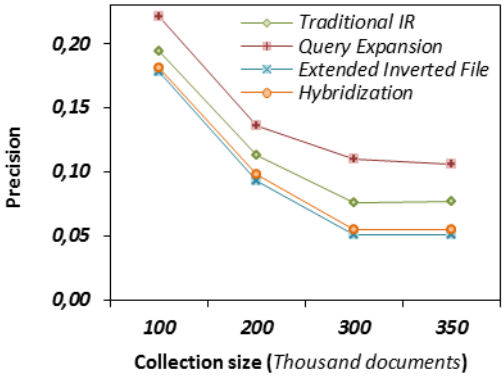


Fig. 2. Comparison of the methods in terms of document relevance quality

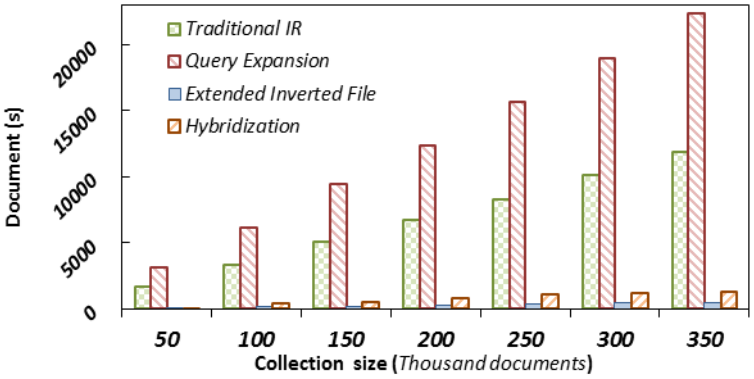


Fig. 3. Comparison of the four retrieval approaches in terms of the number of extracted documents

6 Conclusion

This work allowed us to study and develop an information retrieval system founded on traditional information retrieval background augmented with the developments based on the interpretation of term proximity concept.

As part of this work, we proposed an indexing method based on the grouping of terms using statistical methods and clustering techniques. The realization of this proposal can be summarized by the creation of *Cluster* file and *Extended Inverted* file. These files are used in the interrogation phase by several retrieval methods to quickly find the relevant documents. The retrieval methods, using effective indexes, are very efficient and give very satisfactory results in terms of robustness and computation time.

The system we developed has been tested on OHSUMED and the achieved results are extremely interesting. According to numerical values, the efficiency of our IRS becomes clearly visible when using the Extended Inverted file in the interrogation phase. It indeed shows the importance of exploiting the term proximity concept in information retrieval development.

References

1. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. Addison-Wesley, New York (1999)
2. Chu, H.: Information representation and retrieval in the digital age. Information Today, New Jersey (2010)
3. Cummins, R., O'Riordan, C.: Learning in a pairwise term-term proximity framework for information retrieval. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 251–258 (2009)
4. Drias, H., Khennak, I., Boukhedra, A.: A hybrid genetic algorithm for large scale information retrieval. In: IEEE International Conference on Intelligent Computing and Intelligent Systems, ICIS, pp. 842–846 (2009)
5. He, B., Huang, J.X., Zhou, X.: Modeling Term Proximity for Probabilistic Information Retrieval Models. Information Sciences Journal 181(14) (2011)
6. Kowalski, G.: Information Retrieval Architecture and Algorithms. Springer, New York (2011)
7. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco (2011)
8. Manning, D.M., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008)
9. Mingjie, Z., Shuming, S., Mingjing, L., Ji-Rong, W.: Effective Top-K Computation in Retrieving Structured Documents with Term-Proximity Support. In: CIKM 2007 (2007)
10. Robertson, S.E., Walker, S., Hancock-Beaulieu, M., Gatford, M., Payne, A.: Okapi at TREC-4. In: TREC (1995)
11. Vechtomova, O., Wang, Y.: A study of the effect of term proximity on query expansion. Journal of Information Science 32(4), 324–333 (2006)
12. Wei, X., Croft, W.B.: Modeling Term Associations for Ad-Hoc Retrieval Performance Within Language Modeling Framework. In: Amati, G., Carpineto, C., Romano, G. (eds.) ECiR 2007. LNCS, vol. 4425, pp. 52–63. Springer, Heidelberg (2007)

A Set of Requirements for Business Process Management Suite (BPMS)

Nemésio Freitas Duarte Filho and Norben P.O. Costa

University of São Paulo (USP)
Institute of Mathematics and Computer Science - ICMC
Phone/Fax: + 55 (16) 3373-8177
{nemesiofreitas,shadow.suporterj}@gmail.com

Abstract. In order to automate and manage the processes of business of the organizations, it is advisable to the use of systems of management of processes of business (BPMS - Business Process Management Suite) with a high level of quality and efficiency. We believe that the definition and proposal for a catalog (set) of specific requirements to these systems is one of the important steps in this sense. This work presents a set of requirements for BPMS, serving the guide to your development and improvement. The catalog has been drawn up on the basis of work in the literature and in the experience of experts (including experiences of assessment and elicitation requirements for systems of BPM).

Keywords: Business Process Management Suite (BPMS), Requirements of Software, Development of systems, Catalog of Requirements, Comparison of Tools.

1 Introduction

The software industry has the challenge to create a wide variety of products with increasing quality and decreasing costs. Several initiatives have been developed in order to resolve it. Therefore, an important one is the recognition and the survey of requirements. The software requirements are developed on the basis of specific aspects of the system, taking into account the mission and domain field of software [1].

During the creation process of a software project, specifically in the field of systems of management of business processes (BPMS), the requirements become more concrete, detailed and complex. Many systems of BPM are built in an isolated manner, possessing architectures and requirements that are not normally treated by most applications.

Even with the advent and relevance of BPMS in the market, a lot of them are developed and delivered with a reasonable amount of failures, particularly in relation to the features needed for efficient use and automation in relation to business processes. Currently the requirements for a system of BPM tend to be a model not yet standardized, and complex to be captured and detailed.

The main contribution of this paper is to present a list of requirements for the development and elaboration of systems of BPM, with a focus on process automation

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 487–496.
DOI: 10.1007/978-3-642-36981-0_45

and its features. The preparation and design of the set of requirements were performed with the support of a team of specialists in management of business processes (analysts, consultants, architects and researchers).

The motivation to establish a catalog of requirements for systems to BPM exists by necessity to establish a standard for the features that must be present in these systems. A set of requirements is not an official document, but rather an artifact subject to periodic review and modifications, due to the technological trends, taking into account the new requirements or needs [2]. Initially the set was prepared by means of assessments between BPMS tools, and was subsequently validated by the team of specialists, using various techniques of elicitation of requirements, such as brainstorm and sessions of JAD - Joint application design [3].

This article is structured as follows. Section 2 presents the reasoning of work, addressing the context for the discussion of requirements and systems of BPM. Section 3 presents the methodology of the research. Section 4 presents the details of the catalog of requirements. And finally, in section 5 we have the conclusions along with considerations for future work.

2 Literature Review

2.1 Business Process Management Suite (BPMS)

Many companies conduct their activities in function of their business processes. Some of the processes are mission critical and essential to the success of the organization; they end up providing competitive advantages in relation to competitors. Other processes may not be mission critical, but still are important for the organization. These processes represent "the central axis of the business of the companies. Under this scenario, many organizations are adopting the idea of BPM in their daily lives.

Initially, it is necessary to differentiate the terminologies BPM and BPMS. According to Schick [4], BPM - Business Process Management is the ability to understand and control the many parts of a process of the organization. Hedge [5], define BPMS - Business Process Management Suite as a software platform, which allows the user design, execute and manage a complete business process, in its entirety. BPMS can be considered as an integrated environment of software components that automate business processes, from conception and initial modeling, passing through execution and monitoring, until the incorporation of improvements.

Arora [6] presents other settings for these terminologies. "The process management or BPM, describes capabilities and technologies that enable organizations to model, automate, manage and optimize their business processes, leveraging the infrastructure of information technology". "The management systems of processes or BPMS are platforms that orchestrate the business processes, along with all systems, giving a complete visibility and control to the managers of processes. They are, therefore, the results of automated processes and managed with the use of case management tools" [6].

The organizations have been presenting countless efforts in automation of their business processes [7]. According to Karagiannis [8], the idea is that first systems of existing workflow had similarities to BPMS, possessing the ability to delegate tasks to

the right people, at the right time, using the appropriate information. Most of the projects of BPMS stems from a desire to improve corporate performance, although a good part seek flexibility as the main motivation.

Some difficulties commonly found in the adoption of a BPMS are the reluctance in accepting the need for change and problems when the existing culture of the organization is changed. For Parkes [9], the most important characteristics for the success of the projects of BPMS are the commitment of management, communication with users and with the management, and collaboration of users with the project.

2.2 Context and Influence of Requirements in the Development of BPMS

The requirements detailed and described in this work were developed based on the specification of projects for Business Process Management Suite (BPMS), based on the experiences of design in the industry and market of BPMS. The requirements reflect the experience gained between the evaluators of tools (specialists) and simple users, who use the systems in their daily lives [10, 11].

The systems of BPM currently software's are complex, highly modular and organized in hierarchical structures. Various variants and products in this line are emerging, containing different functions that in many cases affect end users due to the not understanding in relation to the features in the system. A BPMS standardized in relation to their features is extremely relevant, allowing greater interoperability and ease of use.

The efficiency and standardization in projects for systems to BPM are of extreme importance to the market. Users of tools of BPM are more involved in automation and management of processes than in simple features of systems ERP (enterprise resource planning) and CRM (Customer relationship management). Those applications (BPMS), by having they have a quantity of users not determined, together with the use of aspects linked to the business of the organization, have requirements that are not normally treated by most applications.

3 Methodology

The methodology followed in the work was defined primarily in 4 (four) steps:

1. Design and preparation of research;
2. Comparison of tools;
3. Identification and mapping of requirements;
4. Search and relevance of requirements;

Step 1- The first phase concerns the design and initial aspects for the accomplishment of the research. The problem definition took into account the needs in the scientific area, by checking if the efforts are really valid and relevant, in relation to the lack of a catalog for specific requirements and standardized for tools of BPM.

Step 2- The second step consists primarily of studies and comparisons in relation to the specific features contained in a BPMS. In the second phase of the research were

carried out qualitative comparison between tools of BPM, the assessments have the objective to show and identify the features similar between the evaluated systems.

The tools to be evaluated were chosen leading some criteria, such as for example, the popularity of organizations, tools free or payable, the domestic or foreign origin, trying to combine several tools BPM that had different characteristics. The BPMS chosen were the Aqualogic [12], BizAgi [13], Orchestra [14], jBPM [15], Oracle BPM [16] and Intalio [17]. The 6 (six) solutions analyzed have characteristics that make them, in fact, choices very interesting. Present a set of specific features, covering all phases and aspects considered most important in a BPMS. After the evaluations, the authors identified essential features of these systems, classifying them as criteria necessary for the development and implementation of a system of BPM.

Step 3- The third stage refers to the identification and collection of requirements (guidelines) for the construction of BPMS. This survey was done through meetings with specialists in the area (managers, analysts, project managers), all related to the management of business process. In the meetings were used specific techniques for elicitation of requirements, based on the criteria defined in the previous step, the main techniques used were the use of brainstorm and sessions of JAD (Joint application design).

With the end of the meetings, the authors could gather a considerable number of requirements for each criterion defined in the second step, facilitating the design and preparation of the catalog of requirements.

Step 4- In the end, in the fourth step we have the prioritization of the identified requirements. In this step was performed an online search with about 80 people in the area of management of business processes (entrepreneurs, specialists and researchers). The research helped to prioritize the requirements in order of importance, identifying the attributes most relevant to the context. The research was drawn up with a list of requirements (result of step 3), where the specialist reviewed and prioritized each attribute, opting for some criteria (Little Relevant, relevant, Very important).

At the end of the whole process, to facilitate the understanding of proposed catalog, improving its structure and facilitating its maintenance, was used and defined a hierarchical structure in relation to criteria and defined requirements. The detailing of the hierarchy, together with the analysis of requirements and criteria, are highlighted in the next section "Catalog of Requirements for BPMS".

4 Catalog of Requirements for BPMS

To facilitate the understanding of the catalog of requirements, improving its structure and facilitating its maintenance, was used and defined a hierarchical structure similar to the specification for the evaluation of software product: ISO/IEC 9126 [18] and Bertoa and Vallecillo [19], which are based on criteria, requirements (guidelines) and description.

The hierarchical structure defined in this work resembles the structure proposed in the work of Hoffman [20], in order to facilitate the verification of specific requirements for the construction of a BPMS. For the present work were used three

levels of hierarchy, each having a specific purpose. The three levels of the catalog are presented below:

Level 1 – Specific Criteria of a BPMS: In the first level of the hierarchy were defined criteria specific to a system of BPM, and primordial features that must exist in a system of this area, they are:

- Module of Modeling and Orchestration;
- Usability (Interface with users);
- Support for Business Rules Complex;
- BAM (Business Activity Monitoring);
- ESB (Enterprise Service Bus);
- Security;
- Portability.

Level 2 – Requirements (Guidelines): The second level is formed by development requirements. The requirements represent specific guidelines for each criterion defined in the first level, for example, the criterion "safety" can be decomposed into technical requirements in order to provide more detail and ensure greater security for systems to BPM (validating users, managing sections, encrypting, among others). The requirements of the second level of the hierarchy must be prioritized, because from them will be implemented specific features for each criterion. In general, this level is based on the fact that the requirements raised represent guidelines relevant to the development of BPMS, in other words, defining characteristics and features important to the development of these systems.

Level 3 – Description of the Guidelines: The third level corresponds to the text (description) of aspects related with the guidelines (second level), facilitating its implementation and development. After presenting the levels that will be used in the catalog of requirements, will be done a detailing mainly in elements from the third level. To facilitate visualization of the relevance of the requirements, priorities were defined throughout the research with the experts who took part in the sessions of elicitation of requirements, symbolized by the symbols: (+) little relevance; (++) Relevance; and (+++) Very relevant in order to express the importance and the impact that these requirements have on the systems of BPM. Figure 1 shows the graphical representation of the symbols used to prioritize the requirements. We recommend the priorities in the second level, in order to simplify the description.

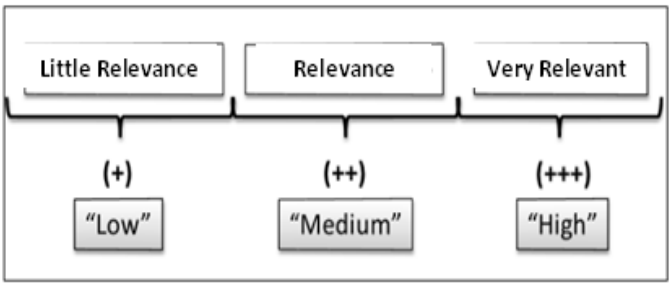


Fig. 1. Graphical representation of the priority requirements

At the end of the whole process, the catalog containing criteria and specific requirements for BPMS, can be viewed below in Table 1.

Table 1. Criteria and specific requirements of a BPMS

Criterion (Level 1)	Requirements (Level 2)
Module of modeling and orchestration	Using a notation language Operations of Users Management Operations of Roles Management Operations of Audit Management
Usability	Intelligibility Learnability Operability Attractiveness
Support for business rules complex	Workflow Client Application Workflow Relevant Data
BAM (*Business Activity Monitoring)*	Supervisory Functions of Processes Functions of States of Processes Repository of Metadata
ESB *(Enterprise Service Bus)*	Scalability Expansibility Performance Interoperability
Security	Reliability Availability
Portability	Adaptability Coexistence Ability to replace Capacity to be installed

Below will be made the analysis and the detailing of each criterion, along with the details of your requirements and priority.

Module of Modeling and Orchestration
Criterion used to graphically develop business processes along with basic features of orchestration, i.e. composition of business processes (via Web Services) where there is the figure of a core process (process master) that controls and coordinates the other processes. In this type of composition, each process participant is not aware of that is part of a composition of processes, with the exception of the master process. The requirements are:

- Using a notation language (+++): language that brings a series of icons standards for the design of processes, which facilitates the understanding of the user. The modeling is an important stage of automation, allowing that processes are discovered and designed;
- Operations of Users Management (+++): create, change, remove privileges of users or group of users.
- Operations of Roles Management (++): set, remove or change roles of the users of workflow;
- Operations of Audit Management (+): create, remove record of progress of the processes.

Usability
Noting that the software interface for the user and one of the strong points for acceptance by the clients, the specific requirements of usability of a BPMS were defined as follows:

- Intelligibility (++): allows the user understand if the software is appropriate and how it can be used for tasks and conditions of use specific;
- Learnability (++): the software should enable the user apprehend its application over time, avoiding that the user will always have recourse to the manual or support services;
- Operability (++): software should enable the user operates it and control it easily and efficiently. Allowing a friendly interaction with the user;
- Attractiveness (++): involves characteristics that allow a greater satisfaction on the part of the clients of the system, including since the adequacy of the information provided for the user, even the refinements visuals used in its graphical interface.

Support for Business Rules Complex
Business rules are defined in various ways, but generally the best definition is: decision points within the business processes. Business rules are described by a logic of the type "if-then-else". The main requirements to be observed are:

- Workflow Client Application (++): applications that relate to the end user. As an example, one may mention the electronic forms of activities, the task manager, control of states and administration functions of processes;
- Workflow Relevant Data (+++): given that controls a particular transition and affects the choice of the next activity to be executed;

BAM (Business Activity Monitoring)
Business Activity Monitoring is also known with the word Real-Time Business Intelligence, referring to business intelligence capable of analyzing and reporting facts and trends in real time. The requirements more relevant to the implementation and use of a BAM are:

- Functions of Supervision of Processes (++): management functions of processes as, for example, change of status and qualification of new version of the process;
- Functions of States of Processes (++): filtration capacity of the processes according some filter, for example, all the processes completed;
- Repository of Metadata (+++): allows a BPM can store information on the processes in its execution.

ESB (Enterprise Service Bus)
Enterprise Service Bus (BUS), also called Corporate Bus Services, is the infrastructure of SOA architecture, considered by many experts the heart of an infrastructure service-oriented. Its responsibility is to define an interface to input and output, processing of messages, treatment of exceptions and monitoring of messages. The main requirements were defined as follows:

- Scalability (+++):capacity with the BPMS support a specific number of users in the same section, observing that these systems can be used by several users at the same time;
- Expandability (+++): ease with which new services can be added or extended within a system of BPM without affecting the other services and features of the system. This characteristic is important mainly for the reason that currently the business environment of a software is in constant change and evolution;
- Performance (++):ensure maximum efficiency and agility as possible in the time of response of a service, i.e. reduce the elapsed time from when the request is received by the BUS until a response (XML) is issued to the consumer;
- Interoperability (+++): it is the ability of a service to communicate transparently with other service or system. For a service to be considered interoperable is very important that he work with patterns such as XSD, BPEL, WSDL, XML.

Security
Relevant criterion for systems to BPM in virtue of its applicability and use. Due to the fact that many systems are accessed via the web (remote access), there are some questions that should be guaranteed for greater security. The requirements raised within the criterion safety can be seen below:

- Reliability (+++): the access to the information/service can only be granted to authorized users, avoiding that the system or information are used by parties that are not tied to the contracting organization of the system;
- Availability (+++): capability of the system to keep your operation online on the server of the provider, minimizing possibility of interruption of service applications.

Portability
Capacity of the system be transferred from one environment to another. We must consider all the factors of adaptation, such as different conditions of infrastructure (operating systems, versions of data banks) and different types and hardware resources (such as taking advantage of a greater number of processors or memory). The requirements defined for this criterion are:

- Adaptability (++): capacity of the system to be adapted to different environments specified, without the need for application of other actions or means in addition to those provided for the purpose of the software;
- Coexistence (++): capacity the BPMS coexist with other independent systems in a common environment, independent of the door of communication used and sharing resources available;
- Ability to replace (+): represents the capacity that the system has to replace another specified system, in a context of use and environment specific;

Ability to be installed (++): identifies the ease with which you can install the system in a new environment.

5 Conclusion

In this work we present a catalog (set) of specific requirements for the construction of functionality for systems of management of business processes (BPMS). The main motivation for this research was the fact the absence of patterns in relation to the specific requirements for BPMS.

The catalog of requirements was defined and detailed through research of related work, comparisons of systems of BPM and participation of experts providing the elicitation of requirements specific to this field. The catalog was developed according to a hierarchy based on three (3) levels. The hierarchy helps to use, update, and exemplify the requirements raised, facilitating the understanding by both developers and users.

The hierarchy proposed for the catalog of requirements also allowed the prioritization and relevance in relation to each one of the requirements. The prioritization was important, because many are the criteria and requirements to be met during the development and implementation of a BPMS. From this, the companies may choose to implement a requirement with greater relevance and impact to the area in question, ensuring higher quality and reduction in time during development.

At the end we believe that the proposed catalog may capture the needs, expectations and constraints not only for developers of BPM systems, but also for those who use such tools. THE catalog must be used as a guide of "best practices" in relation to BPMS and be conducive to changes, because the catalog is subject to periodic reviews and changes due to technological trends and innovations in management of business processes.

In spite of this catalog have been influenced and directed to BPMS, the same requirements can be quite useful in similar systems, examples: Workflows, Enterprise Resource Planning (ERP), CRM (Customer Relationship Management), since they have similar domains. The authors recommend as future work, the survey and the identification of a quality model specific for systems of BPM, the requirements defined in the catalog of this work may serve of guideline for the proposal of this new model of quality. The ISO/IEC standards 9126 and 14598 may be used to support the definition and detailing of this model of quality.

References

1. Jacobson, I., Booch, G., Rumbaugh, J.: The Unified Software Development Process. Addison-Wesley (1999)
2. Weber, M., Weisbrod, J.: Requirements Engineering in Automotive Development - Experiences and Challenges. IEEE Software 20(1), 16–24 (2003)
3. Wood, J., Silver, D.: Joint Application Development. John Wiley & Sons (1995)
4. Schick, S.: Edmonton Power company rewrites billing system. Computing Canada, Willowdale 32(3), 4 (2006)
5. Hedge, A.J.: Business Process Management: Management Tools. AIIM E - Doc Magazine 19(4), 52–53 (2005)
6. Arora, S.: Business Process Management: process is the enterprise. BPM-Strategy (2005)

7. Reijers, H.: Implementing BPM systems: the role of process orientation. Business Process Management Journal 12(4) (January 2006)
8. Karagiannis, D.: BPMS: Business Process Management Systems. SIGOIS Bulletin 16(1) (March 1995)
9. Parkes, A.: Critical Success Factors in Workflow Implementation. Department of Accounting and Business Information Systems. The University of Melbourne, Melbourne (2002)
10. John, G.: Using a Common Information Model as a Methodological Basis for a Tool-Supported Requirements Management Process. In: Fairbairn, A., et al. (eds.) Proceedings of the 9th INCOSE Symposium, pp. 59–67. Ascent Logic Corp., Brighton (1999)
11. Tvete, B.: Introducing Efficient Requirements Management. In: Proceedings of the 10th International Workshop on Database and Expert Systems Applications (1999)
12. Aqualogic - BEA Systems (2012), http://www.oracle.com/bea/ (accessed in September 25, 2012)
13. BizAgi - Business Process Management Software (2012), http://www.bizagi.com/ (accessed in October 12, 2012)
14. Orquestra. Cryo Technologies (2012), http://www.cryo.com.br/Solucoes/ (accessed in September 10, 2012)
15. Jbpm. jBoss jBPM (2012), http://www.jboss.com/products/jbpm/ Accessed in September 15, 2012)
16. Oracle - Oracle Business Process Management (2012), http://www.oracle.com/us/technologies/bpm/ (accessed in September 26, 2012)
17. Intalio. Product Components (2012), http://www.intalio.com (accessed in September 2, 2012
18. ISO/IEC 9126. Information Technology - Software Product Evaluation (The International Standardization Organization) - Quality Characteristics and guidelines for their Use (1991)
19. Bertoa, M.F., Vallecillo, A.: Quality attributes for cots components. In: Proceedings of the 6th ECOOP Workshop on Quantitative Approaches in Object- Oriented Software Engineering (QAOOSE 2002), Spain (2002)
20. Hoffmann, M., Kuhn, N., Weber, M., Bittner, M.: Requirements for requirements management tools. In: Proceedings of the IEEE International Requirements Engineering Conference (RE 2004), Kyoto, Japan, pp. 301–308 (September 2004)

High Level Architecture for Trading Agents in Betting Exchange Markets

Rui Gonçalves[1], Ana Paula Rocha[2], and Fernando Lobo Pereira[1]

[1] Department of Electrical and Computer Engineering
[2] Department of Informatics Engineering
School of Engineering, Porto University (FEUP)
Rua Dr. Roberto Frias s/n, 4200-465 Porto
Portugal
{rjpg,arocha,flp}@fe.up.pt
http://www.fe.up.pt

Abstract. This paper presents a new high level software architecture used to develop automated trading agents in betting exchange markets. *Betfair* betting exchange entity is used as case study. The implementation of the *Scalping, Swing*, and *Trailing-Stop* trading mechanisms is explored in detail. A *rule-based* trading agent, used as test example, is also described together with results of simulations and real interaction with *pre-live* horse race markets.

Keywords: Software Framework, Agents, Automated Trading, Java.

1 Introduction

A betting exchange is an entity which offers trading services to buy and sell bookmaking contracts. These contracts are structured as binary options (win or lose) where the payoff is either some fixed amount of money or nothing at all, depending on the outcome of a future event. Betting exchanges trade heavily on sports events, but also offer markets on elections and other type of events[1]. In analogy to the financial markets the buy and sell operations are replaced by betting for and against (Back and Lay). This paper presents a software framework, developed in Java, used to instantiate trading agents in this kind of markets. Using automated information and trading platforms has become a requirement for successful trading. Its advantages as compared to conventional trading include, for example, an unprecedented speed of processing, level of integration with data providers, and a wide array of built-in instruments [2][3].

The work presented in this paper is based on the use of *Betfair* exchange entity [4] to provide an environment where is possible the application of different trading policies, in particular on horse racing markets. *Betfair* processes more than five million transactions (such as placing a bet) every day which is more than all European stock exchanges combined. *Betfair* is available 24 hours a day 7 days a week.

Section 2 of this paper gives an overview about the trading environment in betting exchange markets. Section 3 describes the trading framework and its components.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 497–510.
DOI: 10.1007/978-3-642-36981-0_46

Section 4 goes into detail over the implementation of the *Trading Mechanisms*. Section 5 explains one instantiation example of a trading agent, and section 6 summarizes the main results.

2 Trading Environment

Betfair betting exchange includes markets, e.g. tennis match or horse race. On each market there are runners, e.g. horses in a horse race. On runners the Back/Lay bets are placed. Back bet means that the runner will win, Lay bet that it will lose. Fig. 1 illustrates the classical view of a runner in the market. In the middle column at green is the price scale. The price can also be referred as odd. Bets are placed on a given price, which represents the chances of the runner win. E.g. the price 2.0 is a 50% of chances (1/2 = 0.5), price 1.01 is a 99% of chances (1/1.01 = 0.99), 1000 is a 0.1% of chances (1/1000 = 0.001). On the left side, next to the prices column, in pink is the amount column formed by the Lay bets. On the right side, the column in blue, next to the prices column, is the amount column formed by the Back bets. The first and fifth column represent our own unmatched bets. In this case (Fig. 1) we have the following bets placed (but not matched yet):

- Lay of 10.00€ at 4.30 (Lay 10@4.3);
- Back of 10.00€ at 4.70 (Back 10@4.7); and
- Back of 5.00€ at 4.60 (Back 5@4.6).

The last column on the right, in gray, is the volume column which represents the amount matched at every price of the ladder since the market was open. If a bet is placed at a price that "the market" is willing to buy, the bet will be matched at the best price offered. For example, in the market state of Fig. 1 if a Back bet 15@4.4 is

		5,4			32
		5,3			119
		5,2			29
		5,1			20
		5,0	250		93
		4,9			68
		4,8	263		24
		4,7	148	10,00	70
		4,6	349	5,00	76
	8	4,5			217
	2	4,4			23
10,00	10	4,3			4
	448	4,2			
	398	4,1			
	335	4,0			
		3,95			
		3,90			
		3,85			
		3,80			
		3,75			
		3,70			

Fig. 1. Ladder data example of a runner in the market

placed (on the blue side) it will match 8.00@4.5 and 2.00@4.4 , and will leave the remaining 5.00@4.4 unmatched (waiting for someone to buy with a Lay bet). The traded volume information will be updated. This is how the prices move in the market. Since this bet was matched at two different prices, the global matched price of this bet can be calculated using equation 1.

$$Price\ Average = \frac{\sum_{n=1}^{n}(Price_n \times Amount_n)}{\sum_{n=1}^{n}(Amount_n)} \quad (1)$$

If a Back is placed above the best offer in the market (4.5 in Fig. 1), for example 15@4.9, it will stay in the market unmatched and, so for, waiting to be matched. The same happens to a Lay bet if it is placed at a lower price than the best offer (counter bet waiting to be matched). Only unmatched or partial unmatched bets can be canceled.

The profit of a Back bet is calculated using equation 2 and the liability (in case of loss) of a Back bet is the amount of the bet itself.

$$Profit\ Back\ Bet = Amount\ Back \times (Price\ Back - 1) \quad (2)$$

The liability or amount in case of loss of a Lay bet is given by equation 3 and the profit is the amount of the bet itself. Basically the Lay is the "mirror" of Back.

$$Liability\ Lay\ Bet = Amount\ Lay \times (Price\ Lay - 1) \quad (3)$$

Using combinations of Back/Lay it is possible to assure profit (or loss) before the end result of an event, using the price movement on the runner (odd variation). Example of a trade where it is not need to know end result of an event to have secure profit:

- Back of 2.00e at 2.12 (Back 2@2.12) Matched; and
- Lay of 2.00e at 2.10 (Lay 2@2.1) Matched;

For a bet to be matched it must become the best offer in the market and it has to be purchased with a counter bet. When the runner is a winner, then the profit (Back bet) – loss (Lay bet) is:

$$2 \times (2.12 - 1) - 2 \times (2{:}10 - 1) = 2.24 \times 2.20 = 0.04$$

When the runner is a looser, then the profit (Lay bet) – loss (Back bet) is:

$$2 - 2 = 0$$

Notice that if we have this kind of Back/Lay bet combination, on the same runner with the same amount at different prices, there will be profit (if the Back price is higher than the Lay price) or loss (if the Back price is lower than the Lay price) only if the runner in question wins the event. If any other runner wins the event and the combination of a Back/Lay bets have the same amount, the profit/loss will be 0. To distribute the profit/loss equal for all runners (outcomes) the amount of the bet used to

close the trade must be recalculated. This process is called "*do the greening*" or "*hedging*". If a Back position is open on the market, the amount to close the position with the corresponded Lay bet is calculated using equation 4.

$$Close\ Amount\ Lay = \frac{Price\ Open\ in\ Back}{Price\ Lay\ to\ Close} \times Amount\ Open\ in\ Back \quad (4)$$

If a Lay position is open on the market, the amount to close the position with the corresponded Back is calculated using equation 5.

$$Close\ Amount\ Back = \frac{Price\ Open\ in\ Lay}{Price\ Back\ to\ Close} \times Amount\ Open\ in\ Lay \quad (5)$$

More information about this type of trading environment can be found in [5] and [6].

3 The Framework

The software framework described in this paper is an event-based architecture [7]. The connections between modules are made through interfaces based on architecture pattern promoting the production, detection, consumption of, and reaction to events. Fig. 2 illustrates the main modules and their connections. Parallel-processing approaches [8][9] were also applied, in the sense that it is possible to instantiate several trading agents, with different policies, running in parallel managing several trades simultaneously. To ensure that the reproduction of your illustrations is of a reasonable quality, we advise against the use of shading. The contrast should be as pronounced as possible. The following sections will provide a brief description of each module.

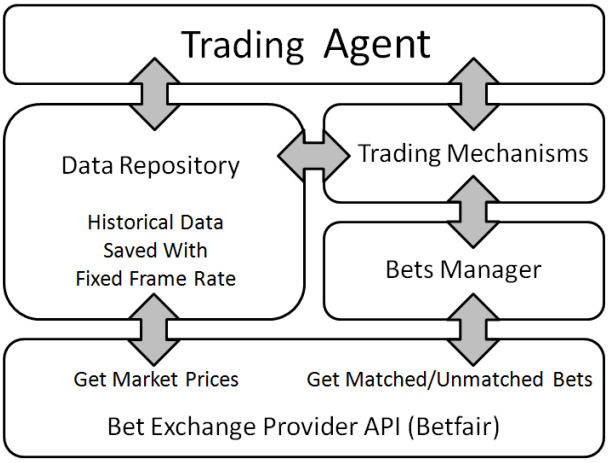


Fig. 2. High level architecture for automated betting exchange

3.1 Betfair API

This section reports the *Betfair* API capabilities and limitations, used as case study for this paper. *Betfair* has its own API to allow software developing companies make software which accesses *Betfair* data. In this case *Betfair* is a service provider and other companies are clients to its services. The API can be used in different programming languages which support web services. Usually it is used to develop trading software or software for tipsters. Its goals are speed and manageability. *Betfair* API is also used to develop 100% autonomous agents.

For the framework described in this paper, *Betfair* API module is used as low layer for communication with *Betfair* server. This low level layer interacts with the *Data Repository* module, providing data about the prices and volumes of each runner, and the *Bets Manager* module, providing data about the states of the bets and also placing and canceling bets. The main *Betfair* API services [10] used for these actions are:

- *Data Repository*
  - Get Complete Market Prices; and
  - Get Market Traded Volume.
- *Bets Manager*
  - Get Matched and Unmatched Bets;
  - Place Bets;
  - Update Bets; and
  - Cancel Bets.

Depending on the type of licensing there are different numbers of calls permitted per minute for each service.

3.2 Data Repository

The *Data Repository* module is responsible for data gathering, inform listeners about new data or events in the market, save the data, and replay saved data. The main interface of this module is shown on listing 1. The *Market Update* event type simply informs listeners that new data about the runners prices and volume has arrived. Since we can do trade before the race or game start (before the market is *in-play* mode) we have the event type *Market live* that is activated when the market turns *in-play*. The *Market Suspended* event type is activated when the market is suspended. The market can be suspended for different reasons depending on the type of market, e.g. in a soccer match the market is suspended after a goal until the game restart. Also, right after the markets become *in-play*, the markets are suspended for a small time. The *Data Repository* module can be instantiated with a new market by an external object. When this happens the *Market New* event is delivered.

This module can be connected to the *Betfair* server, through the *Betfair* API, or to saved files to be used for data review and analysis (using data "visualization agents")[11] or in simulation, as explained in section 4.4.

Listing 1. Market Change Listener Interface

```
1  public interface MarketChangeListener {
2        public static final int MarketUpdate =0;
3        public static final int MarketLive =1;
4        public static final int MarketClose =2;
5        public static final int MarketSuspended =3;
6        public static final int MarketNew =4;
7
8        public void MarketChange ( MarketData md , int marketEventType ) ; }
```

3.3 Data Repository

The *Bet Manager* module assures the placement of bets and manages all objects with a trade interface (i.e. *Trading Mechanisms*) informing them about the state of their bets. It is important to centralize all the bets processing in order to optimize the number of calls to the "*Get Matched and Unmatched Bets*" service, which is limited. Also sometimes the *Betfair* API does not return the ID of a placed bet, leaving the program unclear about the placement. In these cases the *Bets Manager* module tracks the bets without "owner" and re-assigns them correctly.

3.4 Trading Agents

To instantiate a *Trading Agent* object it just has to be extended to the abstract superclass "*Bot*". This class implements all the interfaces and virtual methods in order to interact with the *Data Repository* and *Trading Mechanisms* modules. The *Trading Agents* are normally attached to one market observing one runner but is possible (and useful e.g. for *dutching* and *bookmaking* techniques) to put a *Trading Agent* observing several markets and runners. The *Trading Agent* object can also initialize *Trading Mechanisms* objects (processes of trading, see section 4) whenever it takes some decision about a runner forecasting. When a *Trading Mechanism* starts, running in parallel, the *Trading Agent* is informed about the state of the trade. On top of this high-level element it becomes easy to implement decisions policies interacting with the markets, since simple *rule-based* decisions policies to more complex methodologies e.g. time series predictions.

4 Trading Mechanisms

In this paper we will focus on three of trading methodologies aimed at maximizing trading profits and minimize loss:

- Scalping;
- Swing; and
- Trailing-Stop.

In the framework, these methods are implemented on the *Trading Mechanism* module (fig. 2). After one agent (*Bot*) parameterizes and instantiates one trade mechanism, this trade mechanism will run in parallel and will inform the owner agent along the way about the state of the trade. Ultimately it will inform the agent that trade is over and the amount of profit/loss for the operation.

4.1 Scalping

Scalping is short term trading. A scalping trader looks to make lots of small profits, which in time add up. Scalping relies on lots of active participants in the market. Scalping works better in markets with high liquidity. The concept is simple, if a back bet is matched at a certain price, a lay bet must be placed right in the next lower price, or, if a lay bet is matched at a price a back bet must be placed right in the next higher price in the ladder to make profit. The profit/loss is equal to the difference, or spread, between the Back and Lay price, has explained in section 2. The *Betfair* betting exchange is an ideal place to trade in this way. Mainly in horse racing because there is high liquidity being matched in these markets, in particular just before the start of the race. Scalping the market means trading in the market tick by tick. One tick is one step in the prices scale of the ladder. For example if a Back at 2.12 is placed, one successful scalp will close the position with Lay at 2.10 (one tick down). If a trade starts with a back it means that the price was forecast to go down. If it is predicted to go up the scalp starts with a Lay bet.

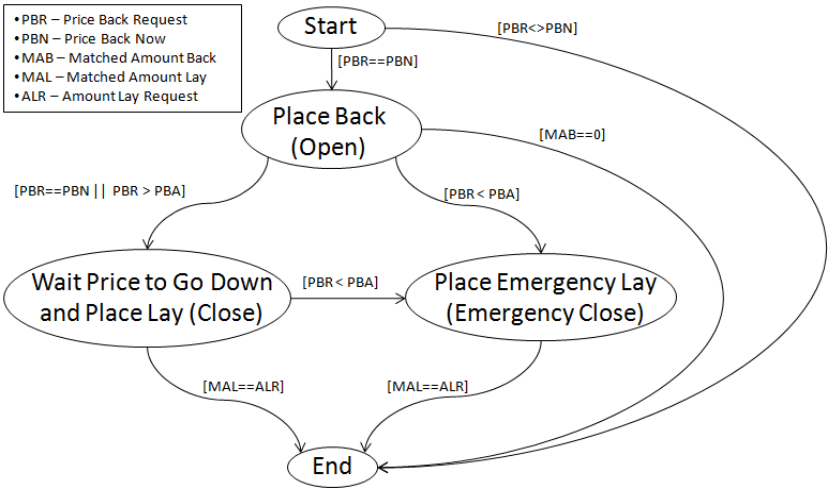


Fig. 3. Simplified graph schema for a Back->Lay scalp implementation

Fig. 3 represents the state machine used to process one Back->Lay scalping (prediction for the prices to go down). One Lay->Back scalp will be a "mirror" of this state machine. The *Price Back Request- PBR* is the price where the agent enters the market. If the prices already moved [*PBR* <> *PBN*] when the order reach the scalp

module (the start state), it will assume the opportunity was lost (the prices already went down) or the prices went in the wrong predicted direction (up) so it ends the process without doing nothing. Otherwise [*PBR == PBN*], opens position on the market with a Back bet. After the bet is ordered to be placed, if the bet was not matched it will end the trade (canceling the bet) because it will assume the prices already move down, while the order was reaching the *Betfair* server, and the opportunity was lost. Otherwise it will try to close the position placing a Lay bet. If the price goes one tick down it will close the trade with profit. If the price does not move it will wait. If the price goes up it will close in "emergency" with loss.

In listing 2 is the declaration of the constructor method of the object that runs the scalping process. The *MarketData md* argument identifies the event (horse race, soccer game, tennis..), the *RunnersData rd* identifies the runner where is about to scalp. *double entryAmount* is the initial stake of the trade and *double entryPrice* is the entry price to open position on the market. The *int waitF ramesNormal* is the number of actualizations received from the *Data Repository* before it tries to close at the same price he entered (no profit or loss). After the *int waitF ramesNormal* expire the *int waitFramesEmergency* also starts a count down and after expire it will close at the best offer available to place the counter bet (emergency close with loss). The *Bot botOwner* is the owner agent of the trade, used to be informed about the state of the scalp. Finally the *int direction* argument indicates the predicted direction of the prices movement.

Listing 2. Parameters for the Scalping process

```
1  public Scalping(MarketData     md,
2                  RunnersData     rd,
3                  double      entryAmount,
4                  double     entryPrice,
5                  int waitFramesNormal,
6                  int waitFramesEmergency,
7                  Bot  botOwner,
8                  int direction);
```

4.2 Swing

The swing methodology is very similar to the scalping. The main difference is the number of ticks the prices have to move in order to enter the close state (profit or loss close). On the swing methodology is possible to define the offset number of ticks to close in profit and the offset number of ticks to close in loss. If the prices stay inside this interval offset (up and down) it does not do nothing. Swing with offset of 1 tick for profit and offset of 1 tick for loss is the same as scalping.

In listing 3 is the constructor for the swing process initialization. Besides the same parameters present on the scalping constructor there are the *int ticksUp* and *int ticksDown* representing the offset number of ticks to close in profit and loss (depending in the direction parameter). There is also *boolean frontLine* and *int waitF ramesOpen* new arguments. These arguments are used when the agent does not want

to enter in the market where offers is available but wants to wait until the market reach the price given in the *entryPrice* argument. If *waitF ramesOpen* expires and the market does not consume/match the entry bet it will cancel the trade process. If *frontLine* = *true* it will ignore this time (*waitF ramesOpen*) and assume the agent wants to enter the market at the *entryPrice* argument where the counter offer is available.

Listing 3. Parameters for the Swing process

```
1  public Swing(MarketData Market,
2                RunnersData rd,
3                double entryAmount,
4                double entryPrice,
5                boolean frontLine,
6                int waitFramesOpen,
7                int waitFramesNormal,
8                int waitFramesEmergency,
9                Bot botOwner,
10               int direction,
11               int ticksUp,
12               int ticksDown);
```

4.3 Trailing-Stop

The trailing-stop methodology is used when the agent is looking to catch a much broader trend in a market but wants to retain a stop loss condition if the trend starts to turn. After a position is open in the market (open bet is matched), the close bet is placed with a tick offset behind, and moves only when the price moves in the predicted direction. Eventually the price will move in the reverse direction reaching the close bet.

Fig. 4 represents the state machine used to process this method, for the price prediction to go down (Back to open -> Lay to close). The state "*Place Lay N Ticks Above PBN*", being *N* the tick offset to follow the price, is repeated when [*PBP* > *PBN*] the runner price moves in the predicted direction (down). If [*MAL* == *CAL*] the close bet is completely matched, it means the price went in the reverse direction (up) and reach the close bet, closing the trade.

Listing 4 is the constructor method of the object that runs the Trailing-Stop process. The *int offset* is the number of ticks offset to follow the runner price.

4.4 Simulation

This section addresses the issues and limitations to achieve simulation in the betting exchange markets. This process implies the simulation of the bet placement. There are two main problems to simulate a bet placement on the market.

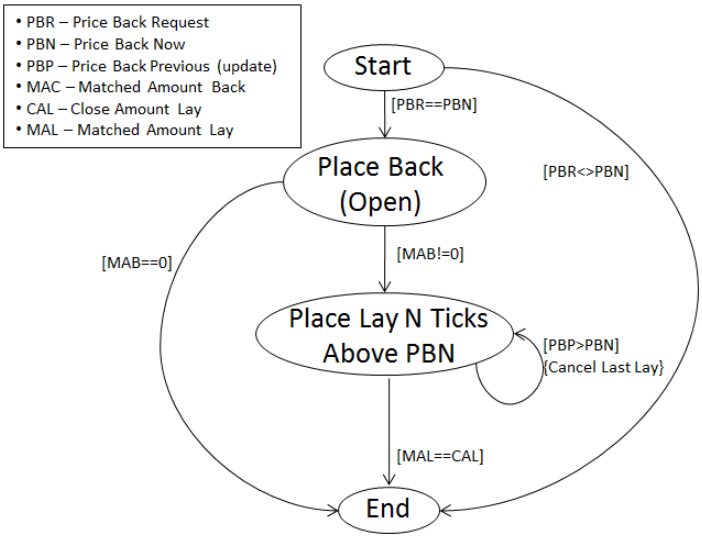


Fig. 4. Simplified graph schema for a Back->Lay Trailing-Stop implementation

Listing 4. Parameters for the Trailing-Stop process

```
1  public TrainlingStop(MarketData      Market,
2                       RunnersData     rd,
3                       double     stakeSize,
4                       double     entryPrice,
5                       boolean frontLine,
6                       int waitFramesOpen,
7                       int waitFramesNormal,
8                       int waitFramesEmergency,
9                       Bot  botOwner,
10                      int direction,
11                      int offset);
```

The first problem is the bet amount influence on the market. The unmatched bets will not appear on the real market. This issue is impossible to bypass, since, in simulation, the actual amount of the bets is not placed. For example, this limitation makes it impossible to simulate and test trading agents relying on *spoofing* methodologies.

The second problem is the simulation of the matching process. The bets, of all traders on the market, are placed in a *FIFO* queue of bets for each price on the runner. It is impossible (there is no data provided by *Betfair* API about it) to know in which position in the queue is our bet. It is possible to have an approximate idea by looking to the volume matched in the placement price and monitoring this volume evolution. But since these markets are of very high frequency trade, it is impossible to know the exact volume on the price when the placement order reaches the *Betfair* server. Also it is impossible to know if canceled bets were ahead or behind our bet, damaging the process of volume monitoring to try to solve this issue.

For the described framework we assume the worst case scenario: the bets are considered matched when the amount available on the specific price reaches 0 (Lay or Back side depending on the type of bet). Having these assumptions, it is possible to simulate and test agents in replay mode or in real-time connection to *Betfair*.

5 Agent Instantiation Example

This section presents the policy of a *rule-based* trading agent. This agent was developed with the purpose of testing the framework in simulation mode and in real-time interaction with *pre-live* (before the start of the race) horse racing markets. *Rule-based systems* consist on defining a set of static conditions that, when verified, all together, the agent enters the market. *Rule-based system* is a type of reactive agent based system [12]. Normally documentation or software found about this type of trading agents is used as a teaching tool.

Listing 5. Kernel of a rule-based agent example

```
1  if(Utils.isValidWindow(rd,MecanicBot.WINDOW_SIZE,StudyBot.WINDOW_SIZE)          &&
2      Utils.isRdConstantOddLayInWindow(rd,MecanicBot.WINDOW_SIZE,0)               &&
3      Math.abs(Utils.oddToIndex(Utils.getOddBackFrame(rd,          0))
4                          Utils.oddToIndex(Utils.getOddLayFrame(rd,  0)))<=1      &&
5      Utils.getOddBackFrame(rd,  0)<20                                            &&
6      Utils.isAmountLayGoingDown(rd,MecanicBot.WINDOW_SIZE,0)                     &&
7      Utils.isAmountBackGoingUp(rd,MecanicBot.WINDOW_SIZE,0)                      &&
8      Utils.isAmountBackBiggerThanLay(rd,0)                                       &&
9      Utils.getWomAVGWindow(rd,MecanicBot.WINDOW_SIZE,MecanicBot.WINDOW_SIZE)
10                 <Utils.getWomAVGWindow(rd,MecanicBot.WINDOW_SIZE,0))
11     {
12       writeMsg("Runner "+rd.getName()+" going Up",Color.BLACK);
13       double odd=Utils.getOddBackFrame(rd,       0);
14       if(odd>10)
15         new Swing(md,rd,2.50,odd,true,0,100,80,this,1,1,1);       // same as scalping
16       else if(odd>5)  ␣                    ␣       ␣
17         new Swing(md,rd,2.50,odd,true,0,100,80,this,1,2,1);
18       else if(odd>4)
19         new Swing(md,rd,2.50,odd,true,0,100,80,this,1,2,2);
20       else if(odd>3)
21         new Swing(md,rd,2.50,odd,true,0,75,60,this,1,3,2);
22       else if(odd>2)
23         new Swing(md,rd,2.50,odd,true,0,75,60,this,1,3,3);
24       else
25         new Swing(md,rd,2.50,odd,true,0,50,40,this,1,4,3);
26      }
```

In listing 5 is the kernel code segment used to decide the entry point on the market. This code is part of an agent class named "*Mecanic Bot*" (*MecanicBot class* extended to the *Bot* super class). The segment of code is called every time a new frame of information, about the runners prices, arrives into the *Data Repository* module. For this example, if a *Trading Mechanisms* (swing in this case) associated with this agent is in trade processing, this code is not processed. This agent looks to one runner, i.e. horse, and processes one trade at time.

For this example there is a conjunction of eight rules to forecast prices to move up. These rules are tested over a small time of data on the near past (*WINDOW_SIZE* = 5 samples of data).

Rule 1 - Line 1 Test if there is enough data available for processing (2 × *WINDOW_SIZE* numbers of samples);
Rule 2 - Line 2 Test if the prices (on the Lay side) were constant on the near past;
Rule 3 - Lines 3 - 4 Test if there if no gap between the prices of Lay and Back bets, on the runner on the last frame of information received;
Rule 4 - Line 5 Test if the runner has the actual price under 20 - works as a filter;
Rule 5 - Line 6 Test if the amount available to match Lay bets is going down, on the best offer price (tests if Back bets are being matched or canceled);
Rule 6 - Line 7 Test if the amount available to match Back bets is going up, on the best offer price (tests if Lay bets are being placed);
Rule 7 - Line 8 Test if the amount available to match Back bets is bigger than the amount available to match Lay bets on the last sample frame received; and
Rule 8 - Lines 9 - 10 Test if the evolution of the *Weight of Money* (WOM) if going up - If there are more Lay bets being placed then Back bets at different prices around the best offer.

If these rules are all verified the agent instantiate a different swing *Trading Mechanism* parameterization depending on the type of the actual price (odd) of the runner (horse). To understand the arguments of the swing mechanism instantiation consult listing 1.3. Horses with lower prices (odds) are favorites for winning the race. In this type of horses the liquidity is bigger and the price movements are wider and faster than on loser horses. The tuning process of choosing the parameters used in listing 1.5 was made by hand, through a series of tries, using the simulation mode.

6 Conclusions

For testing the framework, described in this paper, 20 instantiations of *Trading Agents* were used. All running in parallel and using the same policy, described on the previous section. The policy was also mirrored to forecast price movements in both directions. The agents were used in UK horse race events and activated every 10 minutes before the start and deactivated at the race start (*pre-live* trading). There were 20 agents running in parallel because these races can have up to 20 horses (one agent per runner). If the race, in processing, has less than 20 horses the agents without

assigned horses stay in standby. The sample frame rate of retrieved data was fixed to 2 samples per second, on the *Data Repository* module. The retrieved data was also saved to be tested in simulation mode. The agents stayed in processing seven days a week 24 hours per day. We used stakes of 2,50€ for testing. This low amount does not influence the general trends of the markets. Fig. 5 shows the test results, in real interaction with the markets and in simulation, made between November 16th and December 17th of 2011.

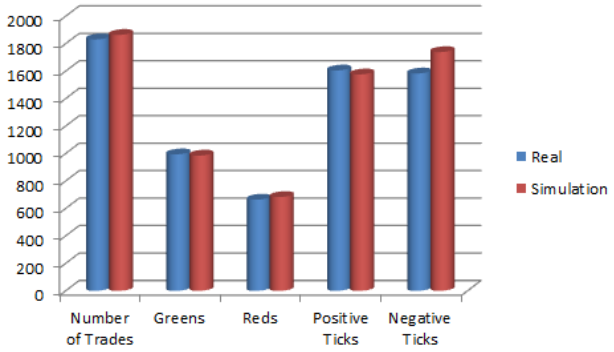


Fig. 5. Results of the trading agents in real and simulation *pre-live* UK horse racing

The "Number of trades" is the number of times one of the agents execute one *Trading Mechanism*, swing in this case. This value is not equal in real races and simulation because many times when the opening bet placement of the trade reach the *Betfair* engine, on the server, the prices already moved and the trade is canceled (as explained in section 4) - in simulation it always catches the actual price available to be matched. The "Greens" is the number of times the Trading Mechanism closes in profit. The "Reds" is the number of times the Trading Mechanism had to close in loss. The sum of the "Greens" and "Reds" is not equal to the "Number of trades" because sometimes the *Trading Mechanism* can close the trade on the same entry price, without profit or loss. This happens when the *Trading Mechanism* reach the timeout exposure on the market (parameter explained in section 4).

The "Positive Ticks" is the number of total ticks that result in profitable trades. The "Negative Ticks" is the total number of ticks that result in loss. These are the main values to get to conclusions about how well succeeded is one trading policy. The number of greens is higher than the number of reds, however the total number of ticks in profit ("Positive Ticks" – "Negative Ticks") is almost zero in real tests and a small negative value in simulation. What happens is that, in the Swing mechanism, many times the prices reach the complete offset ticks in loss but do not reach the total offset ticks in green. Many Swing trades close in green before the price reaches the complete offset to close in maximum potential profit, due to the timeout exposure parameter. This is explained by the fact that *pre-live* trading in horse racing relies a lot on traders that are *spoofing* the market, yielding small displacements on the predicted trend and then big movements on the opposite direction. The total number

of ticks that result in positive trades is lower in simulation because in simulation we assume the worst scenario in the matching process, as explained in section 4.4.

The presented software framework has proven to be resilient enough to execute trading policies continuously, instantiating and running different *Trading Mechanisms* in parallel. For future test we intend to develop a *Trading Agent* based on artificial intelligence techniques, namely NNs, to capture trends more consistently.

Acknowledgment. The research leading to these results has received funding from Fundação para a Ciência e a Tecnologia (FCT - REF : SFRH / BD / 70987 / 2010) and ISRP - Instituto de Sistemas e Robótica do Porto (R&D Unit 147).

References

1. Chen, Y., Goel, S., Pennock, D.M.: Pricing combinatorial markets for tournaments. In: Proceedings of the 40th Annual ACM Symposium on Theory of Computing, STOC 2008, pp. 305–314 (2008)
2. Williams, L., Babcock Jr., B.: Guide to trading systems. Dow Jones-Irwin (March 1989)
3. Deboeck, G.: Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets. Wiley Finance
4. Daniel Shapiro Daniel Shapiro, Daniel Shapiro: `www.betfair.com`: World-wide wagering. Communications of the Association for Information Systems, 149–161 (2005)
5. Davies, M., Pitt, L., Shapiro, D., Watson, R.: Betfair.com: Five technology forces revolutionize worldwide wagering. European Management Journal 23(5), 533–541 (2005)
6. Laffey, D.: Entrepreneurship and innovation in the uk betting industry: The rise of personto-person betting. European Management Journal 23(3), 351–359 (2005)
7. Haussermann, K., Levi, P., Zweigle, O., Kappeler, U.: Event based distributed real-time communication architecture for multi-agent systems. In: 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT), November 30-December 2 (2010)
8. Kramer, J., Magee, J., Dulay, N.: A constructive development environment for parallel and distributed programs. In: Proceedings of 2nd International Workshop on Configurable Distributed Systems, March 21-23 (1994)
9. Wang, J., Yau, S.S., Bae, D.-H.: An architecture-independent software development approach for parallel processing systems. In: Proceedings of Nineteenth Annual International Computer Software and Applications Conference, COMPSAC 1995, August 09-11 (1995)
10. Betfair. Sports API Reference Guide - v1.101. The Sports Exchange API Documentation
11. Chen, H., Schumaker, R.P., Solieman, O.K.: Sports data mining. Integrated Series in Information Systems, vol. 26. Springer (2010)
12. Eleftherakis, G., Gheorghe, M., Kefalas, P., Holcombe, M.: Formal development of reactive agent-based systems. In: Encyclopedia of Information Science and Technology, IAV 2007, 2nd edn. IGI Global (2009)

An MDA Approach to Develop Web Components

José Luis Herrero Agustin, Pablo Carmona, and Fabiola Lucio

Department of Computer and Telematics Systems Engineering
University of Extremadura
Avda. de Elvas s/n, 06006, Badajoz, Spain
{jherrero,pablo,flucio}@unex.es

Abstract. Model-Driven Development (MDD) supports the automating of code generation by performing a set of transfomations between models. This approach is currently been applied to specific domains and in particular in the web domain, and this is because web aplications have evolved with the appearance of AJAX and Web 2.0 technology, and a new breed of applications for the Internet has emerged. However, as web applications become more and more complex, the performance degree is negatively affected, since the initial stages of software life cycle are not incorporated into the development process of this type of applications. In order to solve this problem, this paper proposes a model-driven architecture to support web application development from the design to the implementation model. With this objective, the following tasks have been performed: first a new profile extends UML with new concepts from the web domain, next a new framework supports web application development by composing web components, and finally, a transformation model generates web applications from the UML extension proposed. The main contributions of this work is a cost and complexity reduction of web applications, and a high reusability degree achieved, since web components can be reused in different web applications.

Keywords: Model-driven architecture, web services, web applications, rich internet applications, distributed systems, AJAX.

1 Introduction

Model-Driven Development (MDD) is an approach to software development that goes beyond just modeling, and proposes automating as much of the code generation as possible. The fundamentals of this new trend are shown in Figure 1.

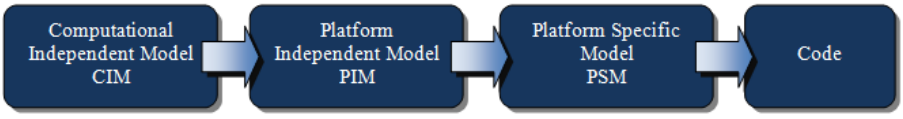


Fig. 1. MDE approach

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 511–522.
DOI: 10.1007/978-3-642-36981-0_47

A Computational Independent Model (CIM) is a model of the business rules of the system and has no computer implementation details specification. The Platform Independent Model (PIM) is a high level model tied to computational concepts but not to a computer platform specific implementation. Finally the Platform Specific Model (PSM) is a model of a specific computer platform implementation. Also a transformation tool preforms the translation between one step and another. Model-driven development is currently been applied in domain specific applications, and web applications is a good example of this new trend.

Web applications have evolved in the last few years, and with the appearance of Web 2.0 and the introduction of Asynchronous JavaScript and XML (AJAX) [1], a new breed of web applications with a high degree of interactivity has emerged. Rich Internet Applications (RIA) [2] have emerged under the umbrella of these new technologies, and have gained much attention and acceptance, since web browsers are the only execution environment required. This is the reason why, software vendors are actually adapting their applications to this new trend: Adobe has presented AIR[1] (Adobe Integrated Runtime) to extend web development to the desktop, Google has introduced Google Docs[2] as a new technology to work with online documents, and Microsoft has developed Windows Presentation Foundation[3] to provide support for web browser applications (WBAs). The interest exposed by all of these software vendors demonstrates the attention this technology is gaining.

However, while web applications become more and more complex, the development cost has been incremented, and this is because initial stages of software life cycle are not considered in the development process of this type of applications.

In order to overcome these problems, the Model Driven Architecture (MDA) [3] is proposed in this paper because simplifies modeling, design, implementation, and integration of applications by defining software mainly at the model level. The primary goals of MDA are portability, interoperability, and reusability through architectural separation of concerns [4], making product development more cost efficient by increasing automation in software development [5].

The objective of this paper is to propose a model-driven architecture to develop web applications with a high quality degree. This task has been achieved at three levels: first, a new profile includes web application concepts and extends UML [6], then a new framework supports web application development by composing web components, and finally, a transformation model generates web applications from the proposed UML profile.

This paper is organized as follows: first, section 2 presents the motivation of this work, an example of a scenario, and related works. Then, a UML extension is proposed in Section 3, which defines a new profile to design web applications. Next, the framework and performance evaluations are displayed in Section 4, while Section 5 describes the transformation model to generate web applications from a UML design. Finally, conclusion and future works are shown in Section 6.

---

[1] `http://www.adobe.com/es/products/air.html`
[2] `http://docs.google.com`
[3] `http://msdn.microsoft.com/es-es/library/ms754130.aspx`

2 Motivation and Related Works

Different AJAX frameworks have been proposed with the aim of providing new capacities for web applications. In this regard, Echo3[4] is a platform for building web applications that approaches the capabilities of rich clients, and presents a JavaScript client-side, which is executed directly by the client's web browser, while a Java server-side is stored in a Servlet Container. GWT[5] is a Google proposal for building and optimizing complex web applications, AJAX applications are written in Java and compiled into a highly optimized JavaScript code that runs across all browsers. Dojo Offline[6] is a cross-framework that enables web applications to work offline, where the information generated by the user is stored inside the web browser. In [7] it is presented a comparative of these approaches, and proposes a new architectural style (SPIAR) which captures the essence of AJAX applications.

Another interesting research studies how to provide connection to remote web services, in this sense [8] defines a framework based on asynchronous invocation of web services. In this work, two proxies (synchronous and asynchronous) support different types of connections at the client side, while the specialization of asynchronous calls is supported by handlers. Asynchronous callbacks in web services have also been proposed by Qian [9], in this work a web-service-callback proxy is defined. Also, focused in the correlation problem, asynchronous web services interactions have been analyzed in [10].

Trends in web application development are not limited to the specification of new frameworks, and the design level is also considered. A review of the principles of structured programming and the preferred characteristics of web applications is explained by [11]. This work presents a mapping of how the traditional guidelines may be applied to web applications development. In the field of MDA, there is a consensus about the benefits that this technology offers to web application development: a reduction of sensitivity to the inevitable changes that affect a software system [12], a reduction of cost and complexity [13], and an increase of abstraction [14]. An interesting analysis about the existing problems in the field of web engineering and how they can be solved by model-driven development approaches is presented by [15], which identifies the problems encountered in the development process of web applications such as their dependence on the HTTP protocol, compatibility issues due to the heterogeneity of web browsers, and the lack of performance because of the increase in the latency degree.

Different proposals extend web engineering methods for developing web applications; [16] presents a survey of existing web engineering methods to develop this type of applications. The Object-Oriented Hypermedia Design Model (OOHDM) uses abstraction and composition mechanisms in an object-oriented framework to, on one hand, allow a concise description of complex information items, and on the other hand, allow the specification of complex navigation patterns and interface

[4] http://echo.nextapp.com/site/echo3
[5] http://code.google.com/webtoolkit
[6] http://dojotoolkit.org/offline

transformations. The RUX-Model [17] is a representational model that offers a method for engineering the adaptation of legacy model-based Web 1.0 applications to Web 2.0 UI expectations. An extension of this model is proposed by [18] where a model-driven approach to web application development by combining the UML based Web Engineering (UWE) method for data and business logic modeling, with the RUX-Method for the user interface modeling of RIAs, is defined.

Focused on the presentation layer Abstract Data Views [19] are defined to specify the structure and behavior of user interfaces in web applications, and according to this model [20] proposes an interaction model, based on interaction patterns, to define new semantic to deal with the model-driven web application development.

WebML [21] is a conceptual model for specifying the content structure of web applications and the organization and presentation of contents in one or more hypertexts. WebML has been extended in different ways; [22] proposed an extension to design semantically web applications and [23] defined a UML Profile to model web applications using WebML.

A novel method for designing web applications that employs a UML notation is proposed by Dolog [24]. In this model, interaction spaces represent identifiable elements of a software system that support a specific part of the user interaction.

Another perspective for developing web applications is the data-driven model, where the state of the applications resides in a database, and users interact with this persistent state through web clients. According to this model, in [25] a framework for personalizing web applications is proposed, and suggested that users have the capability of customizing web applications to fit her requirements.

Currently, web applications are complex software composed by different types of components, and this is the reason why the development process of this type of software is so difficult. In this paper, a web application is considered as a composition of software components that can be activated at client or server side. A client component is downloaded and activated in the user's web browser, while server components are external resources, implemented in a variety of different languages.

3 Web Application Modeling

An extension of UML is proposed in the section, and a new profile (Figure 2) has been developed to specific concepts of the web application domain, adding new stereotypes and tagged values to facilitate the modeling process of this type of applications.

The main element of this profile is the *<<Web Resource>>* stereotype that represents a general element in the web. A software component is represented by the *<<URL Resource>>* stereotype. Software applications for the web are represented by the *<<Web application>>* stereotype, and according to previous considerations, they can be composed by other components, represented by the *<<Web Element>>* stereotype. Web applications, and also web elements, that can be bound to web services in order to invoke remote operations. Web components are represented by stereotypes with the same name. Remote services are referenced by the *<<Web Service>>*.

Requests sent by web resources to web services are represented by the <<*Web Service Request*>> stereotype, which can be classified according to the following communication patterns: synchronous (<<*Synchronous web service request*>>), and asynchronous (<<*Asynchronous web service request*>>). The asynchronous communication is extended with the <<*Asynchronous call back request*>> stereotype, which includes the operation that will be executed when the callback is activated. The <<*Asynchronous pull request*>> stereotype incorporates the *pool* property to hold responses, and the *push* and *pop* method to store and retrieve elements from the pool.

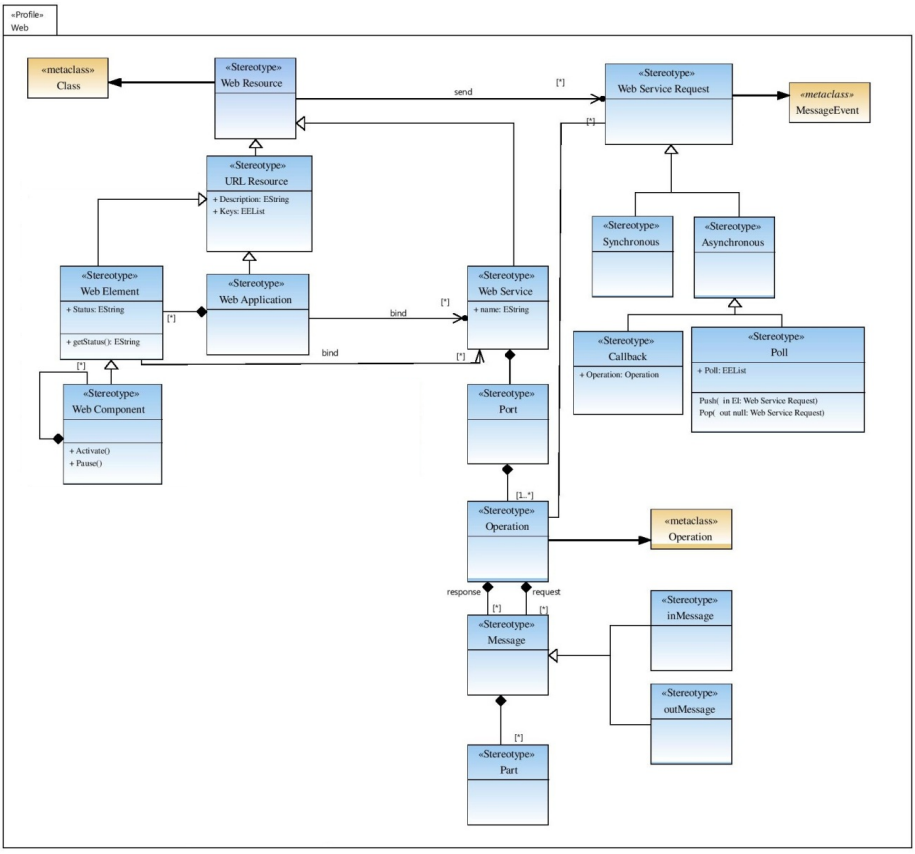


Fig. 2. UML profile

In order to test the proposed profile, the following scenario is suggested: a holiday web portal offers hotel vacancies in different locations around the world, and users can select the best alternative according to their preferences (price, date, location...). This web portal also provides a shopping cart to save client preferences and a payment service. Different tools are provided in order to facilitate the selection of a specific date and currency conversions. Figures (3a, 3b and 3c) show the web portal

model according to the proposed UML profile. As several web sources are required to be supported in this scenario, hotel information and vacancies (Hotel Source). Other important elements in this application are a shopping cart to save user's selections, a payment component used to connect with a bank and provide payment services. Finally, different tools are designed to allow users the selection of a specific date (calendar) and currency conversions (currency converter).

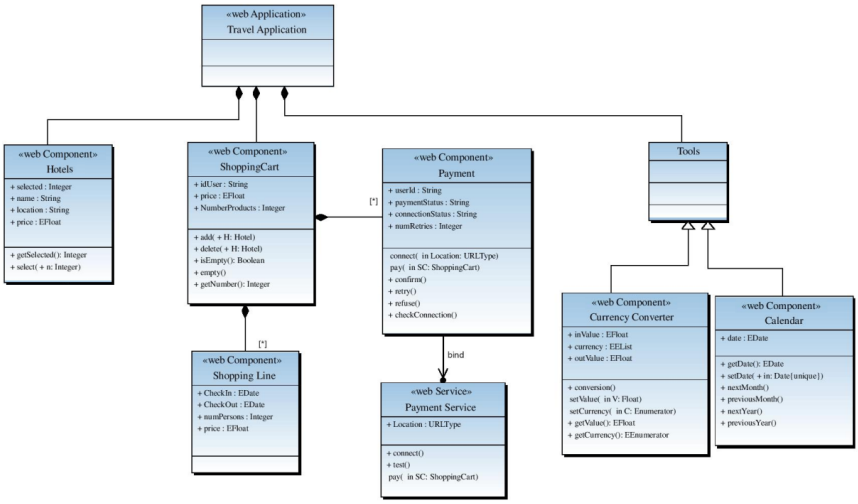


Fig. 3a. Holiday web portal design

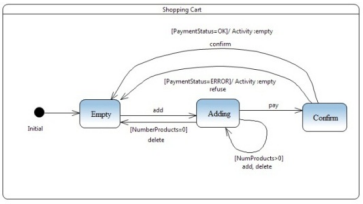


Fig. 3b. Shopping cart statechart

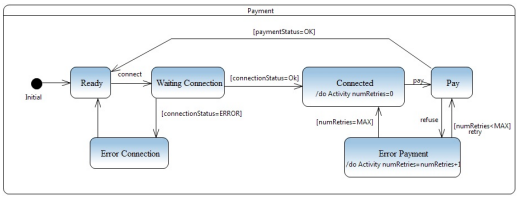


Fig. 3c. Payment statechart

4 The Framework

Web applications considered in this paper are characterized by the following properties:

- Downloaded: they are downloaded from the web and activated in a web browser environment.
- Distributed: web components and web services are distributed in the web.

In order to provide all this features, we propose the definition of a new framework (Figure 4) that supports the development of web applications by composing remote

resources, which can be spread over the Internet. As such, web applications are built in the client's web browser, while resources can be dynamically downloaded and activated only when they are required. This framework has been deeply explained in our previous work [26].

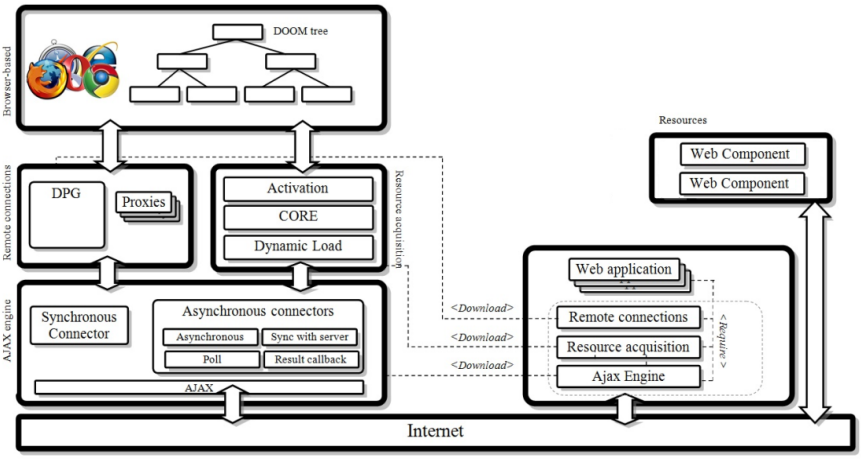


Fig. 4. The framework

4.1 Web Components

Currently web applications are composed by different type of components that can interact with each other or with the application. However, it is very difficult to develop a web application when interactions mechanisms have not been defined previously. Web components are proposed as a new type of web resource in the sense that they are downloaded and activated in a web application, and also, interface is specified in order to provide successful interaction mechanisms. A definition structure is defined in order to specify full information about the resource. The specification of a web component is divided into the following elements:

- Definition: the name of the web component, according to a namespace, a full description, and the set of elements required to activate the web component are specified in this section. XML is the language selected to specify this component, and the following tags have been defined:
    - <WEBCOMPONENT>: represents the beginning of a web component.
    - <NAME>: is the name of the web component.
    - <TYPE>: defines the type of the web component.
    - <DESCRIPTION>: includes the description of the web component
    - <AUTHOR>: contains the web component developer's name.
    - <REQUIREMENTS>: represents the beginning of the requirement structure.
    - <REQUIRES>: identifies a specific requirement.

- Interface: this element specifies the set of operations provided by the web component using IDL language.
- Functionality: the code of all the operations is stored in this element.

According to the example proposed in this paper, the shopping cart definition (Figure 5a) and interface (Figure 5b) are shown.

```
<WEBCOMPONENT>
  <NAME>ShoppingCart</NAME>
  <TYPE>LOCAL</TYPE>
  <DESCRIPTION> Save user's selections </DESCRIPTION>
  <AUTHOR> </AUTHOR>
  <REQUIREMENTS>
        <REQUIRES> shopping Line </REQUIRES>
        <REQUIRES> payment </REQUIRES>
  </REQUIREMENTS>
</WEBCOMPONENT>
```

Fig. 5a. Shopping cart definition

```
Interface ShoppingCart
{
   add(Hotel H);
   delete(Hotel H);
   boolean isEmpty();
   empty();
   Integer getNumber();
}
```

Fig. 5b. Shopping Cart Interface

5 Model Transformation

One of the keys of MDA is the capacity of defining transformations from higher-level models to platform specific models guided by a set of transformation rules. With this aim, a generation tool to transform models according to the proposed profile has been developed. In the following figure the structure of this tool is presented.

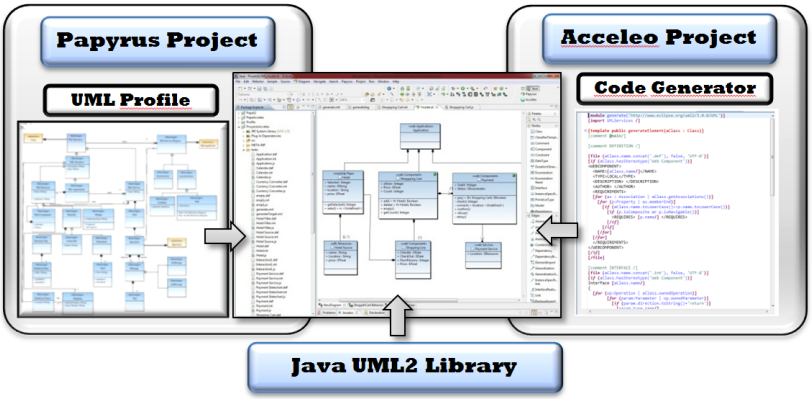


Fig. 6. Transformation tool

The algorithm to transform web components into final code is divided in three steps: definition, interface and functional code.

```
Input: a UML model (M)
Output: the definition of a web component.
1  foreach Cm ϵ Elements(M) do
2    if typeof(Cm)=WebComponent
3     generate(Header(Cm))
4     foreach As ϵ Associations(Cm) do
5      foreach End ϵ AssociationEnds(As) do
6        if (typeof(As)=Aggregation or
          typeof(AS)=Association)
7          generate("<REQUIRES>"+End+</REQUIRES>")
8      endfor
9     endfor
10   endif
11 endfor
```

Where *Cm* is an element from the model, *As* is an Association and *End* is a component of an Association. The algorithm searches each *Cm* of the model tagged as a web component, and looks for its associations. Every association or aggregation is transformed into a requirement.

The following section of the algorithm generates the interface according to IDL language, where Op is a method and *Param* is a parameter of a method. In this case, for each Cm all its methods and parameters are identified and transformed into IDL.

```
Input: a UML model (M)
Output: the interface of a web component.
1  foreach Cm ϵ Elements(M) do
2    if typeof(Cm)=WebComponent
3     generate("Interface"+Cm+"{")
4     foreach Op ϵ Operations(Cm) do
5       foreach Param ϵ Params(Op) do
6         if (typeof(Param)=return)
7           generate(typeof(Param))
8       endfor
9       generate(Param+"(")
10      foreach Param ϵ Params(Op) do
11        if (typeof(Param)<>return)
12          generate(typeof(Param)+Param)
13      endfor
14      generate(")}")
15   endif
16 endfor
```

Finally, the following section of the algorithm generates the javascript code of the web component where *At* is an attribute. Each *Cm* is transformed into a JavaScript class, and all its methods and attributes (single or multiple) are identified

```
Input: a UML model (M).
Output: the functional code of a web component.
1    foreach Cm ϵ Elements(M) do
2       if typeof(Cm)=WebComponent
3          generate("function"+Cm +"{")
4          foreach At ϵ Attribute(Cm) do
5             generate("this."+At)
6          endfor
7          foreach As ϵ Associations(Wc) do
8             foreach End ϵ AssociationEnds(As) do
9                if (typeof(As)=Aggregation)
10                  if (Ocurrences(End,As)>1)
11                     generate("this."+End+i+"= new Array()")
12                  else
13                     generate("this."+End+i)
14                     generate("=new "+End+"()")
15                  endif
16               endif
17            endfor
18            generate("}")
19            for (each Operation(Wc) Op)
20               generate (Op+"=function(")
21            endfor
22            for (each Param(Op) Param)
23               if (typeof(Param)<>return)
24                  generate(Param)
25         endfor
26         generate("{ }")
27      endif
28   endfor
```

The generated code for the web component *ShoppingCart,* modeled in the example exposed in section 3, is shown in Figure 5a and 5b.

6 Conclusions and Future Works

In this paper a model-driven architecture for developing web applications is proposed and, with this aim, different stages of software development process are considered. At the design level, a new profile extends UML with new concepts from the web application domain, and it is also proposed a new framework to support the dynamic download and activations of web components. Finally, a transformation model generates web applications from the UML extension proposed. The main benefits provided by this proposal are:

- A cost and complexity reduction of web applications through the incorporation of the model-driven approach.
- As web elements are deployed independently, the can be reused in different web applications, and as such the reusability degree is increased.

A prototype of a holiday web portal has been presented along this paper, the scenario of the example has been introduced, the design, according to the UML extension, has been modeled, and finally an example of a generated web component has been exposed.

Future works will try to incorporate new types of web entities such as applets or dynamic web contents generated by server languages into the development process of web applications. Moreover, historical information will be included in the prefetching algorithm in order to predict more efficiently the web elements required to be downloaded in advance.

References

1. Garrett, J.J.: Ajax: A new approach to web applications. Adaptive path (2005), http://www.adaptivepath.com/publications/essays/archives/000385.php (last accessed November 2012)
2. Paulson, L.D.: Building rich web applications with Ajax. IEEE Computer 38(10), 14–17 (2005)
3. http://www.omg.org/mda/ (last accessed November 2012)
4. Estefan, J.: Survey of Model-Based Systems Engineering (MBSE) methodologies, INCOSE-TD-2007-003-01 (2008)
5. Teppola, S., et al.: Challenges in Deployment of Model Driven Development. In: Fourth International Conference on Software Engineering Advances, ICSEA 2009, pp. 15–20, 20–25 (2009), doi:10.1109/ICSEA.2009.11
6. http://www.uml.org/ (last accessed November 2012)
7. Mesbah, A., Van Deursen, A.: A component-and push-based architectural style for AJAX applications. Journal of Systems and Software 81(12), 2194–2209 (2008), doi:10.1016/j.jss.2008.04.005
8. Zdun, U., Völter, M., Kircher, M.: Design and Implementation of an Asynchronous Invocation Framework for Web Services. In: Jeckle, M. (LJ) Zhang, L.-J. (eds.) ICWS-Europe 2003. LNCS, vol. 2853, pp. 64–78. Springer, Heidelberg (2003)

9. Qian, K., et al.: Asynchronous Callback in Web Services. In: Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, pp. 375–380 (2006), doi:10.1109/SNPD-SAWN.2006.23
10. Brambilla, M., et al.: Managing asynchronous Web services interactions. In: Proceedings of the IEEE International Conference on Web Services, ICWS 2004, p. 80 (2004), doi:10.1109/ICWS.2004.73
11. Marquis, G.: Application of traditional system design techniques to web site design. Information and Software Technology 44(9), 507–512 (2002), doi:10.1016/S0950-5849(02)00050-2
12. Atkinson, C., Kühne, T.: The role of metamodeling in MDA. In: International Workshop in Software Model Engineering (2002)
13. Mukerji, J., Miller, J.: MDA. Guide version 1.0.1, OMG document number: omg/2003-06-01, http://www.omg.org/cgi-bin/doc?omg/2003-06-01 (last accessed November 2012)
14. Booch, G., et al.: An MDA manifesto. MDA Journal, http://www.bptrends.com/publicationfiles/05-04COLIBMManifest20-Frankel-3.pdf (last accessed November 2012)
15. Gitzel, R., et al.: Using established Web Engineering knowledge in model-driven approaches. Science of Computer Programming 66(2), 105–124 (2007), doi:10.1016/j.scico.2006.09.001
16. Fraternali, P., et al.: Engineering rich internet applications with a model-driven approach. Journal ACM Transactions on the Web (TWEB) 4(2) (2010), doi:10.1145/1734200.1734204
17. Linaje, M., et al.: Engineering Rich Internet Application User Interfaces over Legacy Web Models. Journal IEEE Internet Computing Archive 11(6), 53–59 (2007), doi:10.1109/MIC.2007.123
18. Preciado, J.C., et al.: Designing Rich Internet Applications Combining UWE and RUX-Method. In: Eighth International Conference on Web Engineering, pp. 148–154 (2008), doi:10.1109/ICWE.2008.26
19. Urbieta, M., et al.: Designing the Interface of Rich Internet Applications. In: Fifth Latin American Web Congress, pp. 144–153 (2007), doi:10.1109/LA-Web.2007.14
20. Valverde, F., Pastor, O.: Applying Interaction Patterns: Towards a Model-Driven Approach for Rich Internet Applications Development. In: 7th International Workshop on Web-Oriented Software Technologies, pp. 13–18 (2008)
21. Ceri, S., et al.: Designing Data-Intensive Web Applications. Morgan Kaufmann Publishers Inc. (2002)
22. Brambilla, M., et al.: Model-driven design and development of semantic Web service applications. ACM Transactions on Internet Technology (TOIT) 8(1), article No. 3 (2007), doi:10.1145/1294148.1294151
23. Moreno, N., et al.: A UML 2.0 profile for WebML modeling. In: ICWE 2006 Workshop Proceedings of the Sixth International Conference on Web Engineering (2006), doi:10.1145/1149993.1149998
24. Dolog, P., Stage, J.: Designing interaction spaces for rich internet applications with UML. In: 7th International Conference on Web Engineering, pp. 358–363 (2007)
25. Reischuk, R., et al.: SAFE Extensibility for Data-Driven Web Applications. In: 21st International Conference on World Wide Web, pp. 799–808 (2012), doi:10.1145/2187836.2187944
26. Herrero, J.L., et al.: Web services and web components. In: Seventh International Conference on Next Generation Web Services Practices, pp. 164–170 (2011)

Towards a Conceptual Framework for Early Warning Information Systems (EWIS) for Crisis Preparedness

Mohamed Saad[1], Sherif Mazen[1], Ehab Ezzat[1], and Hegazy Zaher[2]

[1] Faculty of Computers and Information, Cairo University, Giza, Egypt
mmsaad@hotmail.com, smazen@mailer.eun.org, admin@drehab.net
[2] Faculty of Management and Information Systems, French University, Egypt
hgsabry@yahoo.com

Abstract. This paper highlights the need of many organizations nowadays for early warning information systems (EWIS) that can predict the future and help prevent crises or reduce their negative effects. These EWIS should be based on a reliable and consistent framework. The frameworks currently available are mostly deterministic, simplified or inconsistent in application and assumption; thus making them unreliable and impractical. The goal of this paper is twofold. Firstly, it provides guidelines for system analysts, designers, engineers and experts seeking to deal with crisis or disaster information systems. Secondly, it aims to present a novel framework for EWIS that can be adapted to the dynamic needs of the field of crisis management, and that can also be used efficiently in crisis preparedness. Finally, the paper will describe a case study in the law enforcement sector as a proof-of-concept for the conceptual framework; to demonstrate both the theoretical and practical approaches.

Keywords: Early Warning, Framework, Crises, Forecasting, Model, Indicators.

1 Introduction

Although information systems are used in order to collect and share information at a time of crisis, such systems are not always effective and do not always bring about the intended outcomes. While *Assilzadeha* believes that the development and implementation of application software for early warning especially for disaster data and information management is very important [1], *Glantz* claims that there is no perfect Early Warning System (EWS), except on paper, or in governmental plans, or in a PowerPoint presentation and that most of the current systems are not as effective as they should be [2]. Along the same lines, *Sanada* also agrees that the information systems which are used in order to collect and share information at a time of disaster are not always effective [3]. Furthermore, *Harff* argues that at present, early warnings are rarely "early," seldom accurate, and moreover lack the capacity to distinguish among different kinds of crises [4]. In this context, after analyzing the current problems in existing early warning information systems used in crisis or disaster preparedness, we have found that in many cases these systems tend to be fairly narrow in scope and do not have an adequate or clear framework for collecting,

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 523–534.
DOI: 10.1007/978-3-642-36981-0_48

classifying, processing and producing accurate forecasting information. The limitations of existing EWIS suggest the need for a more comprehensive framework. This is the goal we aim to achieve from this paper. Furthermore, other motivations for this paper are:

1. The scarcity of specialized EWIS researches and early warning applications used in crisis preparedness.
2. Most of the previous studies did not address the key steps used in building EWIS.
3. There is no agreement on the ideal structure or functions of EWIS.

The proposed framework provides a guideline for any organization or sector in the country (such as: health, security, education, .etc) that needs to have an EWIS for crisis preparedness. Moreover, this framework will support any information system in producing more effective and accurate predictions of the future. In addition, it will provide decision makers with a reliable and manageable amount of warning information for taking preventive actions.

2 Early Warning Information Systems (EWIS)

The idea of early warning emerged in the fifties of the past century, and was used for the first time in military domains to predict risks and potential attacks before they occur. Until the early eighties; the concept of early warning had not evolved noticeably due to a number of reasons; such as the difficulty of creating its applications and its high cost. However, the concept has been rediscovered again after a series of crises and disasters had taken place in the world and after witnessing their major impact on lives and property [5]. The expression 'Early Warning' is used in many fields to mean the provision of information on an emerging dangerous circumstance where that information can enable action in advance to reduce the risks involved [6]. A universally accepted definition of an EWS does not yet exist and most probably never will [7]. There are many definitions of an EWS that are used to guide the actions of individuals, groups, and governments. The formal UN definition is as follows: "The provision of timely and effective information, through identifying institutions, that allows individuals exposed to a hazard to take action to avoid or reduce their risk and prepare for effective response" [8]. An EWS can also be defined as "a social process for generating maximally accurate information about possible future harm and for ensuring that this information reaches the people threatened by this harm, as well as others disposed to protect them from the harm" [2][9]. An *'Early Warning Information System (EWIS)' (*see Fig. 1.*)* can be understood as a set of institutional and technical solutions designed and

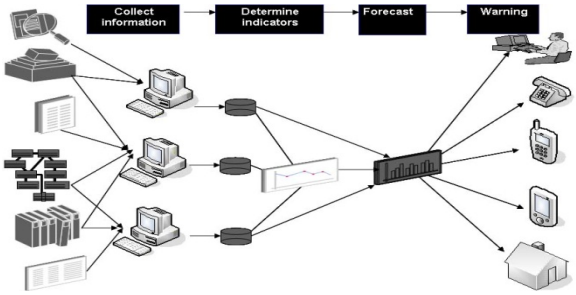


Fig. 1. EWIS Architecture

implemented in a coherent way to make available, to a wide range of users and more particularly to decision makers, information useful to carry out vulnerability analyses, to evaluate and manage the risk of a hazard that can become a disaster, and to manage disasters from prevention to recovery and rehabilitation [10][11][12]. The objective of EWIS is to generate accurate information to empower individuals and communities threatened by hazards to act in sufficient time and in an appropriate manner so as to reduce the possibility of personal injury, loss of life and damage to property or the environment. We can use the term EWIS for any information system that collects, shares, analyzes data, produces future predictions about potential crises and gives recommendations or warnings for those involved. The effectiveness of any EWIS largely depends on the transformation of the event[1] recognition into the report of warning to the population or people at risk [13][14], also depends on how it supports the organization to reach its objectives as a whole [15]. The best known EWIS are the HEWS- Humanitarian Early Warning System- used by the Department of Humanitarian Affairs in the United Nations, and the GIEWS -The Global Information and Early Warning System- used by the Food and Agriculture Organization of the United Nations [16].

3 EWIS Framework

Various writers have identified what they considered should be the components of an EWIS framework. For example, *Verstegen* suggested that the EWIS framework should have five components: selection of indicators; communication of warnings; reception of warnings; early warning education; generation and maintenance of awareness [16]. However, this framework did not specify the methods or steps of data collection. Moreover, it did not explain how to measure the precursors, evaluate the event or specify the forecasting models. Along the same lines, *Lundin* suggests that an EWIS is responsible for issuing forecasts, warnings, and responses [17]. Yet, this framework does not clarify how the data is collected and analyzed; explain how to prepare the future forecasts or how to select the most suitable model for forecasting. Obviously, there is no agreement on the ideal structure or function of an early warning system [18]. This means that the structure and functions of EWIS may vary from one organization to another and from one field to another. Hence, we present in this paper a comprehensive framework with general characteristics that can be applied in different fields and institutions. The suggested EWIS framework *(*see Fig. 2.*)* combines the positive aspects of previous frameworks and represents the core functions of EWIS across four Models. The first model (Societal Detection of Events sub-model) includes

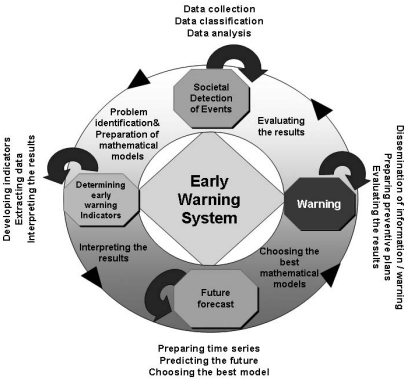


Fig. 2. EWIS Framework

[1] Event is defined as an observable occurrence, phenomenon or an extraordinary occurrence.

functions that capture and analyze the event/crisis information; the second model (Determining Early Warning Indicators sub-model) determines the set of mathematical indicators that should be measured frequently; the third model (Future Forecasting sub-model) provides future forecasts depending on the data calculated from the previous model, and finally the fourth model (Issuance of Warnings sub-model) is concerned with sending warnings (alerts) to users. Each of the four models is explained below in detail.

3.1 Work Methodology in Societal Detection of Events (Sub-model 1)

The first sub-model is divided into a series of processes (see Fig. 3.) including:

1. Designing a data entry form to collect data from different sources.
2. Data collection.
3. Filling out the prepared form with the data collected in the previous step.
4. Checking any data missing in the form.
5. Verifying and validating the data.
6. Classifying and analyzing the data.

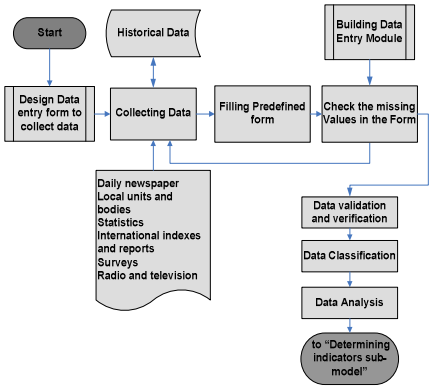


Fig. 3. Societal Detection of Events – Sub-model (1)

3.2 Work Methodology in Determining Early Warning Indicators (Sub-model 2)

An indicator is defined as a number or ratio derived from a series of observed facts; it can reveal relative changes as a function of time [19]. The value of an indicator in EWIS is very important to evaluate any event/crisis. This sub-model is divided into a series of processes (see Fig. 4.) such as:

1. Determining the indicator.
2. Defining and describing the indicator.
3. Determining the variables involved in calculating the indicator.
4. Analyzing the links between the variables and determining the measuring unit.
5. Revision and validation of information.
6. Testing the indicator with real data.

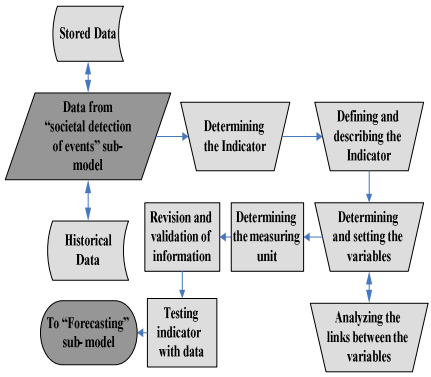


Fig. 4. Determining Early Warning Indicators – Sub-model (2)

3.3 Work Methodology in Future Forecasting (Sub-model 3)

The third sub-model is divided into a series of important processes (see Fig. 5.) such as:

1. Creating a database for time series data.
2. Selecting specific data for each indicator (year & value).
3. Checking the validity of data.
4. Forecasting models specification.
5. Creating mathematical equations for different forecasting models.
6. Forecast generation.
7. Obtaining the results.
8. Validating the results.
9. Displaying different forecasting models with different values.
10. Choosing the best model.

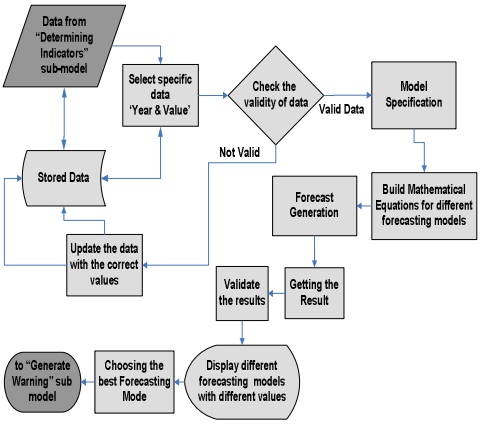


Fig. 5. Future Forecasting – Sub-model (3)

3.4 Work Methodology in the Issuance of Warnings (Sub-model 4)

This sub-model is divided into a series of important processes (see Fig. 6.) such as:

1. Calculating the probability of a crisis/event.
2. Calculating the length of time remaining to the emergence of the crisis/event.
3. Applying the early warning statistical equations.
4. Applying the color coded scale.
5. Checking the color of each variable.
6. Sending warning alerts to those concerned (the public, decision makers, government organizations, etc) in a time frame that varies according to the severity of the crisis expressed by the color of each variable (red, orange, yellow, green and blue).

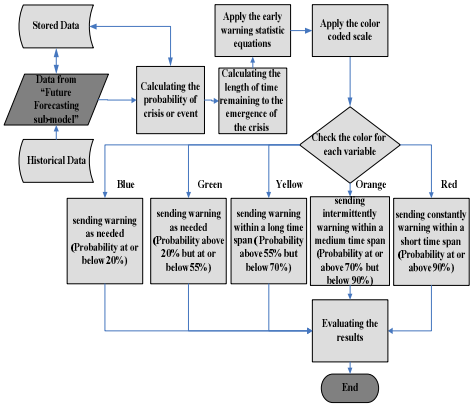


Fig. 6. Issuance of Warnings – Sub-model (4)

4 Case Study

In order to demonstrate some of the key concepts introduced in the framework described above, we have implemented a proof-of-concept prototype purely in

software. The key objectives for the proof of concept are to demonstrate the early warning theoretical concept, and to show how the theoretical concept can be implemented practically against data from the law enforcement sector. The next few pages will explain the case study in more detail.

4.1 Societal Detection of Event

1. The event data is collected from different sources and is regularly inputted into the EWIS.
2. The number of occurrences is calculated for each event in a specific time frame.
3. The EWIS selected the most frequent event; which is "widespread drug abuse among the youth" (see Table 1.), because this event has the highest number of occurrences. This event was detected from multiple sources[2] through the period of time from January 2006 to December 2010 (see Table 2.).
4. Data is collected from different data sources and entered into the EWIS.
5. After analyzing the data, the system found that this event is increasing on an annual basis (see Table 2).

Table 1. Event Occurrences (2010)

Event /Phenomena	No. of Occurrences
Drug abuse (youth)	64
Spinsterhood	24
The collapse of buildings	22
Hooliganism	21
Child molestation	20
Luxury consumption	19
Unknown parentage	18
Train accidents	18
Drug trafficking	14
Trafficking in Persons	12

Table 2. Event (Drug Abuse among youth) Occurrences (2006-2010)

Year	2006	2007	2008	2009	2010
No of occurrences	36	40	43	52	64

4.2 Determining the Event Indicators

Determining the Indicators that Best Describe the Event

1. Total number of drug users in the country.
2. Total number of drug cases.
3. The percentage of people arrested in drug cases to the total population.
4. The percentage of local people arrested to the total of arrests (all nationalities).

[2] Sources such as: News Media, Newspaper, Statistics, Reports, Databases, Radio & T.V, Blogs, Internet, Surveys, Studies, Local Communities, Local Government, etc.

Defining the Indicators

1. Total number of youth population addicted to drugs in the country.
2. Total number of drug cases that is registered in police statistics.
3. The percentage of youth arrested in drug cases to the total youth population.
4. The percentage of local people arrested to the total of arrests of other nationalities.

Determining the Variables Involved in Calculating the Indicator

1. Total number of drug users in the country = (total number of drug users in the youth population).
2. Total number of drug cases = (total number of youth drug cases).
3. The percentage of people arrested in drug cases to the total population= (total number of people arrested/ total population) * 100.
4. The percentage of local youth population arrested to the total of arrests of youth from all nationalities= (total number of local youth population arrested in the drug abuse cases/ total youth arrested in the drug abuse cases) * 100.

4.3 Designing a Forecasting Model for Each Indicator

Creating a Time Series for Each Indicator. Table (3) shows the time series data for all indicators and Figure (7) shows the same table implemented in the EWIS.

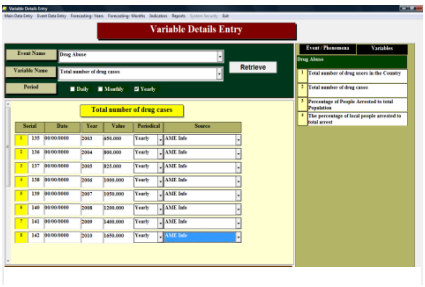


Fig. 7. Create time Series Data

Specifying the Mathematical Model Specification. The EWIS will use three forecasting models to be applied (Linear, Quadratic and Cubic Models).

Building Mathematical Equations for Different Forecasting Models

1. Linear equation (y) = mx +b (where m and b designate constants, x is a slope)
2. Quadratic equation = $ax^2 + bx + c$ (where a, b, c are constants with $a \neq 0$)
3. Cubic equation = $ax^3 + bx^2 + cx + d$ (where a, b, c, d are constants with $a \neq 0$).
4. MAPE (Mean Absolute Percentage Error) =

$$\frac{\sum \left| \frac{\text{Actual Value} - \text{Fitted Value}}{\text{Actual Value}} \right|}{n} \times 100$$ (Where n = total number of actual values)

R^2 : The coefficient of determination[3].

[3] The coefficient of determination: is used in the context of statistical models whose main purpose is the prediction of future outcomes on the basis of other related information. R^2 is most often seen as a number between 0 and 1.0. An R^2 near 1.0 indicates that a regression line fits the data well, while an R^2 closer to 0 indicates a regression line does not fit the data very well.

Table 3. Indicators Time Series Data (2003-2010)

Year	Indicator A	Indicator B	Indicator C	Indicator D
2003	700	650	4%	40%
2004	800	800	5%	45%
2005	850	825	5.5%	48%
2006	1050	1000	7%	49%
2007	1110	1050	7.2%	55%
2008	1206	1200	7.7%	60%
2009	1708	1400	9%	62%
2010	2000	1650	10.6%	69%

Creating Forecasts for Each Indicator (Forecast Generation). Calculating forecasting data for each indicator as follows:

Indicator (A). Table (4) shows the correlation of variables for each model, the EWIS will automatically choose the cubic model because its correlation value is the largest among other models (=0.981). In addition the value of (MAPE) is the smallest (=3.5), hence the forecasting data for the cubic model will be more accurate and Table (5) shows the comparison between the various forecasting models for indicator (A).

Table 4. Mathematical Measures for Indicator (A)

Model	Fitted Trend Equation	MAPE	Correlation
Linear	$y = 386.857 + 175.81\ x$	11	$R^2 = 0.891$
Quadratic	$y = 783.286 - 62.048\ x + 26.429\ x^2$	5.2	$R^2 = 0.971$
Cubic	$y = 547.286 + 183.488\ x - 37.935\ x^2 + 4.768\ x^3$	3.5	$R^2 = 0.981$

Table 5. Forecasting Models for Indicator (A) for the years (2011-2014)

Year	X	Actual Data (Indicator A)	Linear Model	Quadratic Model	Cubic Model
2003	1	700	562.667	747.67	697.607
2004	2	800	738.477	764.9	800.666
2005	3	850	914.287	835	885.1
2006	4	1050	1090	958	979
2007	5	1110	1266	1134	1112.4
2008	6	1206	1442	1362.43	1312
2009	7	1708	1618	1643.95	1608
2010	8	2000	1793	1978.33	2029
2011	9		1969	2366	2602
2012	10		2145	2806	3357
2013	11		2321	3299	4322
2014	12		2497	3844	5526

Indicator (B). Table (6) shows the correlation of variables for each model, the EWIS will automatically choose the cubic model because its correlation value is the largest among other models (=0.994) as well as the value of (MAPE) is the smallest (=2), so the forecasting data for the cubic model will be more accurate, Table (7) shows the comparison between the forecasting data for indicator (B) and Figure (8) shows an example of an EWIS designed to verify the framework.

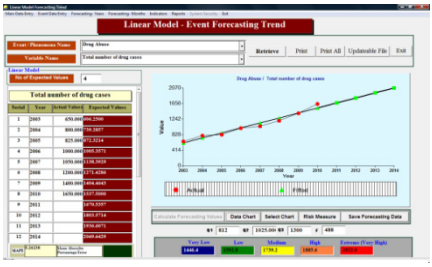


Fig. 8. EWIS (Designed by the Researcher) - Linear Forecasting Trend Example

Table 6. Mathematical Measures for Indicator (B)

Model	Fitted Trend Equation	MAPE	Correlation
Linear	y=473.214+133.036x	5.2	$R^2 = 0.957$
Quadratic	$y = 649.554+27.232\ x + 11.756\ x^2$	3.5	$R^2 = 0.987$
Cubic	$y = 496.429 + 186.544\ x – 30.005\ x^2 + 3.093\ x^3$	2	$R^2 = 0.994$

Table 7. Forecasting Models for Indicator (B) for the years (2011-2014)

Year	X	Actual Data (Indicator B)	Linear	Quadratic	Cubic
2003	1	650	606	689	656.1
2004	2	800	739	751	774.2
2005	3	825	872	837	869.5
2006	4	1000	1005	947	960.5
2007	5	1050	1138	1080	1065.6
2008	6	1200	1271	1236	1203.6
2009	7	1400	1404	1416	1392.9
2010	8	1650	1538	1620	1652.1
2011	9		1671	1847	1999.7
2012	10		1804	2097	2454.4
2013	11		1937	2372	3034.6
2014	12		2070	2669	3658.9

Indicator (C) and Indicator (D). The system will carry out the same processes as indicators (A & B); the EWIS will automatically choose the cubic model because its correlation value is the largest among other models. In addition, the value of (MAPE) is the smallest, so the forecasting data for the cubic model will be more accurate.

Charting the Forecasting Data for Each Indicator. The EWIS will chart the forecasting data for each indicator (see Fig. 9.a.)

Choosing the Best Model. After analyzing and validating the data, the EWIS will choose the best forecasting model (see Fig 9.b.) depending on the following criteria:

1. The Lowest Mean Absolute Percentage Error (MAPE).
2. The Highest correlation value between variables.

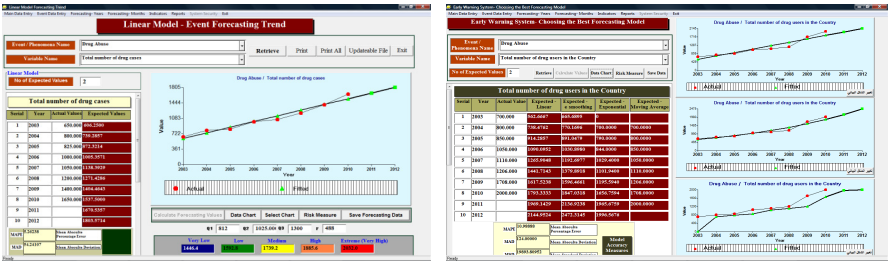


Fig. 9. (a) Linear Trend Forecasting for Indicator (A); (b) Choosing the best model

4.4 Sending Warning Messages

Calculating the Probability of a Crisis

1. Calculating the probability of each indicator by setting a range for each one. The probabilities will have values between 0.1 and 1 as shown in (Table 8).
2. Determining the level of danger (see Table 9., Fig. 10.), the level of danger equation will be :

   [(Probability of indicator A + Probability of indicator B + Probability of indicator C + Probability of indicator D) / total Number of indicators] * 100.
3. Converting the result into the following description:
   1. ***Dangerous*** (Red Color) = Probability at or above 90%.
   2. ***High*** (Orange Color) **=** Probability at or above 70% but below 90%.
   3. ***Medium*** (Yellow Color) = Probability above 55% but below 70%.
   4. ***Low*** (Green Color) = Probability above 20% but at or below 55%.
   5. ***Nil*** (Blue Color) **=** Probability at or below 20%.

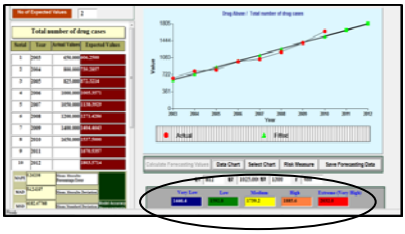


Fig. 10. Level of Danger

Calculating the Length of Time Remaining to the Emergence of the Crisis. The system found that in the year 2014 the detected event will reach a dangerous level.

Table 8. Probability Table

Indicator A		Indicator B		Indicator C		Indicator D	
Range	Probab -ility	Range	Probabi -lity	range	Prob- ability	range	Proba- bility
<500	0.1	<400	0.1	< 1	0.1	<10%	0.1
501-1000	0.2	401-800	0.2	1- 3	0.2	10-19	0.2
1001-1500	0.3	801-1200	0.3	>3-6	0.3	20-29	0.3
1501-2000	0.4	1201-1600	0.4	>6-9	0.4	30-39	0.4
2001-2500	0.5	1601-2000	0.5	>9-12	0.5	40-49	0.5
2501-3000	0.6	2001-2400	0.6	>12-15	0.6	50-59	0.6
3001-3500	0.7	2401-2800	0.7	>15-18	0.7	60-69	0.7
3501-4000	0.8	2801-3200	0.8	>18-21	0.8	70-79	0.8
4001-4500	0.9	3201-3600	0.9	>21-24	0.9	80-89	0.9
> 4500	1	> 3600	1	>24	1	90-100	1

Charting the Data from the Previous Table to Determine the Level of Danger. The EWIS will start charting the data for each indicator *(*see Fig. 10*.)* to determine the level of danger (there are five levels of danger: dangerous, high, medium, low, and nil), the probability of a crisis and the time frame remaining its emergence.

Matching the Level of Danger to the Color Coded Scale. The EWIS will match the level of danger (see Table 8. and Table 9.) to the color coded scale.

Generating Alerts. During this process, the system will automatically generate alerts based on the forecasting data calculated through the system. Once alerts are generated, they can be distributed through different channels to many parties, such as decision makers, governmental sectors, and the people exposed to hazards.

Table 9. Final Results

Year	Indicator A	Probability	Indicator B	Probability	Indicator C	Probability	Indicator D	Probability	Level of Danger (%)	Description
2003	700	0.2	650	0.2	4%	0.3	40%	0.5	30	Low
2004	800	0.2	800	0.2	5%	0.3	45%	0.5	30	Low
2005	850	0.2	825	0.3	5.5%	0.3	48%	0.5	32.5	Low
2006	1050	0.3	1000	0.3	7%	0.4	49%	0.5	37.5	Low
2007	1110	0.3	1050	0.3	7.2%	0.4	55%	0.6	40	Low
2008	1206	0.3	1200	0.3	7.7%	0.4	60%	0.7	42.5	Low
2009	1708	0.4	1400	0.4	9%	0.4	62%	0.7	47.5	Low
2010	2000	0.4	1650	0.5	10.6%	0.5	69%	0.7	52.5	Low
2011	2602	0.6	1999.7	0.5	12.9%	0.6	75.2%	0.8	62.5	Medium
2012	3357	0.7	2454.4	0.7	15.9%	0.7	83%	0.8	72.5	High
2013	4322	0.9	3034.6	0.8	19.8%	0.8	92.9%	1	87.5	High
2014	5526	1	3658.9	1	24.8%	1	104.4%	1	100	Dangerous

5 Conclusion

In this paper, we have proposed a conceptual framework to build any EWIS. We have also presented the detailed structure of this framework and how it will be implemented. The framework was tested through a case study to proof the concept. A list of issues related to the EWIS was also presented. Based on our literature review, we have concluded that hitherto there had been no comprehensive framework worldwide for building an efficient EWIS. Hence, an information system built using this new framework will be effective and beneficial due to a number of reasons. Firstly, it will provide information on the past, present and future and on relevant events inside and outside any organization. Secondly, it will be an integrated system for gathering relevant data, converting it to warning information and supplying the same to concerned executives and decision makers. Thirdly, it will reduce the time needed to build a sophisticated EWIS from scratch. Fourthly, it will select the best model for forecasting; which leads up to accurate results. And finally, it will strengthen the ability of the organization to prevent disasters and crises before they occur.

References

[1] Assilzadeha, H., Mansor, S.B.: Natural Disaster Data and Information Management System. In: International Society for Photogrammetry and Remote Sensing, Commission VII, Working Group 5, Methods for Change Detection and Process Modeling, 2008-2012 (2008)

[2] Glantz, M.: Early Warning Systems: Dos and Don'ts. In: The Early Warning Systems Workshop, Shanghai, China (2004)

[3] Sanada, A., Kusakabe, T., Tamura, K.: Establishing a Disaster Information System to Practically Support Disaster Countermeasures: Conducting and Reviewing Demonstration Experiment. In: 38th Joint Meeting Panel on Wind and Seismic Effects. Public Works Research Institute (2006)

[4] Harff, B.: Early Warning of Humanitarian Crises: Sequential Models and the Role of Accelerators in Preventive Measures: Building Risk Assessment and Crisis Early Warning Systems. In: Davies, J.L., Gurr, T.R. (eds.), p. 71. Rowman & Littlefield Publishers, Lanham (1998)

[5] Saad Eldin, M.: Early Warning Systems and the Positive Effects in Supporting the Capacity Development. Dubai police General Headquarters, Decision Making Support Centre, Dubai (2011)

[6] Basher, R.: Global Early Warning Systems for Natural Hazards: Systematic and People-centered. In: The Platform for the Promotion of Early Warning (PPEW), Bonn, Germany (2006)

[7] Sivakuma, M.: Early Warning Systems for Drought: Past and Present, World Meteorological Organization, Climate and Water Department, Geneva2, Switzerland (2009)

[8] ISDR (UN International Strategy for Disaster Reduction), Terminology: Basic Terms of Disaster Risk Reduction (2003), `http://www.unisdr.org/` (accessed January 15, 2012)

[9] Davies, J., Gurr, T.: Preventive Measures Building Risk Assessment and Crisis Early Warning System. Rowman & littlefield Publishers, Unites states of America (1998)

[10] Scott, J.: Input for Second International Conference on Early Warning (EWC II) Planning, Bonn, Germany, January 23 (2003)

[11] Intelligent Systems for Crisis and Disaster Management -ISCRAM. 5th International Conference on Information Systems for Crisis Response and Management Washington, DC, USA, May 4-7 (2008)

[12] Intergovernmental Authority on Development (2002). Disaster Risk Management Program, Strengthening of EWISs and Vulnerability Analysis (June 2002)

[13] Martin, B., Wolfgang, R., Eva, O., Stephan, M., et al.: Development of suitable information systems for early warning systems. FZI research center, Germany (2008)

[14] Wikipedia. (n.d.). Definition of Event, from Wikipedia the free encyclopedia: `http://en.wikipedia.org/wiki/Event` (retrieved September 15, 2012)

[15] Malik, K.: Information Systems Effectiveness: An Integrated Approach. Paper presented at the Change Management and the New Industrial Revolution, International Engineering Management Conference IEMC 2001, NY, USA, October 9 (2001)

[16] Verstegen, S.: Conflict Prognostication: Toward a Tentative Framework for Conflict Assessment. Clingendael Institute, Netherland (1999)

[17] Lundin, H.: Crisis and Conflict Prevention with an Internet based Early Warning System. Royal Institute of Technology (KTH), Sweden (2008)

[18] Shrestha, B.K.: Planning Food Security in Nepal. In: ADB-FAO Workshop on Agricultural Statistics and Agriculture Census, Bangkok, Thailand, November 2-5 (2009)

[19] Saad Eldin, M.: International Indicators and Their Role in Supporting Decision Making, Dubai police General Headquarters, Decision Making Support Center, Dubai, United Arab Emirates (2009)

Aggregation Operators and Interval-Valued Fuzzy Numbers in Decision Making

József Mezei and Robin Wikström

IAMSR, Åbo Akademi University
Joukahainengatan 3-5A, Turku, Finland
{jmezei,rowikstr}@abo.fi

Abstract. Aggregation operators play a fundamental role in decision making, especially when there are numerous (conflicting) criteria present. In case of uncertain data, an important task is to develop appropriate solutions for the aggregation process. In many applications the knowledge and data provided by the experts tend to be vague, as experts express their knowledge in non-structured and ambiguous ways, for instance by using linguistic terms. We combine interval-valued fuzzy sets and OWA operators to create new aggregation methods and we prove that the new operators satisfy some important properties. In this article we present novel approaches for aggregating vague and imprecise information.

Keywords: Aggregation, Uncertainty, Induced OWA, Interval-valued Fuzzy Numbers, Project Selection.

1 Introduction

Decision making is a significant part of life, used in both everyday events as well as for advanced, organization wide decisions. These actions demand aggregation operations, which combine the different preferences into a general value, taking all the different parts into account, but still making it possible for different options to be differed amongst each other [1,2]. By definition, aggregation is the process of combining numerical values into a single number, representing the original set of numbers. Aggregation and the aggregation function performing the aggregation process have been successfully applied in several fields. The OWA operator [3] has, since its introduction, been used in many successful projects and applications [4]. Several successful extensions of OWA operators have also been introduced.

Type-2 fuzzy sets provide an extensive knowledge representation and approximate reasoning compared to what type-1 fuzzy sets can provide, especially regarding computing with words and human perception. Type-2 fuzzy sets offer an effective way of modelling experts' preferences, which can be quite vague and imprecise. In other words, fuzzy sets make it possible to use uncertain information but still produce accurate results [5]. Interval type-2 fuzzy sets (IT2FS) have become the most used subclass of type-2 fuzzy sets, eliminating some of the problems regarding type-2 fuzzy sets. Interval-valued fuzzy set theory is based on the notion that there is a lack

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 535–544.
DOI: 10.1007/978-3-642-36981-0_49

of objective procedures for selecting a crisp membership degree for the elements in the fuzzy set; therefore IT2FS specifies an interval-valued degree of membership to each element [6,7].

Project selection problem are complex decision-making processes as there are many factors that have to be taken into consideration. However, project selection is still a task that has to be undertaken regularly in many organizations. The methods used when making decisions are therefore crucial [8]. Project selection involves the simultaneous comparison of a number of projects based on certain measurements, where the goal is to find a desirable ranking of the projects. Due to the versatility of the Project selection cases, the problem often becomes NP-hard [9]. Naturally, the selection process becomes even more complex as multiple objectives are included in the model.

Many project-selection models have problems when dealing with uncertainty. Also, decision making becomes difficult when the available information is incomplete or imprecise. Another factor affecting the project selection process is that the actual decision has to be made by humans, even though there have been extensive calculations performed before the decisions. Modelling the human aspect has turned out to be a challenge for the traditional methods. By including fuzzy logic in the models, it opens for possibilities to deal with uncertainties in project selection models [8]. A good example regarding the above mentioned methods is the framework developed by Lin and Hsieh [10] which includes a portfolio matrix model, fuzzy weighted average, and fuzzy integer linear programming. The goal of this framework is to improve strategic project portfolio selection.

The rest of the paper is structured as follows. In Section 2, we introduce basic concepts from the theory of interval-valued fuzzy sets and OWA operators. Section 3 presents new definitions of interval-valued fuzzy number aggregation operators with the basic properties and the new definitions are illustrated with an example from project selection in Section 4. Finally Section 5 provides the conclusion and future research directions.

2 Preliminaries

In this section we shortly present the basic definitions and concepts that will be used in the paper: interval-valued fuzzy numbers, OWA operators and various generalizations of the OWA operator.

2.1 Interval-Valued Fuzzy Numbers

A *fuzzy number A* is a fuzzy set in $\mathbb{R}$ with a normal, fuzzy convex and continuous membership function of bounded support. The family of fuzzy numbers is denoted by F. Fuzzy numbers can be considered as possibility distributions.

Originally introduced by Zadeh in 1975 [6], type-2 fuzzy sets make it possible to model and minimize the uncertainty, in a more effective way than type-1 fuzzy sets. In comparison to type-1 fuzzy sets, where the membership values are crisp, type-2

fuzzy sets are characterized by having fuzzy values as membership functions. The three-dimensional membership function of type-2 fuzzy sets enables for better modelling of uncertainties. A special class of type-2 fuzzy sets is the class of interval-valued fuzzy sets.

Definition 1. [11] *An interval-valued fuzzy set A defined on X is given by*

$$A = \left\{\left(x, \left[\mu_A^L(x), \mu_A^U(x)\right]\right)\right\}, x \in X,$$

where $\mu_A^L(x), \mu_A^U(x) : X \to [0,1]; \forall x \in X, \mu_A^L(x) \le \mu_A^U(x)$ *and the ordinary fuzzy sets* $\mu_A^L(x)$ *and* $\mu_A^U(x)$ *are called lower fuzzy set and upper fuzzy set about A, respectively.*

We use the notation $\bar{\mu}_A(x) = [\mu_A^L(x), \mu_A^U(x)]$ for the interval assigned to x. $\bar{\mu}_A(x)$ can be seen as an interval-valued function from X to $[I] = \{[a, b]: a \le b, a, b \in I\}$. All interval-valued fuzzy sets on X are denoted by *IVF(X).* Since every $A \in IVF$ $(\mathbb{R})$ is uniquely associated with the corresponding membership function, throughout the paper we will use the notation $A(x) = \bar{\mu}_A(x)$ and similarly for the upper and lower fuzzy sets. We consider a subclass of *IVF* $(\mathbb{R})$: interval-valued fuzzy numbers (*IVFN*), which is simply the case when $A^L(x)$ and $A^U(x)$ are ordinary fuzzy numbers.

For the α-level sets of $A^L(x)$ and $A^U(x)$ we will use the notations $[A^L(x)]^\alpha = [a_1(\alpha), a_2(\alpha)]$, $[A^U(x)]^\alpha = [a^1(\alpha), a^2(\alpha)]$ and $[A]^\alpha = ([A^L(x)]^\alpha, [A^U(x)]^\alpha)$. The arithmetic operations of interval-valued fuzzy numbers can be defined using γ-cuts and the Extension Principle [6]. If $A, B \in IVFN$ with upper and lower membership functions $A^L(x)$, $A^U(x)$ and $B^L(x)$, $B^U(x)$, then the α-cuts of the upper and lower membership functions of $A * B$, where $* \in \{+, -, *\}$ are the following:

$$[(A * B)^U]^\alpha = [a^1(\alpha), a^2(\alpha)] * [b^1(\alpha), b^2(\alpha)]$$

and

$$[(A * B)^L]^\alpha = [a_1(\alpha), a_2(\alpha)] * [b_1(\alpha), b_2(\alpha)].$$

If $A \in IVFN$, then $B \in F$ is an embedded fuzzy number of A if

$$A^L(x) \le B(x) \le A^U(x),$$

for all $x \in \mathbb{R}$. The set of all the embedded fuzzy numbers of $A \in IVFN$ will be denoted by $F(A)$.

In the application of OWA operators, the reordering procedure is a crucial step. There exist several methods for ranking fuzzy quantities, specifically fuzzy numbers, but the literature on ranking procedures for interval-valued fuzzy sets does not offer many different approaches. Wu and Mendel [12] analysed two ranking methods for interval-valued fuzzy sets: Mitchell's method [13] and a centroid-based ranking method. In this paper we use the mean value of $A \in IVFN$ [14] to obtain the orderings:

Definition 2. *The mean (or expected) value of* $A \in$ *IVFN is defined as*

$$E(A) = \int_0^1 \alpha\left(M(U_\alpha) + M(L_\alpha)\right)\mathrm{d}\alpha \tag{1}$$

where U_α *and* L_α *are uniform probability distributions defined on* $[A^U]^\alpha$ *and* $[A^L]^\alpha$*, respectively, and M stands for the probabilistic mean operator.*

2.2 OWA Operators

OWA operators were introduced by Yager [3] as a new aggregation technique. It provides a family of aggregation operators, including the maximum and minimum. The OWA operators are especially useful in cases where the query cannot be formulated logically: in many cases, it is not possible to create the query by using logical connectors (e.g. *and* and *or*). OWA operators have proven to be especially useful when extending the purpose of t-norms and t-conorms. In some cases, t-conorms can often be too polarized towards the *and* and *or* operators.

Definition 3. *An OWA operator of dimension n is a mapping F:* $\mathbb{R}^n$ *->* $\mathbb{R}$ *that has an associated vector* $W = (w_1, \ldots, w_n)$ *such as* $w_i \in [0,1], 1 \le i \le n, \sum_{i=1}^n w_i = 1$. *Furthermore*

$$F(a_1,\ldots,a_n) = \sum_{j=1}^n w_j b_j = W^T B$$

$B = (b_1,\ldots,b_n)$*, where* b_j *is the j-th largest element of the bag* $\langle a_1,\ldots,a_n \rangle$.

2.3 Induced OWA

The induced ordered weighted averaging (IOWA) operator was introduced by Yager and Filev [15]. In this generalization of the original OWA operator, an order inducing value (u_i) is associated with each of the argument values. In this approach, the arguments are ordered based on the u_i values.

The procedure for calculating the OWA aggregation of these OWA pairs is the following:

$$F(\langle u_1, a_1 \rangle,\ldots,\langle u_n, a_n \rangle) = \sum_{j=1}^n w_j b_j$$

where b_j is the a_i value of the OWA pair having the *j*-th largest u_i value. The use of order introducing variables combined with the OWA operator allows decision makers to aggregate complex objects whose ordering may not be easily accomplished but which may be ordered with respect to some properties [15].

3 Aggregation of IVFN's with IOWA Operators

In this section the definition of IOWA operator is generalized for IVFN's. The definitions of different fuzzy extensions of the OWA operator [16,17,18] are special cases of the new definitions introduce in this section.

3.1 Interval-Valued Fuzzy Number-IOWA

Definition 4. *An IVFN-IOWA operator of dimension n is a mapping IVFN-IOWA:* $\mathbb{R}^n \times IVFN^n \rightarrow IVFN$ *that has an associated weighting vector W of dimension n with* $w_j \in [0,1], 1 \leq i \leq n, \sum_{j=1}^{n} w_j$, *such that*

$$IVFN - IOWA(\langle u_1, A_1\rangle,...,\langle u_n, A_n\rangle) = \sum_{j=1}^{n} w_j B_j \tag{2}$$

where, B_j *is the* A_i *value of the FIOWA pair* $\langle u_i, A_i\rangle$ *having the j-th largest* u_i *value,* u_i *is the order inducing variable and* A_i *is the argument variable represented in the form of IVFN's.*

Theorem 1. *If f is an IVFN-IOWA operator, then the following properties are satisfied:*

1. *f is commutative :*
$$f(\langle u_1, A_1\rangle,...,\langle u_n, A_n\rangle) = f(\langle u'_1, A'_1\rangle,...,\langle u'_n, A'_n\rangle)$$
where $\langle u'_1, A'_1\rangle,...,\langle u'_n, A'_n\rangle$ *is any permutation of the arguments.*
2. *f is monotone: if* $A_i \geq B_i, \forall i,$ *then*
$$f(\langle u_1, A_1\rangle,...,\langle u_n, A_n\rangle) \geq f(\langle u_1, B_1\rangle,...,\langle u_n, B_n\rangle)$$
3. *f is idempotent:*
$$f(\langle u_1, A\rangle,...,\langle u_n, A\rangle) = A$$
4. *f is bounded:*
$$\min_i\{A_i\} \leq f(\langle u_1, A_1\rangle,...,\langle u_n, A_n\rangle) \leq \max_i\{A_i\}.$$

Proof. The first and third statements follow from the definition of the IVFN-IOWA operator and the arithmetical operations of IVFN's. For statements 2 and 3, we can use that the mean value of IVFN's is a linear operator [14]. The monotonicity of f can be shown as

$$E(f(\langle u_1, A_1\rangle,...,\langle u_n, A_n\rangle)) = E(\sum_{j=1}^{n} w_j B_j) = \sum_{j=1}^{n} w_j E(C_j) \geq \sum_{j=1}^{n} w_j E(D_j)$$

$$= E(\sum_{j=1}^{n} w_j D_j) = E(f(\langle u_1, B_1\rangle,...,\langle u_n, B_n\rangle))$$

where $C_j(D_j)$ is the $A_i(B_i)$ value of the pair $\langle u_i, A_i\rangle(\langle u_i, B_i\rangle)$ having the j-th largest u_i. The boundedness can be proven by comparing the mean value of the aggregated value to the minimum and maximum as follows:

$$E(\min_i\{A_i\}) = \sum_{j=1}^{n} w_j E(\min_i\{A_i\}) \le \sum_{j=1}^{n} w_j E(B_j) = E(\sum_{j=1}^{n} w_j B_j) = E(f(\langle u_1, A_1\rangle,...,\langle u_n, A_n\rangle))$$

and

$$E(\max_i\{A_i\}) = \sum_{j=1}^{n} w_j E(\max_i\{A_i\}) \ge \sum_{j=1}^{n} w_j E(B_j) = E(\sum_{j=1}^{n} w_j B_j) = E(f(\langle u_1, A_1\rangle,...,\langle u_n, A_n\rangle))$$

where B_j is the A_i value of the pair (u_i, A_i) having the j-th largest u_i.

Example 1. To illustrate the concept, we will calculate the aggregation of triangular shaped IVFN's (the upper and lower fuzzy numbers are triangular fuzzy numbers). The upper and lower triangular fuzzy numbers can be represented as $A^L = (a, \alpha, \beta)$ and $A^U = (a, \theta, \tau)$ respectively, where a stands for the centre, (α, β) and (θ, τ) denotes the left and right width of the fuzzy numbers. The mean value of a trapezoidal IVFN can be expressed as [14]

$$E(A) = a + \frac{\beta - \alpha}{12} + \frac{\tau - \theta}{12}.$$

In the example we will use the following three trapezoidal IVFN's: $A_1^L = (5, 1, 2)$, $A_1^U = (5, 3, 3)$, $A_2^L = (3, 2, 4)$, $A_2^U = (3, 3, 6)$, $A_3^L = (8, 2, 2)$, $A_3^U = (8, 3, 3)$, and the corresponding order inducing variables are $u_1 = 3$, $u_2 = 2$, $u_3 = 5$. The weights are defined as W = (0.3, 0.4, 0.3). The aggregation can be calculated as IVFN-IOWA $(\langle 3, A_1\rangle, \langle 2, A_2\rangle, \langle 5, A_3\rangle) = 0.3 * A_3 + 0.4 * A_1 + 0.3 * A_2$. Using the arithmetic of IVFN's, we obtain that the aggregated value, A, can be described by the lower fuzzy number $A^L = (5.3, 1.6, 2.6)$ and the upper fuzzy number $A^U = (5.3, 3, 3.9)$.

3.2 Quasi Interval-Valued Fuzzy Number-IOWA

Definition 5. *A Quasi IVFN-IOWA operator of dimension n is a mapping QIVFN-IOWA: $\mathbb{R}^n \times IVFN^n \rightarrow IVFN$ that has an associated weighting vector W of dimension n with $w_j \in [0,1], 1 \le i \le n, \sum_{j=1}^{n} w_j$ such that:*

$$Quasi\ IVFN - IOWA(\langle u_1, A_1\rangle,...,\langle u_n, A_n\rangle) = g^{-1}\left(\sum_{j=1}^{n} w_j g(B_j)\right) \quad (3)$$

where, B_j is the A_i value of the pair $\langle u_i, A_i\rangle$ having the j-th largest u_i value, u_i is the order inducing variable and A_i is the argument variable represented in the form of IVFN's and g: $IVFN^n \rightarrow IVFN$ is a continuous strictly monotone function.

Theorem 2. *If f is a Quasi IVFN-IOWA operator, then it is commutative, monotone, idempotent, and bounded.*

Proof. The first and third statements follow from the definition of the Quasi IVFN-IOWA operator and the arithmetical operations of IVFN's. The mono-tonicity follows from the properties of function g and the linearity of the mean value. The

boundedness can be proven by comparing the mean value of the aggregated value to the minimum and maximum as follows:

$$E(f(\langle u_1, A_1\rangle,...,\langle u_n, A_n\rangle)) = E\left[g^{-1}\left(\sum_{j=1}^{n} w_j g(B_j) \right)\right] \geq E\left[g^{-1}\left(\sum_{j=1}^{n} w_j g(\min_i\{A_i\}) \right)\right]$$

$$= E\left[g^{-1}\left(g(\min_i\{A_i\}) \sum_{j=1}^{n} w_j \right)\right] = E(\min_i\{A_i\})$$

using that $\sum_{j=1}^{n} w_j = 1$, *and*

$$E(f(\langle u_1, A_1\rangle,...,\langle u_n, A_n\rangle)) = E\left[g^{-1}\left(\sum_{j=1}^{n} w_j g(B_j) \right)\right] \leq E\left[g^{-1}\left(\sum_{j=1}^{n} w_j g(\max_i\{A_i\}) \right)\right]$$

$$= E\left[g^{-1}\left(g(\max_i\{A_i\}) \sum_{j=1}^{n} w_j \right)\right] = E(\max_i\{A_i\})$$

where B_i *is the* A_i *value of the pair* $\langle u_i, A_i\rangle$ *having the jth largest* u_i.

4 Project Selection with the IVFN-IOWA Operator

To illustrate the applicability and usefulness of the proposed aggregating operators, we will consider a project selection problem. In general, a multi-attribute decision making problem with multiple experts can be processed through the following steps:

- Step 1. Selection of criteria and alternatives: we define the appropriate set of selection criteria C = $\{c_1,..., c_m\}$ and the set of potential alternative solutions $A = \{a_1,..., a_n\}$.

- Step 2. Defining the evaluation measure: in this case, experts express their opinionintermsoflinguisticlabelsrepresentedbyinterval-valuedfuzzynumbers. For example, the experts can provide their evaluation about the degree to which a given alternative satisfies a criterion. One possible representation using IVFN's is described in Table 1. We use trapezoidal shaped upper and lower fuzzy numbers and we use the notation $A = (a, b, c, d)$, where $[b, c]$ is the core of the fuzzy number and a and d are the left and right endpoints of the support, respectively.

- Step 3.Experts specify their opinion: every expert $E_k, k \in \{1,...,l\}$ provides his/her evaluation in the form of a matrix $(A_{ij}^k)_{n \times m}$ where $A_{ij}^k \in \{$*very low*, *low*, *medium*, *high*, *very high*$\}$.

- Step 4. Aggregate into one decision matrix: every expert E_k is associated with a weight s_k such that $s_k \in [0, 1]$, $\sum_{k=1}^{l} s_k = 1$. The weights define the importance (impact) of the experts within the group. Using these weights, the individual evaluations are aggregated using the arithmetic operations of IVFN's; the result is denoted by $(A_{ij})_{n \times m}$
- Step 5. Aggregate individually for every alternative: using the matrix of aggregated payoffs, we obtain the overall evaluation for every alternative individually by employing the IVFN-IOWA operator. The alternative with the highest value will be selected and an ordering for the alternatives is established.

Table 1. Linguistic labels represented by trapezoidal IVFN's

	Upper fuzzy number	Lower fuzzy number
Very Low	(0,0,0.2,0.4)	(0,0,0.15,0.3)
Low	(0,0.2,0.4.,0.6)	(0.1,0.25,0.35,0.5)
Medium	(0.2,0.4,0.6,0.8)	(0.3,0.45,0.5,0.7)
High	(0.4,0.6,0.8,1)	(0.5,0.65,0.75,0.9)
Very High	(0.6,0.8,1,1)	(0.7,0.85,0.95,1)

Many scholars have emphasized the importance of research and development (R&D) projects in the field of information technology (IT)[19].As innovations are crucial to ensure the profitability of a company in the IT sector, selecting the best R&D projects constitutes an important part of the decision making processes within companies. In this example, we assume that the decision-makers (experts) consider the following 5 criteria when evaluating candidate projects:

1. Competitiveness of technology
2. The potential size of market
3. Environmental and safety benefits
4. Return on development cost
5. Opportunity of project result implementation

Table 2. The preference relation of the 4 alternatives constructed by 3 experts

E_1	C_1	C_2	C_3	C_4	C_5
A_1	low	medium	very low	low	medium
A_2	high	low	very high	medium	very low
A_3	medium	medium	medium	medium	high
A_4	low	very high	high	medium	medium

E_2	C_1	C_2	C_3	C_4	C_5
A_1	high	medium	low	very high	medium
A_2	low	very low	high	medium	high
A_3	low	very high	medium	very low	high
A_4	high	very high	low	very low	medium

E_3	C_1	C_2	C_3	C_4	C_5
A_1	medium	high	very low	high	medium
A_2	low	medium	very high	very low	medium
A_3	low	high	very high	low	medium
A_4	very high	medium	high	medium	medium

The preferences are expressed using trapezoidal IVFN's as described in Table 1. The preference matrices specified by 3 experts for the 4 alternative projects considered are listed in Table 2.

The weights that represent the importance of the experts in this example are the following: S = (0.2, 0.5, 0.3). The aggregated payoff matrices for the four alternatives can be obtained using the importance weights of the experts and arithmetical operation on IVFN's (weighted average).

Table 3. Order inducing values

Lower	C_1	C_2	C_3	C_4	C_5
A_1	3	4	1	5	2
A_2	2	4	3	5	1
A_3	1	3	5	4	2
A_4	5	4	1	3	2

Table 4. The obtained evaluation of the 4 alternatives

Alternative	Lower fuzzy number	Upper fuzzy number	Mean value	Ranking
A_1	(0.34,0.49,0.59,0.73)	(0.25,0.44,0.64,0.82)	0.54	2
A_2	(0.31,0.44,0.56,0.69)	(0.23,0.40,0.60,0.77)	0.50	4
A_3	(0.38,0.52,0.64,0.76)	(0.29,0.48,0.68,0.84)	0.57	1
A_4	(0.34,0.47,0.59,0.71)	(0.25,0.43,0.63,0.80)	0.53	3

The weights of the IVFN-IOWA are specified as W = (0.2, 0.15, 0.25, 0.3, 0.1). Considering possible different characteristics of R&D projects, induced ordering variables are introduced to emphasize different criteria for different alternatives (see Table 3). The final results of the project selection process are shown in Table 4. The mean value is employed to obtain the ordering. The alternative with the best payoff is A3, and the order is the following: A3 > A1 > A4 > A2.

5 Conclusion

In this paper we introduced a generalization of the OWA operator which allows for using interval-valued fuzzy numbers to be used as the arguments of the aggregation process. The use of induced ordering variables in the aggregation make it possible to model complex problems in the presence of imprecision and vagueness. The new aggregation operators satisfy the important properties of commutativity, monotonicity, idempotency and boundedness. We illustrated the use of the new aggregation operator through an R&D project selection problem. In the future, the operators will be employed to facilitate the use of interval-valued fuzzy numbers in fuzzy ontology construction and reasoning procedures.

References

1. Chiclana, F., Herrera-Viedma, E., Herrera, F., Alonso, S.: Some induced ordered weighted averaging operators and their use for solving group decision-making problems based on fuzzy preference relations. European Journal of Operational Research 182(1), 383–399 (2007)
2. Xu, Z., Da, Q.: An overview of operators for aggregating information. International Journal of Intelligent Systems 18(9), 953–969 (2003)
3. Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Transactions on Systems, Man, and Cybernetics 18(1), 183–190 (1988)
4. Yager, R.R., Kacprzyk, J., Beliakov, G. (eds.): Recent Developments in the Ordered Weighted Averaging Operators: Theory and Practice. STUDFUZZ, vol. 265. Springer, Heidelberg (2011)
5. Zhou, S., John, R., Chiclana, F., Garibaldi, J.: On aggregating uncertain information by type-2 owa operators for soft decision making. International Journal of Intelligent Systems 25(6), 540–558 (2010)
6. Zadeh, L.A.: The concept of a linguistic variable and its applications to approximate reasoning. Information Sciences, Part I, II, III, 8, 8, 9,199–249, 301–357, 43–80 (1975)
7. Zhang, Z.: On characterization of generalized interval type-2 fuzzy rough sets. Information Sciences 219(0), 124–150 (2013)
8. Machacha, L., Bhattacharya, P.: A fuzzy-logic-based approach to project selec¬tion. IEEE Transactions on Engineering Management 47(1), 65–73 (2000)
9. Gutjahr, W., Katzensteiner, S., Reiter, P., Stummer, C., Denk, M.: Multi-objective decision analysis for competence-oriented project portfolio selection. European Journal of Operational Research 205(3), 670–679 (2010)
10. Lin, C., Hsieh, P.: A fuzzy decision support system for strategic portfolio man-agement. Decision Support Systems 38(3), 383–398 (2004)
11. Gorzałczany, M.: A method of inference in approximate reasoning based on interval-valued fuzzy sets. Fuzzy sets and systems 21(1), 1–17 (1987)
12. Wu, D., Mendel, J.: A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets. Information Sciences 179(8), 1169–1192 (2009)
13. Mitchell, H.: Ranking type-2 fuzzy numbers. IEEE Transactions on Fuzzy Systems 14(2), 287–294 (2006)
14. Carlsson, C., Fuller, R., Mezei, J.: Project selection with interval-valued fuzzy numbers. In: 2011 IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI), pp. 23–26. IEEE (2011)
15. Yager, R.R., Filev, D.P.: Induced ordered weighted averaging operators. IEEE Transactions on Systems, Man, and Cybernetics 29(2), 141–150 (1999)
16. Merigó, J., Gil-Lafuente, A.: Fuzzy induced generalized aggregation operators and its application in multi-person decision making. Expert Systems with Applications 38(8), 9761–9772 (2011)
17. Merigó, J., Casanovas, M.: The fuzzy generalized owa operator and its application in strategic decision making. Cybernetics and Systems: An International Journal 41(5), 359–370 (2010)
18. Chen, S.J., Chen, S.M.: A new method for handling multicriteria fuzzy decision -making problems using fn-iowa operators. Cybernetics & Systems 34(2), 109–137 (2003)
19. Lee, Y., Kim, S., Lee, H.: The impact of service r&d on the performance of Korean information communication technology small and medium enterprises. Journal of Engineering and Technology Management 28(1), 77–92 (2011)

Deriving Weights from Group Fuzzy Pairwise Comparison Judgement Matrices

Tarifa S. Almulhim, Ludmil Mikhailov, and Dong-Ling Xu

Manchester Business School, University of Manchester, Booth Street East, Manchester M15 6PB, UK
Tarifa.Almulhim@postgrad.mbs.ac.uk
ludmil.mikhailov@manchester.ac.uk
ling.xu@mbs.ac.uk

Abstract. Several Multi-Criteria Decision Making (MCDM) methods involve pairwise comparisons to obtain the preferences of decision makers (DMs). This paper proposes a fuzzy group prioritization method for deriving group priorities/weights from fuzzy pairwise comparison matrices. The proposed method considers the different importance weights of multiple DMs by extending the Fuzzy Preferences Programming Method (FPP). The elements of the group pairwise comparison matrices are presented as fuzzy numbers rather than exact numerical values in order to model the uncertainty and imprecision in the DMs' judgments. Unlike the known fuzzy prioritization techniques, the proposed method is able to derive crisp weights from incomplete and fuzzy set of comparison judgments and doesn't require additional aggregation procedures. A prototype of a decision tool is developed to assist DMs to use the proposed method for solving fuzzy group prioritization problems. A detailed numerical example is used to illustrate the proposed approach.

Keywords: Fuzzy Non-linear Programming. Fuzzy Preferences Programming Method. Multiple Criteria Decision-Making. Triangular Fuzzy Number.

1 Introduction

There are various techniques for deriving priorities/weights for decision elements (e.g. attributes/criteria), see [1] and [2] for a review. These techniques are based on either direct weighting or on pairwise comparison methods.

In direct weighting, the decision maker (DM) is directly asked to give values between 0 and 1 to each decision element to assign their importance. Some methods for deriving attributes/criteria weights by direct assigning techniques are: the Simple Multi-Attribute Rating Technique (SMART) [3], SWING weighting methods [4], and SMART Exploiting Ranks (SMARTER) [5].

When the DM or a group of DMs is unable to directly assign decision elements' weights, the Pairwise Comparison (PC) method proposed in [6] can be used.

Psychological experiments have shown that weight derivation from PC is much more accurate than direct weighting [8]. Therefore, the PC methods are often used as an intermediate step in many MCDM methods, as Analytic Hierarchy Process (AHP)

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 545–555.
DOI: 10.1007/978-3-642-36981-0_50

[7], Analytic Network Process (ANP) [8], PROMETHEE [9], and Evidential Reasoning (ER) [10].

The PC methods require construction of Pairwise Comparisons Judgment Matrices (PCJMs). In order to construct a PCJM, the DM is asked to compare pairwisely any two decision elements and provide a numerical / linguistic judgment for their relative importance. Thus, the DM gives a set of ratio judgments to indicate the strength of his/her preferences, which are structured in a reciprocal PCJM. Then, the weights or priority vectors of the decision elements can be derived from the PCJM by applying some prioritization method.

There are numerous Pairwise Comparisons Prioritization Methods (PCPMs), as the Eigenvector Method [7], the Direct Least Squares Method [11], the rank-ordering method [7] and the Logarithmic Least Square Method [12]. Choo and Wedley [1] summarized and analyzed 18 PCPMs for deriving a priority vector from PCJMs. They discussed that no method performs best in all situations and no method dominates the other methods.

However, in many practical cases, in the process of prioritization the DMs are unable to provide crisp values for comparison ratios. A natural way to deal with the uncertainty and imprecision in the DMs' judgments is to apply the fuzzy set theory [14] and to represent the uncertain DMs' judgments as fuzzy numbers. Thus, Fuzzy PCJMs can be constructed and used to derive the priority vectors by applying some Fuzzy PCPMs. Such methods are proposed by Laarhoven and Pedrycz's [15], Buckley [14], Chang [16]; Mikhailov [17], and applied for group decision making.

The existing fuzzy PCPMs have some drawbacks. They require an additional defuzzification procedure to convert fuzzy weights into crisp (non-fuzzy) weights. However, different defuzzification procedures will often give different solutions.

The linear and non-linear variants of the Fuzzy Preference Programming (FPP) method [17] do not require such defuzzification procedures, but their group modifications assume that all the DMs have the same weight of importance. However, in the real group decision making problems, sometimes some experts are more experienced than others. Therefore the final results should be influenced by the degree of importance of each expert.

In order to overcome some of the limitation of the group FPP method, a new group version of the FPP method is proposed by introducing importance weights of DMs in order to derive weights for decision elements in group decision problems. The proposed method has some attractive features. It does not require any aggregation procedures. Moreover, it does not require a defuzzification procedure and derives crisp priorities/weights from an incomplete set of fuzzy judgments and incomplete fuzzy PCJMs.

For applying the proposed method and solving prioritization problems, a Non-Linear FPP Solver is developed based on the MATLAB Optimisation Toolbox. This decision tool is used for solving a specific numerical example.

The remainder of this paper is organized as follows. In Section 2, representation of the fuzzy group prioritization problem is briefly explained. Then, the proposed method is presented in Section 3 and illustrated by a numerical example in section 4. The developed Non-Linear FPP Solver is presented in section 5, followed by conclusions.

2 Representation of the Fuzzy Group Prioritization Problem

Consider a group of K DMs ($DM_k, k = 1,2,..., K$) that evaluates n elements E_1,.. E_n (in MCDM, these elements could be clusters, criteria, sub-criteria or alternatives). With respect to some fixed preference scale, each DM assesses the relative importance of any two elements (E_i, E_j) $(i, j = 1,2,..,n)$ by providing a ratio judgment a_{ijk}, specifying by how much E_i is preferred/not preferred to E_j.

In a fuzzy environment, suppose that each DM provides a set of y fuzzy comparison judgements $A^k = \{\tilde{a}_{ijk}\}$, $y \le n(n-1)/2$, where $i = 1,2,..,n-1,$, $j \succ i$, $j = 2,3,.....n,$ $k = 1,2,.., K$ and those judgments are represented as Triangular Fuzzy Numbers (TFNs) $\tilde{a}_{ijk} = (l_{ijk}, m_{ijk}, u_{ijk})$, where l_{ijk}, m_{ijk} and u_{ijk} are the lower bound, the mode and the upper bound, respectively.

The set A^k can be used to form a Fuzzy PCJM of the form (1):

$$A^k = \begin{bmatrix} (1,1,1) & (l_{12k}, m_{12k}, u_{12k}) & .... & (l_{1jk}, m_{1jk}, u_{1jk}) \\ (l_{21k}, m_{21k}, u_{21k}) & (1,1,1) & .... & (l_{2jk}, m_{2jk}, u_{2jk}) \\ ... & ... & .... & ... \\ (l_{i1k}, m_{i1k}, u_{i1k}) & (l_{i2k}, m_{i2k}, u_{i2k}) & .... & (1,1,1) \end{bmatrix} \quad (1)$$

Then, the fuzzy group prioritization problem is to determine a crisp priority vector (crisp weights) $w = (w_1, w_2,..., w_n)^T$ from all A^k, $k = 1,2,..., K$, which represents the relative importance of the n elements.

3 Group Fuzzy Preference Programming Method

The non-linear FPP method [17] derives a priority vector $w = (w_1, w_2,..., w_n)^T$, which satisfies:

$$l_{ij} \tilde{\le} w_i / w_j \tilde{\le} u_{ij} \quad (2)$$

where $\tilde{\le}$ denotes 'fuzzy less or equal to'. If M is the overall number of fuzzy group comparison judgments, then $2M$ fuzzy constraints of the type (3) are obtained.

$$\begin{aligned} -w_i + w_j l_{ij} &\tilde{\le} 0 \\ w_i - w_j u_{ij} &\tilde{\le} 0 \end{aligned} \quad (3)$$

For each fuzzy judgement, a membership function, which represents the DMs' satisfaction with different crisp solution ratios, is introduced:

$$\mu_{ij}(w_i/w_j) = \begin{cases} \dfrac{(w_i/w_j) - l_{ij}}{m_{ij} - l_{ij}}, & w_i/w_j \le m_{ij} \\ \dfrac{u_{ij} - (w_i/w_j)}{u_{ij} - m_{ij}}, & w_i/w_j \ge m_{ij} \end{cases} \tag{4}$$

The solution to the prioritization problem by the FPP method is based on two assumptions. The first on requires the existence of a *non-empty fuzzy feasible area* $\tilde{P}$ on the $(n-1)$ dimensional simplex Q^{n-1},

$$Q^{n-1} = \{(w_1, w_2, ..., w_n), w_i \succ 0, \sum_{i=1}^{n} w_i = 1\} \tag{5}$$

The fuzzy feasible area $\tilde{P}$ is defined as an intersection of the membership functions (4). The membership function of the fuzzy feasible area $\tilde{P}$ is given by:

$$\mu_{\tilde{P}}(w) = [Min\{\mu_1(w), \mu_2(w), ...\mu_{2M}(w)\} \setminus \sum_{i=1}^{n} w_i = 1] \tag{6}$$

The second assumption identifies a selection rule, which determines a priority vector, having the highest degree of membership in the aggregated membership function (6). Thus, there is *a maximizing solution* w^* (a crisp priority vector) that has a maximum degree of membership λ^* in $\tilde{P}$, such that :

$$\lambda^* = \mu_{\tilde{P}}(w^*) = Max[Min\{\mu_1(w), ..., \mu_{2M}(w)\} \setminus \sum_{i=1}^{n} w_i = 1] \tag{7}$$

A new decision variable λ is introduced which measures the maximum degree of membership in the fuzzy feasible area $\tilde{P}$. Then, the optimization problem (7) is represented as

$$\begin{aligned} &Max \quad \lambda \\ &s.t. \\ &\lambda \le \mu_{ij}(w) \\ &\sum_{i=1}^{n} w_i = 1, \quad w_i \succ 0, \quad i = 1,2,...,n, \quad j = 1,2,..n, \quad j \succ i \end{aligned} \tag{8}$$

The above max-min optimization problem (8) is transformed into the following non-linear optimization problem:

$$\begin{aligned} &Max \quad \lambda \\ &s.t. \\ &(m_{ij} - l_{ij})\lambda w_j - w_i + l_{ij} w_j \le 0 \\ &(u_{ij} - m_{ij})\lambda w_j + w_i - u_{ij} w_j \le 0 \\ &i = 1,2,...n-1; \quad j = 2,3,..n \;\; ; \;\; j \succ i \;\; ; \\ &\textstyle\sum_{i=1}^{n} w_i = 1 \;\; ; \quad w_i \succ 0; \quad i = 1,2,...,n \end{aligned} \tag{9}$$

The non-linear FPP method can be extended for solving group prioritization problems. Mikhailov *el. at.* [20] propose a Weighted FPP method to fuzzy group

prioritization problem by introducing the importance weights of DMs. However, Weighted FPP method requires an additional aggregation technique to obtain the priority vector at different α - threshold. Consequently, this process is time consuming due to several computation steps needed for applying the α - threshold concept. Therefore, this paper modifies the non-linear FPP method [17], which can derive crisp weights without using α - threshold and by introducing the DMs' importance weights.

When we have a group of K DMs, the problem is to derive a crisp priority vector, such that priority ratios w_i / w_j are approximately within the scope of the initial fuzzy judgments a_{ijk} provided by those DMs, i.e.

$$l_{ijk} \lesssim w_i / w_j \lesssim u_{ijk} \tag{10}$$

The ratios w_i / w_j can also express the satisfaction of the decision makers, because ratios explain how similar the crisp solutions are close to the initial judgments from the DMs.

The inequality (10) can be represented as two single-side fuzzy constraints of the type (3):

$$R_q^{\,k} W \lesssim 0, \quad k = 1,.., K \;\; q = 1,2,..2M_k \tag{11}$$

The degree of the DMs' satisfaction can be measured by a membership function with respect to the unknown ratio w_i / w_j :

$$\mu_q^{\,k}(R_q^{\,k} W) = \begin{cases} \dfrac{\left(w_i^k / w_j^k\right) - l_{ijk}}{m_{ijk} - l_{ijk}}, & w_i^k / w_j^k \le m_{ijk} \\ \dfrac{u_{ijk} - \left(w_i^k / w_j^k\right)}{u_{ijk} - m_{ijk}}, & w_i^k / w_j^k \ge m_{ijk} \end{cases} \tag{12}$$

We can define K fuzzy feasible areas, $\tilde{P}_k$ as intersection of the membership functions (12), corresponding to the k -th DMs' fuzzy judgments and define the group fuzzy feasible area $\tilde{P} = \bigcap \tilde{P}_k$.

By introducing a new decision variable λ_k , which measures the maximum degree of membership of a given priority vector in the fuzzy feasible area $\tilde{P}_k$, we can formulate a max-min optimization problem of the type (8), which can be represented into:

$$\begin{aligned} &Max \;\; \lambda_k \\ &s.t. \\ &\lambda_k \le \mu_q^{\,k}(R_q^{\,k} W) \\ &\sum_{i=1}^{n} w_i = 1, \quad w_i \succ 0, \quad i = 1,2,...,n, \quad k = 1,.., K \;\; q = 1,2,..2M_k \end{aligned} \tag{13}$$

For introducing the DMs' importance weights, let us define I_k as the importance weight of the $DM_k; k = 1,2,..., K$. For aggregating all individual models of type (13) into a single group model a weighted additive goal-programming (WAGP) model [18] is applied.

The WAGP model transforms the multi-objective decision-making problem to a single objective problem. Therefore, it can be used to combine all individual models (13) into a new single model by taking into account the DMs' importance weights.

The WAGP model considers the different importance weights of goals and constraints and is formulated as:

$$\begin{aligned} &\mu_D(x) = \textstyle\sum_{s=1}^{p} \alpha_s \mu_{z_s}(x) + \sum_{r=1}^{h} \beta_r \mu_{g_r}(x) \\ &\textstyle\sum_{s=1}^{p} \alpha_s + \sum_{r=1}^{h} \beta_r = 1 \end{aligned} \tag{14}$$

Where:

μ_{z_S} are membership functions for the p–th fuzzy goal $z_S, s = 1,2,...p$;

μ_{g_r} are membership functions of the h -th fuzzy constraints $g_r, r = 1,2,...h$;

x is the vector of decision variables;

α_s are weighting coefficients that show the relative important of the fuzzy goals;

β_r are weighting coefficients that show the relative important of the fuzzy constraints.

A single objective model in WAMP is the maximization of the weighted sum of the membership functions μ_{z_S} and μ_{g_r} . By introducing new decision variables λ_s and γ_r, the model (14) can be transformed into a crisp single objective model, as follows:

$$\begin{aligned} &Max \ \ \textstyle\sum_{s=1}^{p} \alpha_s \lambda_s + \sum_{r=1}^{h} \beta_r \gamma_r \\ &s.t. \\ &\lambda_s \le \mu_{z_s}(x), \qquad s = 1,2,...p \\ &\gamma_r \le \mu_{g_r}(x), \qquad r = 1,2,...h \\ &\textstyle\sum_{s=1}^{p} \alpha_s + \sum_{r=1}^{h} \beta_r = 1 \\ &\lambda_s, \gamma_r \in [0,1]; \qquad \alpha_s, \beta_r \ge 0 \end{aligned} \tag{15}$$

In order to derive a group model, where the DMs have different importance weights, we exploit the similarity between the models (13) and (15). However, the non-linear FPP model (13) does not deal with fuzzy goals; it just represents the non-linear fuzzy constraints. Thus, by taking into the account the specific form of $R_q^{\,k} W \lesssim 0$, and introducing the important weights of the DMs, the problem can be further presented into a non-linear program by utilizing WAGP model as:

$$
\begin{aligned}
&Max \quad Z = \textstyle\sum_{k=1}^{K} I_k \lambda_k \\
&s.t. \\
&(m_{ijk} - l_{ijk})\lambda_k w_j - w_i + l_{ijk} w_j \le 0 \\
&(u_{ijk} - m_{ijk})\lambda_k w_j + w_i - u_{ijk} w_j \le 0 \\
&i = 1,2,...n-1; \quad j = 2,3,..n \quad ; \; j \succ i \; ; \; k = 1,2,...K \\
&\textstyle\sum_{i=1}^{n} w_i = 1 \; ; \quad w_i \succ 0; \qquad i = 1,2,...,n
\end{aligned}
\tag{16}
$$

Where the decision variable λ_k measures the degree of the DM's satisfaction with the final priority vector $w = (w_1, w_2, ..., w_n)^T$, I_k denotes the importance weight of the k-th DM, $k = 1,2,...K$.

In (16), the value of Z can be considered as a consistency index, as it measures the overall consistency of the initial set of fuzzy judgments. When the set of fuzzy judgments is consistent, the optimal value of Z is greater or equal to one. For the inconsistent fuzzy judgments, the maximum value of Z takes a value less than one.

For solving the non-linear optimization problem (16), an appropriate numerical method should be employed. In this paper, the solution is obtained by using MATLAB Optimization Toolbox, and a Non-linear FPP solver is developed to solve the prioritization problem.

4 An Illustrative Example

This example is given to illustrate the proposed method and also the solution by using the Non-linear FPP Solver. Moreover, this example demonstrates how the importance weights of DMs influence the final group ranking.

We consider the example in [20], where three DMs ($K = 3$) assess three elements ($n = 3$), and the importance weights of DMs are given: $I_1 = 0.3$, $I_2 = 0.2$, $I_3 = 0.5$.

The DMs provide an incomplete set of five fuzzy judgments, presented as TFNs:

DM 1: $a_{121} = (1,2,3)$; $a_{131} = (2,3,4)$.

DM 2: $a_{122} = (1.5,2.5,3.5)$; $a_{132} = (3,4,5)$.

DM 3: $a_{123} = (2,3,4)$.

The group fuzzy prioritization problem is to derive a crisp priority vector $w = (w_1, w_2, w_3)^T$ that approximately satisfies the following fuzzy constraints:

For DM 1: $1 \tilde{\le} w_1 / w_2 \tilde{\le} 3$; $2 \tilde{\le} w_1 / w_3 \tilde{\le} 4$.

For DM 2: $1.5 \tilde{\le} w_1 / w_2 \tilde{\le} 3.5$; $3 \tilde{\le} w_1 / w_3 \tilde{\le} 5$.

For DM 3: $2 \tilde{\le} w_1 / w_2 \tilde{\le} 4$.

The weights obtained by applying the method proposed in the previous section are $w_1 = 0.621$, $w_2 = 0.212, w_3 = 0.167$.

This solution can be compared with the crisp results from the example in [20] as shown in Table 1. We may observe that we have the same final ranking $w_1 \succ w_2 \succ w_3$, from applying the two different prioritization methods. However, the Weighted FPP method [20] applies an aggregation procedure for obtaining the crisp vector from different values of priorities at different α - threshold, while, the proposed non-linear group FPP method does not require an additional aggregation procedure.

Table 1. Results from the two prioritization methods

Methods	w_1	w_2	w_3
Weighted FPP method [a]	0.615	0.205	0.179
Non-linear FPP method [b]	0.623	0.216	0.161

[a] The method proposed in [16] with applying α - threshold.

[b] The method proposed in this paper without applying α - threshold.

If the third DM, who has the highest important weight provides a new fuzzy comparison judgment $a_{323} = (1,2,3)$, which means that the third element is about two times more important than the second element, the weights obtained by using the proposed Non-Linear FFP method are: $w_1 = 0.538$, $w_2 = 0.170$, $w_3 = 0.292$ and the final ranking is $w_1 \succ w_3 \succ w_2$. Consequently, it can be observed that the third DM's judgments strongly influence the final ranking. However, if the importance weight of the third DM is lower to the first two DMs' weights, then the new fuzzy comparison judgment does not change the final ranking. Thus, we can notice the significance of introducing importance weights of the DMs to the fuzzy group prioritization problem.

According to the computation time for solving the fuzzy group prioritization problem, the proposed method does not need an additional procedure to aggregate the priorities at the different α -levels. Therefore, the proposed method in this paper demands less computation time than the Weighted FPP method [20].

The computation time of the proposed method has been investigated by using the Non-Linear FFP Solver. It was found that the group non-linear FFP method performs significantly faster compared to the Weighted FPP [20] with different α - threshold ($\alpha = 0, 0.2 , 0.5 , 0.8 , 1$), as seen in Fig. 1.

We can conclude that the average of computation time (Minutes) for the Weighted FPP method highly increases as the number of decision elements n increases, comparing with the proposed method. Hence, these results show that the method proposed in this paper is more efficient with respect to the computation time.

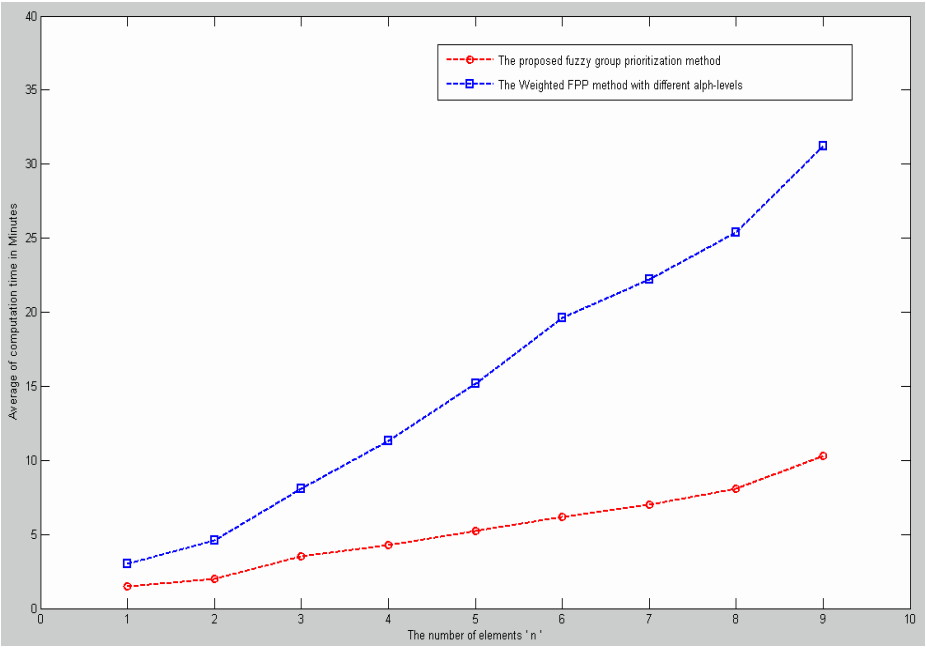


Fig. 1. Average Computation Time (Minutes)

5 Software Implementation Using MATLAB

We use the Optimization Toolbox in MATLAB functions and the functions MATLAB Graphical User Interface (GUI) to implement the proposed group non-linear FPP method. Essentially, there are three steps for programming and developing of the Non-Linear FFP solver:

Step 1: Coding the model into the system. A number of functions are available in MATLAB to solve the non-linear programming problem. In our prototype, the optimization problem is solved using sequential quadratic programming procedure [19].

Step 2: Creating a basic user interface. In this step the interface is designed which can run in the MATLAB command window. The aim of this user interface is to obtain the input from the DMs. The input information which should be acquired includes the total number of decision elements, the name of these elements, the total number of DMs, the importance weights of the DMs, and the fuzzy judgments.

The main feature in the developed interface is that the user can input the fuzzy judgments into the system directly and easily. According to the example from the previous section, the fuzzy judgments for the DM 1 are illustrated in Fig.2. However, if the fuzzy judgments between two elements are missing, the user can click the **'Missing Data'** button then the system will temporarily put -1 for the comparison, the negative value is not a true judgment in the real world, it just indicates that those

elements should not be included in the further calculations. For instance, in the given example, the judgment a_{231} is missing for the DM1 and can be sorted as $(-1,-1,-1)$, Fig. 2.

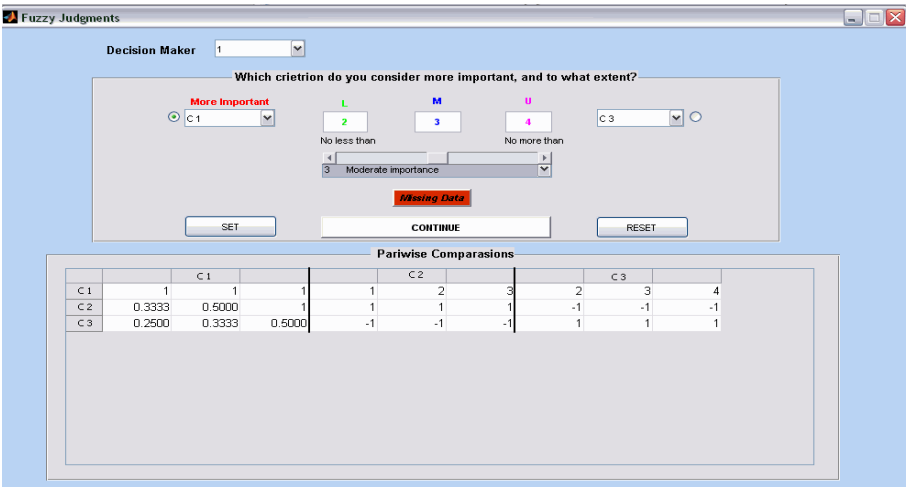


Fig. 2. The fuzzy comparison judgments window for the DM 1

Step 3: Developing the system based on the GUI functions. In this step, the MATLAB Graphical User Interface (GUI) functions are employed to develop a more user-friendly system.

6 Conclusions

This paper proposes a new method for solving fuzzy group prioritisation problems. The non-linear FPP is modified for group decision-making by introducing DMs' importance weights. The proposed method derives crisp priorities/weights from a set of fuzzy judgements and it does not require defuzzification procedures. Moreover, the proposed method is capable to derive crisp priorities from an incomplete set of DMs' fuzzy pairwise comparison judgments. The method is very efficient from a computational point of view, and a promising alternative to existing fuzzy group prioritization methods.

A Non-Linear FPP Solver is developed for solving group prioritization problems, which provide a user-friendly and efficient way to obtain the group priorities.

References

1. Choo, E., Wedley, W.: A common framework for deriving preference values from pairwise comparison matrices. Computers & Operations Research 31, 893–908 (2004)
2. Ittersum, K.V., Pennings, J.M., Wansink, B., van Trijp, H.C.: The validity of attribute-importance measurement: A review. Journa of Business Research 60, 1177–1190 (2007)

3. Edwards, W.: How to use multiattribute utility measurement for social decision making. IEEE Transactions on Systems, Man, and Cybernetics 7, 326–340 (1977)
4. Von Winterfeldt, D., Edwards, W.: Decision analysis and behavioural research. Cambridge University Press, Cambridge (1986)
5. Edwards, W., Barron, F.H.: SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement. Organizational Behavior and Human Decision Processes 60, 306–325 (1994)
6. Thurstone, L.: A law of comparative judgment. Psychological Review 34, 273–286 (1927)
7. Saaty, T.L.: The analytic hierarchy process: Planning, Priority setting, resource allocation. McGraw-Hill, NY (1980)
8. Saaty, T.L.: Decision making with dependence and feedback: the Analytic Network Process. RWA Publications, PA (1996)
9. Brans, J., Mareschal, B.: Promethee methods: In multiple criteria decision analysis: state of the art surveys. International Series in Operations Research & Management Science 78, 163–186 (2005)
10. Xu, D.L., Yang, J.B., Lama, J.: Group-based ER–AHP system for product project screening. Expert Systems with Applications 35, 1909–1929 (2008)
11. Chu, A., Kalaba, R., Springam, K.: A comparison of two methods for determining the weights of belonging to fuzzy sets. Journal of Optimization Theory and Application 27, 531–541 (1979)
12. Saaty, T.L., Vargas, L.G.: Comparison of Eigenvalue, Logarithmic Least Squares and Least Squares Methods in Estimating Rations. Mathematical Modelling 5, 309–324 (1984)
13. Mikhailov, L.: A Fuzzy Programming Method for Deriving Prioritises in Analytic Hierarchy Process. The Journal of Operational Research Society 51, 341–349 (2000)
14. Buckley, J.J.: Fuzzy Hierarchical Analysis. Fuzzy Sets and System 17, 233–247 (1984)
15. Laarhoven, P.J.M., Pedrycz, W.: A fuzzy extension of Saaty's priority theory. Fuzzy Sets and System 11, 229–241 (1983)
16. Chang, D.: Applications of the extent analysis method on fuzzy AHP. European Journal of Operational Research 95, 649–655 (1996)
17. Mikhailov, L.: Deriving priorities from fuzzy pair-wise comparison judgments. Fuzzy Sets and System 134, 365–385 (2003)
18. Tiwari, R.N., Dharmar, S., Rao, J.R.: Fuzzy goal programming-an additive model. Fuzzy Sets and System 24, 27–34 (1987)
19. Higham, D.J., Higham, N.J.: MATLAB Guide, 2nd edn. SIAM (2005)
20. Mikhailov, L., Didehkhani, H., Sadi-Nezhad, S.: Weighted Prioritization Models in the Fuzzy Analytic Hierarchy Process. International Journal of Information Technology & Decision Making 10, 681–694 (2011)

An Economic Production Quantity Problem with Fuzzy Backorder and Fuzzy Demand

József Mezei[1] and Kaj-Mikael Björk[1,2]

[1] IAMSR, Åbo Akademi University,
Joukahaisenkatu 3-5 A, FIN-20520 Turku
[2] Arcada University of Applied Sciences, Helsinki
{jmeze,kaj-mikael.bjork}@abo.fi

Abstract. Optimization models based on fuzzy set theory are relevant to the process industry, where there are many uncertainties that are inherently fuzzy. In this paper, we incorporate backorders (i.e the inventory to go below zero) and cycle time in a fuzzy Economic Production Quantity (EPQ) model. The uncertainties in the backorders and in the demand for different products are modeled using triangular possibility distributions. We illustrate the model with an example that describes a typical decision making problem in the paper industry.

Keywords: Economic Production Quantity, Fuzzy Sets, Signed Distance, Production optimization, Supply Chain Management.

1 Introduction

The contemporary business environment has increased the strategic importance of Decision Support Systems (DSS) in improving the efficiency of a company. In the optimal situation, the decision making steps and procedures and the supporting systems should fit the specific needs of the industry under consideration to identify the "optimal" (most efficient) decisions in different (business) processes instead of using some very generic models. In this paper, we will specifically focus on the specific features and needs of production-inventory problems. According to the APICS dictionary [1], process industries are businesses that add value to materials by mixing, separating, forming, or triggering chemical reactions.

Process industries account for a significant portion of the Gross Domestic Product (GDP) in most of the countries. For example in Finland, manufacturing (17.3 % of the GDP in 2011 [2]) is the second most important sector of the economy after services and the key sector considering foreign trades. The key branches [3] include pulp and paper industry (9.3 %), electronic products (9.8 %), and chemicals and chemical products (5.6 %).

Process industry in the Nordic countries is facing severe challenges. To meet up with the added logistics costs, the production efficiency needs to be extremely high. But not even that is enough and therefore, new ways of looking at the supply chain is needed. Production and distribution planning constitute very important task in the supply chain. Traditional methods, like the EOQ (Economic Order Quantity) model

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 557–566.
DOI: 10.1007/978-3-642-36981-0_51

first presented by [4] and its extensions still play an important role in the decision making processes in present days. On the other hand, many schemes, like the Bullwhip counter-acting models, cannot be introduced unaltered in the process industries [5,6].

Finding the answer to the "when" and "how much" questions in different industries is always a difficult task when the uncertainty present in the processes is significant. Uncertainties often stem from different aspects in both production and markets. These uncertainties can sometimes be captured with probabilistic measures, but quite often they derive from expert opinions in the production planning or marketing departments. Possibilistic measures and fuzzy EOQ-models provide an appropriate tool to handle these uncertainties in many applications [7].

The uncertainties in the production-inventory management decisions are today taken into consideration as a standard procedure. In the EOQ literature already [8] started to use probabilistic measures for EOQ models with backorders. However, the context often calls out for a possibilistic way of dealing with the uncertainties [9,10]. Chang et al. [11] were the first to solve a fuzzy EOQ model with backorders. They solved the model numerically, however. [9] and [12] solved the same model analytically using the signed distance defuzzification method [13]. Yao et al. [14] introduced an EOQ-model, without backorders, but for two replaceable merchandizes. [15] used the signed distance method for a fuzzy demand EOQ-model without backorders. In the references above, none account for a model, where the production rate is finite (i.e. models for producing entities). However [16] and [17] solved a fuzzy EPQ problem, where the entire cycle times are assumed to be fuzzy.

However, none of the above has allowed for backorders. [18] considered a EOQ problem with one item, backorders and fuzzy parameters. One of the first applications of fuzzy theory in EPQ models is [19], where the authors solve the EPQ model with fuzzy demand and production quantity. Bag et al. [20] introduce a model with inventory model with flexibility and reliability (of production process) consideration; the demand is modeled as a fuzzy random variable. The authors in [21] consider the EPQ problem with backorder with the setup cost, the holding cost and the backorder cost as fuzzy variables.

This paper presents a multi-item (one machine) EPQ (Economic Production Quantity) model with fuzzy backorders and demand values. It is an extension of the model in [17]. As analytical solution to the problem cannot be obtained, the optimal values are determined numerically. The rest of the paper is structured as follows. We describe the crisp and fuzzy models and check the convexity of the defuzzified version in Section 2. Section 3 provides an illustrative example using a typical decision making problem from paper industry. Finally, conclusions and future research directions are discussed in Section 4.

2 The Model

In this section, we are first going to present the crisp model, and the fuzzy model along with some basic assumptions and a suitable defuzzification strategy. Finally, we are going to solve the defuzzified optimization problem to global. The parameters and variables (can be assumed strictly greater than zero) in the classical multi-item EPQ

model with shared cycle time and backorders are the following (where the index $i \in I = \{1,2,...,|I|\}$ denotes the products):

- Q_i is the production batch size (variable)
- K_i is the fixed cost per production batch (parameter)
- D_i is the annual demand of the product (parameter)
- B_i is maximum shortage (just after a production run starts, variable)
- b_i is the unit shortage penalty cost per year (parameter)
- h_i is the unit holding cost per year (parameter)
- T is the cycle time (variable)
- P_i is the annual production rate (parameter)

The total cost function (TCU), including production setup costs and inventory holding costs for all products, is given by

$$TCU(Q_1,...,Q_n,B_1,...,B_n) = \sum_{i=1}^{n} \frac{K_i D_i}{Q_i} + \sum_{i=1}^{n} \frac{B_i^2 b_i}{2Q_i\rho_i} + \sum_{i=1}^{n} \frac{h_i(Q_i\rho_i - B_i)^2}{2Q_i\rho_i} \quad (1)$$

where $\rho_i = 1 - D_i/P_i$. The production batch size Q_i can also be replaced with the cycle time T according the formula $Q_i=TD_i$. The insertion of this formula into Eq. (1) yields the total cost function to minimize

$$TCU(T,B_1,...,B_n) = \sum_{i=1}^{n} \frac{K_i}{T} + \sum_{i=1}^{n} \frac{B_i^2(b_i+h_i)}{2TD_i\rho_i} + \sum_{i=1}^{n} \frac{Th_iD_i\rho_i}{2} - \sum_{i=1}^{n} B_i h_i \quad (2)$$

Eq. (2) is one version of the crisp (classical) multi-item EOQ-model with shared production capacity, cycle time and backorders. This problem can be solved using the derivatives, since all the terms in Eq. (2) are convex.

In order to present the fuzzy model, we will start by assuming that the cycle time is uncertain but it is possible to describe it with a triangular fuzzy number (symmetric).

Definition 1. Consider the fuzzy set $\tilde{A} = (a,b,c)$where $a < b < c$ and defined on R, which is called a triangular fuzzy number, if the membership function of $\tilde{A}$ is given by

$$\mu_{\tilde{A}}(x) = \begin{cases} \dfrac{x-a}{b-a} & a \le x \le b \\ \dfrac{c-x}{c-b} & b \le x \le c \\ 0 & otherwise \end{cases}$$

The fuzzy shortage $\tilde{B}_i$ and fuzzy demand $\tilde{D}_i$ will then be $\tilde{B}_i = (B_i - \Delta_i, B_i, B_i + \Delta_i)$ and $\tilde{D}_i = (D_i - \Lambda_i, D_i, D_i + \Lambda_i)$. In order to find non-fuzzy values for the model, we need to use some distance measures, and as in [11] we will use the signed distance [13].

Before the definition of this distance, we need to introduce the concept of α-cut of a fuzzy set.

Definition 2. Let $\tilde{B}$ be a fuzzy set on R and $0 \le \alpha \le 1$. The α-cut of $\tilde{B}$ is the set of all the points x such that $\mu_{\tilde{B}}(x) \ge \alpha$, i.e. $\tilde{B}(\alpha) = \{x | \mu_{\tilde{B}}(x) \ge \alpha\}$.

Let Ω be the family of all fuzzy sets $\tilde{B}$ defined on R for which the α-cut $\tilde{B}(\alpha) = \left[\tilde{B}_l(\alpha), \tilde{B}_u(\alpha)\right]$ exists for every $0 \le \alpha \le 1$, and both $\tilde{B}_l(\alpha)$ and $\tilde{B}_u(\alpha)$ are continuous functions on $\alpha \in [0,1]$.

Definition 3. For $\tilde{B} \in \Omega$ define the signed distance of $\tilde{B}$ to $\tilde{0}$ as

$$d(\tilde{B}, \tilde{0}) = \frac{1}{2} \int_0^1 \left[\tilde{B}_l(\alpha) + \tilde{B}_u(\alpha)\right] d\alpha$$

The Total Annual Cost in the fuzzy sense will be

$$TCU(T, \tilde{B}_1, ..., \tilde{B}_n) = \sum_{i=1}^{n} \frac{K_i}{T} + \sum_{i=1}^{n} \frac{\tilde{B}_i^2 (b_i + h_i)}{2T\tilde{D}_i \rho_i} + \sum_{i=1}^{n} \frac{Th_i \tilde{D}_i \rho_i}{2} - \sum_{i=1}^{n} \tilde{B}_i h_i \tag{3}$$

The signed distance between TCU and $\tilde{0}$ is given by

$$TCU(T, \tilde{B}_1, ..., \tilde{B}_n) = \sum_{i=1}^{n} \frac{K_i}{T} + \sum_{i=1}^{n} \frac{(b_i + h_i)}{2T\rho_i} d(\tilde{B}_i^2 / \tilde{D}_i, \tilde{0}) + \sum_{i=1}^{n} \frac{h_i T \rho_i}{2} d(\tilde{D}_i, \tilde{0}) - \sum_{i=1}^{n} h_i d(\tilde{B}_i, \tilde{0}) \tag{4}$$

If we calculate the signed distances, we obtain that

$$d(\tilde{B}_i, \tilde{0}) = \frac{1}{2} \int_0^1 \left[(\tilde{B}_i)_l(\alpha) + (\tilde{B}_i)_u(\alpha)\right] d\alpha = \frac{1}{2} \int_0^1 \left[(B_i - \Delta_i + \Delta_i \alpha) + (B_i + \Delta_i - \Delta_i \alpha)\right] d\alpha = B_i \tag{5}$$

$$d(\tilde{D}_i, \tilde{0}) = \frac{1}{2} \int_0^1 \left[(\tilde{D}_i)_l(\alpha) + (\tilde{D}_i)_u(\alpha)\right] d\alpha = \frac{1}{2} \int_0^1 \left[(D_i - \Lambda_i + \Lambda_i \alpha) + (D_i + \Lambda_i - \Lambda_i \alpha)\right] d\alpha = D_i \tag{6}$$

$$\begin{aligned} d(\tilde{B}_i^2 / \tilde{D}_i, \tilde{0}) &= \frac{1}{2} \int_0^1 \left[(\tilde{B}_i^2 / \tilde{D}_i)_l(\alpha) + (\tilde{B}_i^2 / \tilde{D}_i)_u(\alpha)\right] d\alpha = \frac{1}{2} \int_0^1 \left[\frac{(B_i - \Delta_i + \Delta_i \alpha)^2}{(D_i - \Lambda_i + \Lambda_i \alpha)} + \frac{(B_i + \Delta_i - \Delta_i \alpha)^2}{(D_i + \Lambda_i - \Lambda_i \alpha)}\right] d\alpha = \\ &- \frac{2B_i \Delta_i}{\Lambda_i} - \frac{D_i \Delta_i^2}{\Lambda_i^2} + \ln\left(\frac{D_i + \Lambda_i}{D_i - \Lambda_i}\right)\left[\frac{B_i^2 \Delta_i}{\Lambda_i} + \frac{B_i D_i \Delta_i}{\Lambda_i^2} + \frac{D_i^2 \Delta_i}{\Lambda_i^3}\right] \end{aligned} \tag{7}$$

The defuzzified total cost function is

$$\begin{aligned} TCU = &\sum_{i=1}^{n} \frac{K_i}{T} + \sum_{i=1}^{n} \frac{Th_i D_i \rho_i}{2} - \sum_{i=1}^{n} B_i h_i - \sum_{i=1}^{n} \frac{B_i \Delta_i (b_i + h_i)}{T \rho_i \Lambda_i} - \sum_{i=1}^{n} \frac{D_i \Delta_i^2 (b_i + h_i)}{2T \Lambda_i^2 \rho_i} \\ &+ \sum_{i=1}^{n} \frac{(b_i + h_i)}{2T\rho_i} \ln\left(\frac{D_i + \Lambda_i}{D_i - \Lambda_i}\right)\left[\frac{B_i^2 \Delta_i}{\Lambda_i} + \frac{B_i D_i \Delta_i}{\Lambda_i^2} + \frac{D_i^2 \Delta_i}{\Lambda_i^3}\right] \end{aligned} \tag{8}$$

To find the optimal solutions of this problem, first we have to examine the convexity of the defuzzified cost function. For this the Hessian matrix of the second derivatives needs to be computed. We will calculate the derivatives for a fixed *i*, to check under which conditions will be the terms in Eq. (8) convex. The Hessian matrix of the following function needs to be calculated:

$$f(T,B_i)=\frac{K_i}{T}+\frac{Th_iD_i\rho_i}{2}-B_ih_i-\frac{B_i\Delta_i(b_i+h_i)}{T\rho_i\Lambda_i}-\frac{D_i\Delta_i^2(b_i+h_i)}{2T\rho_i\Lambda_i^2}$$
$$+\frac{(b_i+h_i)}{2T\rho_i}\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\left[\frac{B_i^2\Delta_i}{\Lambda_i}+\frac{B_iD_i\Delta_i}{\Lambda_i^2}+\frac{D_i^2\Delta_i}{\Lambda_i^3}\right] \tag{9}$$

The partial derivatives are the following:

$$\frac{\partial f}{\partial T}=-\frac{K_i}{T^2}+\frac{h_iD_i\rho_i}{2}+\frac{B_i\Delta_i(b_i+h_i)}{T^2\rho_i\Lambda_i}+\frac{D_i\Delta_i^2(b_i+h_i)}{2T^2\rho_i\Lambda_i^2}$$
$$-\frac{(b_i+h_i)}{2T^2\rho_i}\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\left[\frac{B_i^2\Delta_i}{\Lambda_i}+\frac{B_iD_i\Delta_i}{\Lambda_i^2}+\frac{D_i^2\Delta_i}{\Lambda_i^3}\right] \tag{10}$$

$$\frac{\partial f}{\partial B_i}=-h_i-\frac{\Delta_i(b_i+h_i)}{T\rho_i\Lambda_i}+\frac{(b_i+h_i)}{2T\rho_i}\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\left[\frac{2B_i\Delta_i}{\Lambda_i}+\frac{D_i\Delta_i}{\Lambda_i^2}\right] \tag{11}$$

$$\frac{\partial^2 f}{\partial T^2}=\frac{2K_i}{T^3}-\frac{2B_i\Delta_i(b_i+h_i)}{T^3\rho_i\Lambda_i}-\frac{D_i\Delta_i^2(b_i+h_i)}{T^3\rho_i\Lambda_i^2}$$
$$+\frac{(b_i+h_i)}{T^3\rho_i}\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\left[\frac{B_i^2\Delta_i}{\Lambda_i}+\frac{B_iD_i\Delta_i}{\Lambda_i^2}+\frac{D_i^2\Delta_i}{\Lambda_i^3}\right] \tag{12}$$

$$\frac{\partial^2 f}{\partial B_i^2}=\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\left[\frac{\Delta_i(b_i+h_i)}{T\rho_i\Lambda_i}\right] \tag{13}$$

$$\frac{\partial^2 f}{\partial B_i\partial T}=\frac{\Delta_i(b_i+h_i)}{T^2\rho_i\Lambda_i}-\frac{B_i(b_i+h_i)}{T^2\rho_i\Lambda_i}\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)-\frac{D_i\Delta_i(b_i+h_i)}{2T^2\rho_i\Lambda_i^2}\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right) \tag{14}$$

In Eq. (13) all the terms are non-negative which implies that the first principal minor of the matrix is non-negative. To check the second determinant, we need to calculate

$$\frac{\partial^2 f}{\partial B_i^2}\frac{\partial^2 f}{\partial T^2}-\left(\frac{\partial^2 f}{\partial B_i\partial T}\right)^2$$

and check when this expression is non-negative:

$$\frac{\partial^2 f}{\partial B_i^2}\frac{\partial^2 f}{\partial T^2}=\left[\frac{\Delta_i^2(b_i+h_i)^2}{T^4\rho_i^2\Lambda_i^2}\right](\frac{2K_i\rho_i\Lambda_i}{\Delta_i(b_i+h_i)}\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)-2B_i\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)$$
$$-\frac{D_i\Delta_i}{\Lambda_i}\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)+B_i^2\left(\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\right)^2+\frac{D_iB_i}{\Lambda_i}\left(\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\right)^2+\frac{D_i^2\Delta_i}{\Lambda_i^2}\left(\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\right)^2) \quad (15)$$

and

$$\left(\frac{\partial^2 f}{\partial B_i\partial T}\right)^2=\left[\frac{\Delta_i^2(b_i+h_i)^2}{T^4\rho_i^2\Lambda_i^2}\right](1-2B_i\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)-\frac{D_i}{\Lambda_i}\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)$$
$$+B_i^2\left(\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\right)^2+\frac{D_iB_i}{\Lambda_i}\left(\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\right)^2+\frac{D_i^2}{4\Lambda_i^2}\left(\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\right)^2) \quad (16)$$

After some calculations one can obtain that the second determinant is non-negative if the following inequality holds:

$$\frac{2K_i\rho_i\Lambda_i}{\Delta_i(b_i+h_i)}\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)+\frac{D_i(1-\Delta_i)}{\Lambda_i}\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)+\frac{D_i^2(\Delta_i-0.25)}{\Lambda_i^2}\left(\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\right)^2\geq 1 \quad (17)$$

Since this formula depends only on the parameters, before applying the model, one needs to check whether Eq. (17) is satisfied for every $i=1,...,n$. Therefore, to obtain the minimum of Eq. (9), the system of *(n+1)* equations to be solved is the following:

$$\frac{\partial TCU}{\partial T}=0, \qquad \frac{\partial TCU}{\partial B_i}=0, \forall i=1,...,n.$$

If we look at the derivatives, we can formulate the equations for T:

$$\frac{\partial TCU}{\partial T}=-\sum_{i=1}^{n}\frac{K_i}{T^2}+\sum_{i=1}^{n}\frac{h_iD_i\rho_i}{2}+\sum_{i=1}^{n}\frac{B_i\Delta_i(b_i+h_i)}{T^2\rho_i\Lambda_i}+\sum_{i=1}^{n}\frac{D_i\Delta_i^2(b_i+h_i)}{2T^2\rho_i\Lambda_i^2}$$
$$-\sum_{i=1}^{n}\frac{(b_i+h_i)}{2T^2\rho_i}\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\left[\frac{B_i^2\Delta_i}{\Lambda_i}+\frac{B_iD_i\Delta_i}{\Lambda_i^2}+\frac{D_i^2\Delta_i}{\Lambda_i^3}\right] \quad (18)$$

and for $B_i, \forall i=1,...,n$:

$$\frac{\partial TCU}{\partial B_i}=-h_i-\frac{\Delta_i(b_i+h_i)}{T\rho_i\Lambda_i}+\frac{(b_i+h_i)}{2T\rho_i}\ln\left(\frac{D_i+\Lambda_i}{D_i-\Lambda_i}\right)\left[\frac{2B_i\Delta_i}{\Lambda_i}+\frac{D_i\Delta_i}{\Lambda_i^2}\right] \quad (19)$$

Hence the (numerical) solution of this system of equations will give us the optimal value of T and $B_i, \forall i=1,...,n$:. In the next section, we will calculate the solution for a numerical example.

3 Numerical Example

In this section we illustrate the model with a numerical example. This problem is a fictive one, even if the numbers are in the likely range of a real Finnish paper producer. The problem consists of 8 products, where the an annual demand (D_i) of products are 1200 tons, 1100 tons, 1600 tons, 1100 tons, 1000 tons, 1500 tons, 1200 tons and 1700 tons respectively. The production rates (P_i) are 2900, 2700, 2400, 2800, 3000, 2500, 2100, 2500 tons / year. There is a fixed cost incurring each time a product starts to be produced (setup costs, K_i): 1200, 1100, 800, 1500, 700, 1200, 900, and 1300 euro respectively. The holding costs are 0.8 euro per kg and annum for each product and the unit shortage costs are 1.5 euro per kg and annum for each product. The Δ_i-parameters in the fuzzy case are assumed to be 5% for every product and the Λ_i parameters representing the uncertainty in the demand are specified in this example as 10\% of the demand values. All the requirements for the problem to be convex according to Eq. (17) are satisfied. The optimal solutions for the crisp and fuzzy case are given in Table 1 (the optimal values are obtained using Excel Solver optimization package).

Table 1. Results for the test example

	Crisp	Fuzzy
T	**2,6138**	**2,2093**
B_1	**639,51**	**389,43**
B_2	**592,64**	**382,33**
B_3	**484,89**	**364,93**
B_4	**607,20**	**384,56**
B_5	**606,12**	**384,39**
B_6	**545,43**	**374,94**
B_7	**467,59**	**361,94**
B_8	**494,57**	**366,58**
TCU	**6656,95**	**7500,40**

From Table 1, it can be seen that the cycle time decreases from 2,61 to 2,21 days if the possibility distribution in the cycle time is correctly accounted for. The optimal total cost is increased from 6656,95 to 7500,40 (12,7% increase). Each of the backorders decreased significantly. The increase in the total cost comes mainly from the smaller batch sizes and the uncertainty incorporated in the demand and the backorders.

To analyze the results through sensitivity analysis, we have to examine the optimal solutions and the value of the total cost function for different initial values of the parameters. In this example two important parameters were investigated:

- the b_i values (the unit shortage penalty cost),
- the h_i values (the unit holding cost).

As for the unit shortage penalty costs, the results are listed in Table 2. The results show that the total cost value increases as we increase the unit shortage penalty cost as it is expected, but the difference in the values is not significant. On the other hand we can observe that the optimal cycle time decreases with the penalty cost: if the penalty cost increases, the optimal decision is to avoid shortages as efficiently as possible; this can be achieved by changing the products in the machines more frequently.

The results of the simple sensitivity analysis concerning the unit holding cost can be found in Table 3. As it is clear from the comparison of the crisp and the fuzzy optimal solution and total cost function, if we increase the holding cost in the Economic Production Quantity model, the total cost value is increased significantly more than for the shortage cost. The results in Table 3 reflect this observation: the total cost is an increasing function of the imprecision. On the other hand, the cycle time is much shorter: because of the high holding cost, it is very expensive to produce extra products.

Table 2. The results for different values b_i

	b_i	T	TCU
1	3	4,9848	4965,42
2	2,5	4,6808	4820,49
3	2	4,4157	4696,92
4	1	4,0339	4523,96
5	0,5	3,9324	4479,04

Table 3. The results for different values of h_i

	h_i	T	TCU
1	1,20	4,6960	4801,23
2	1,22	4,6897	4809,09
3	1,25	4,6808	4820,49
4	1,28	4,6722	4831,42
5	1,30	4,6667	4838,47

4 Summary and Future Research Directions

In Supply Chain Management (SCM), track of research that tries to increase collaboration in these supply chains (in order to reduce the Bullwhip effect, for instance), there is a track to create fundamental and generic models for each part of the supply chain. For this track of research, the uncertainties involved should not be neglected. Some of these uncertainties are captured through possibilistic measures based on expert opinions. A promising research direction in this context has been the fuzzy EPQ (Economic Production Quantity) development.

This paper contributes to the track of fuzzy EPQ-theory with a model that takes fuzzy backorders and demands together with cycle times into consideration along with multi-item EPQ (Economic Production Quantity) features. The uncertainty in backorders and demands originates from the imprecise information about different decision process components. The fuzzy model was defuzzified using the signed distance method. The system of equations needed to be solved numerically. The problem was illustrated with a small test example along with some simple sensitivity analysis. This analysis showed that results from the uncertainties in the demand have significant impact on the overall annual cost function value. Future research directions will introduce different types of fuzzy numbers and defuzzification methods in the model.

References

1. APICS Dictionary, 8th ed., American Production and Inventory Control Society, Falls Church, VA (1995)
2. Finland in Figures - National Accounts, Statistics Finland, http://www.stat.fi/tup/suoluk/suoluk_kansantalous_en.html (retrieved September 18, 2012)
3. Finland in Figures - Manufacturing, Statistics Finland, http://www.stat.fi/tup/suoluk/suoluk_teollisuus_en.html (retrieved September 18, 2012)
4. Harris, F.W.: How many parts to make at once. The Magazine of Management 10, 135–136 (1913)
5. Björk, K.-M.: Supply Chain Efficiency with Some Forest Industry Improvements. Doctoral dissertation (Business Administration), Abo Akademi University (2006) ISBN 952-12-1782-0
6. Björk, K.-M., Carlsson, C.: The Effect of Flexible Lead Times on a Paper Producer. International Journal of Production Economics 107(1), 139–150 (2007)
7. Zadeh, L.A.: Outline of a new Approach to the Analysis of Complex Systems and Decision Processes. IEEE Transactions on Systems, Man and Cybernetics, SMC-3, 28–44 (1973)
8. Liberatore, M.J.: The EOQ model under stochastic lead time. Operations Research 27, 391–396 (1979)
9. Björk, K.-M., Carlsson, C.: The Outcome of Imprecise Lead Times on the Distributors. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS 2005), pp. 81–90 (2005)

10. Carlsson, C., Fullér, R.: Soft computing and the Bullwhip effect. Economics and Complexity 2(3), 1–26 (1999)
11. Chang, S.-C., Yao, J.-S., Lee, H.-M.: Economic reorder point for fuzzy backorder quantity. European Journal of Operational Research 109, 183–202 (1998)
12. Björk, K.-M.: An Analytical Solution to a Fuzzy Economic Order Quantity Problem. International Journal of Approximate Reasoning 50, 485–493 (2009)
13. Yao, J.S., Wu, K.: Ranking fuzzy numbers based on decomposition principle and signed distance. Fuzzy Sets and Systems 116, 275–288 (2000)
14. Yao, J.-S., Ouyang, L.-Y., Chiang, J.: Models for a fuzzy inventory of two replaceable merchandises without backorders based on the signed distance of fuzzy sets. European Journal of Operational Research 150, 601–616 (2003)
15. Yao, J.S., Chiang, J.: Inventory without backorder with fuzzy total cost and fuzzy storing cost defuzzified by centroid and signed distance. European Journal of Operational Research 148, 401–409 (2003)
16. Björk, K.-M.: The economic production quantity problem with a finite production rate and fuzzy cycle time. In: 41th Annual Hawaii International Conference on System Sciences, HICSS 2008 (2008)
17. Björk, K.-M.: A Multi-item Fuzzy Economic Production Quantity Problem with a Finite Production Rate. International Journal of Production Economics 135(2), 702–707 (2011)
18. Kazemi, N., Ehsani, E., Jaber, M.Y.: An inventory model with backorders with fuzzy parameters and decision variables. International Journal of Approximate Reasoning 51, 964–972 (2012)
19. Lee, H.M., Yao, J.S.: Economic production quantity for fuzzy demand quantity, and fuzzy production quantity. European Journal of Operational Research 109, 203–211 (1998)
20. Bag, S., Chakraborty, D., Roy, A.R.: A production inventory model with fuzzy random demand and with flexibility and reliability considerations. Computers & Industrial Engineering 56, 411–416 (2009)
21. Wang, X., Tang, W.: Fuzzy EPQ inventory models with backorder. Journal of Systems Science and Complexity 22, 313–323 (2009)

Analyzing Website Content for Improved R&T Collaboration Planning

Dirk Thorleuchter[1] and Dirk Van den Poel[2]

[1] Fraunhofer INT, Appelsgarten 2, D-53879 Euskirchen, Germany
Dirk.Thorleuchter@int.fraunhofer.de
[2] Ghent University, Faculty of Economics and Business Administration, Tweekerkenstraat 2, B-9000 Gent, Belgium
Dirk.Vandenpoel@ugent.be

Abstract. A well-known problem in research and technology (R&T) planning is the selection of suited R&T collaboration partners. We investigate the use of textual information from the website content of possible collaboration candidates to identify their suitability. This improves the selection of collaboration partners and it enables a successful processing of R&T-projects. In a case study 'defense R&T', organizations and companies that have proven their suitability as collaboration partner in former R&T projects are selected (positive examples) as well as organizations and companies that have not. Latent semantic indexing with singular value decomposition and logistic regression modeling is used to identify semantic textual patterns from their websites' content. As a result of prediction modeling, some of these textual patterns are successful in predicting new organizations or companies as (un-) suited R&T collaboration partners. These results support the acquisition of new collaboration partners and thus, they are valuable for the planning of R&T.

Keywords: Collaboration, Research, Technology, Semantic Classification, Text Mining, Defense.

1 Introduction

At present, projects in research and technology (R&T) are often complex [1] and their successful processing requires the collaboration with external partners (organizations or companies) [2]. Based on partners' capabilities, organizations or companies can be classified as suited or unsuited collaboration partners [3]. Thus, an important aspect for the R&T planning of an organization is the selection of suited collaboration partners to enable a successful processing of R&T-projects [4].

Normally, the selection of suited collaboration partners is done by considering gained experience of the organization where the potential candidates have already proven their suitability as collaboration partners in former R&T-projects [5]. This often leads to the selection of already known partners and this excludes the selection of new and unknown partners for future projects [6].

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 567–573.
DOI: 10.1007/978-3-642-36981-0_52

Literature introduces the use of textual information from companies' websites to predict the success of companies in business-to-consumer (B2C) environment [7,8]. Recent research has shown that this information also can be used to predict the success of business-to-business (B2B) commerce transactions [9,10]. Whereas R&T collaborations represent specific B2B commerce transactions, textual website content possibly can be used to predict suited R&T collaboration partners.

Thus, we propose a new methodology that creates such a prediction model to show its success for this specific B2B area. In detail, we investigate textual website content of suited R&T collaboration partners that have proven their suitability in current or former projects (positive examples). We also investigate textual website content of organizations and companies that have not involved in R&T projects in present or past (negative examples). Semantic textual patterns are extracted from the content and prediction modeling is applied.

A case study shows that specific textual patterns can be used for a successful assigning of defense based R&T organizations and companies to the positive or negative examples. As a result, some textual patterns represent existing success factors as known from e-commerce literature while others represent new success factors that specifically can be used to predict R&T collaboration partners. Thus, the created prediction model supports research planners by acquiring new R&T collaboration partners.

2 Methodology

This new methodology consists of several steps as depicted in Fig. 1. The first step is to create two lists of organizations and companies as suited or as unsuited collaboration partners based on experiences of the past. Elements of these lists are divided in training set and test set.

In a second step, the websites standing behind the organizations and companies are identified and textual information from websites' content is crawled. Crawling is done based on a web content mining approach because each website consists of several web pages and only some of them are relevant [11]. Trivial web pages e.g. 'disclaimer' or 'sitemap' are discarded. The identification of relevant web pages is done by selecting the starting page of a website as well as by selecting the four most frequently visited web pages. Whereas information about page visitors is not available, we use results from Google ranking algorithm as indicator for highly visited web pages where high ranked web pages are selected [12]. Additionally, it is supposed that website information about R&T activities might be relevant for identifying the organization or company as suited collaboration partner. Thus, web pages where specific terms ('research', 'technology', 'high-tech', 'science' etc.) occur are selected.

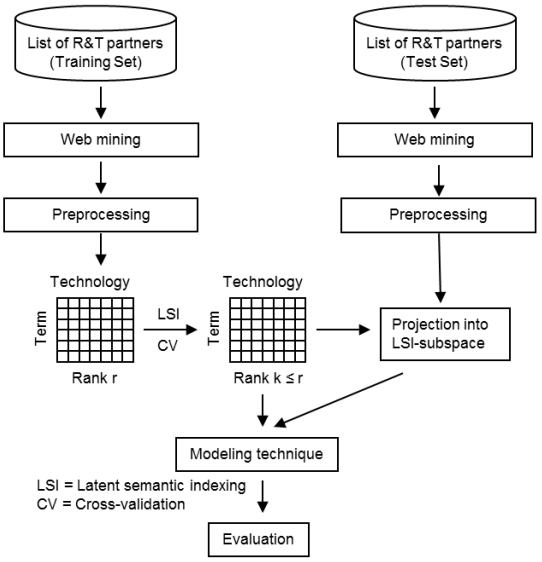


Fig. 1. Processing of the approach

The third step prepares the information crawled from the selected web pages to obtain a specific granularity [13-17]. Typographical errors are corrected, tokenization is used, and all terms are converted in lower case [18]. Further, part-of-speech tagging, stop word filtering, and stemming is applied as well as Zipf distribution [19]. A term vector for each web page is created based on vector space model [20]. Vectors are aggregated to build a vector for each website. Weighted term frequencies are calculated for the vector components because they lead to a significant improvement [21].

The weight of a term is at its maximum if the term occurs very frequently in a small number of websites and if the term does not occur in the other websites. The weights are calculated by multiplying term frequency and inverse object frequency and by dividing it with a length normalization factor [22]. As a result, a term-by-website matrix is created based on the vectors. The matrix consists of high dimensionality and many components of the matrix are zero.

In the fourth step, latent semantic indexing (LSI) is used to reduce the high dimensionality [23]. This summarizes textual patterns concerning their aspect of meaning (semantic textual patterns). Dimensionality reduction can be done by matrix factorization techniques. Thus, we use singular value decomposition (SVD) as the commonly used matrix factorization technique. The term-by-website matrix A is a (m x n) matrix where m equals the number of terms and n the number of websites. The rank r of matrix A is smaller than min(m,n). To reduce the rank r of the term-by-website matrix A to k, SVD splits the matrix in the product of three matrices: a) a term-dimension (m x r) matrix U that shows the impact of each term on the semantic dimensions, b) a website-dimension (n x r) matrix V that shows the impact of each website on the semantic dimensions, and c) a positive singular values (r x r) matrix Σ where the singular values are positioned on the diagonal.

$$A = U \Sigma Vt \quad (1)$$

The rank r of the term-by-website matrix to k can be reduced by selecting the first k singular values and by discarding further singular values.

The value of k is selected based on an optimal predictive performance of the semantic textual patterns. Several rank-k models are constructed using a parameter-selection procedure based on backward selection as main approach in stepwise regression [24,25]. For each rank-k model, a fivefold cross-validation is applied on the training set. We use logistic regression [26] as predictive modeling technique where a maximum likelihood function is produced and maximized. As performance measure, the area under the receiver operating characteristics (ROC) curve (AUC) is used [27-29]. To obtain the optimal number of k, the predictive performance of the model as calculated by the cross-validated AUC is optimized [30].

The fifth step applies prediction modeling on the test set to measure the performance of the approach. Logistic regression is used to identify textual patterns that are characteristic for both the positive and the negative examples. These textual patterns represent success factors and they are compared to existing website success factors known from literature [12]. As a result, existing success factors from literature can be identified and it can be shown that they are also successful in predicting R&T collaboration partners. Further, success factors that are not mentioned in literature can be identified to introduce them as new success factors to the scientific community. For the evaluation, commonly used evaluation criteria are used in a last step: the cumulated lift [31], the precision and recall [32], the cross-validated AUC, as well as the sensitivity and specificity [33].

3 Case Study

In a case study 'Defense R&T', organizations and companies are identified that have been involved as collaboration partners in R&T projects funded by German Ministry of Defense [34,35]. Whereas a participation in a project is always agreed by manual evaluation of research planners, they are assigned to the positive examples. A large number of further organizations and companies with no relation to defense R&T projects are taken over from an existing study in literature [9] as negative examples. The aim of the case study is to identify semantic textual patterns that can be used to predict new (unseen) organizations and companies as member of positive examples. To prevent language translation problems, the case study is restricted to websites in German language. 5.315 positive examples are identified and they are split in 3.720 training examples and 1.595 test examples. 24.897 negative examples are selected and split in 17.428 training examples and in 7.469 test examples. The relative percentage is 17,5% (positive examples) to 82,5% (negative example).

An optimal predictive performance was reached by setting k to 24 dimensions. Four dimensions can be identified as classifier for the positive examples while two dimensions have large impact on the negative examples. Each dimension consists of several textual patterns and thus, it might represent one or several success factors.

Based on a comparing of the identified success factors from the seven dimensions to the success factors from e-commerce literature, we have found three factors that are successful to predict the positive examples (the trustworthiness, reliability and security') and two factors that are successful to predict the negative examples (the website interactivity and the technical support of a website).

We also have identified the occurrence of further success factors (High quality of website content, Usefulness of website content, Website Usability, Website responsiveness, Wide product choice, Website customization, Website promotion initiatives, Addressing emotions, and Website design). However they cannot be used to predict the positive or the negative examples.

Last, further factors have been identified from the dimensions that are not mentioned in literature before. These factors are also successful in predicting the positive examples: These factors can be labeled by 'Data protection policy', 'Conflict monitoring', 'International collaboration', 'Emergency response', and 'Demonstrators und Prototypes'. This is because these factors are of particular interest in defense R&T context.

4 Evaluation

The methodology is applied. Based on the results of the test set, an evaluation is done by use of well-known evaluation criteria: the lift, the precision and recall, the ROC curve that is based on sensitivity and specificity, and the AUC. The results from the test set are compared to the frequent baseline.

The cumulated lift lays above the frequent baseline and it shows that the test set is able to identify more suited collaboration partners than the random baseline within a specific percentile, e.g. the cumulated lift value in the top 10 (30) percentile increases from one to 1.24 (1.13). In the precision and recall diagram, the test set outperforms the baseline at a recall greater than 40 %. The ROC curve of the test set lies above the random baseline and the cross validated AUC of the test set (0.6245) is larger than the baseline (0.5000). This improvement is significant (χ^2=0.02, d.f.=1, $p<0.001$). Overall, the evaluation shows that the proposed methodology outperforms the baseline.

5 Conclusion

The results show that using information from websites' content helps research planners to identify suited R&T collaboration partners with a higher precision. Thus, the proposed methodology can support decision makers to improved research planning. It is shown that some e-commerce success factors are also successful in R&T collaboration planning and that some e-commerce success factors are not successful. Further, new success factors are identified and proposed to the scientific literature.

The AUC of the test set (62%) shows that this approach should not be used alone as predictive model for a collaboration decision. However, the identification process can become more targeted by additionally integrating this information in the decision process as one variable among others.

References

1. Thorleuchter, D., Van den Poel, D.: Technology classification with latent semantic indexing. Expert. Syst. Appl. 40(5), 1786–1795 (2013)
2. Bammer, G.: Enhancing research collaborations: Three key management challenges. Res. Pol. 37(5), 875–887 (2008)
3. Heinze, T., Kuhlmann, S.: Across institutional boundaries?: Research collaboration in German public sector nanoscience. Res. Policy 37(5), 888–899 (2008)
4. Thorleuchter, D., Van den Poel, D.: Semantic Technology Classification. In: International Conference on Uncertainty Reasoning and Knowledge Engineering, pp. 36–39. IEEE Press, New York (2011)
5. Van Rijnsoever, F.J., Hessels, L.K.: Factors associated with disciplinary and interdisciplinary research collaboration. Res. Policy 40(3), 463–472 (2011)
6. He, Z.L., Geng, X.S., Campbell-Hunt, C.: Research collaboration and research output: A longitudinal study of 65 biomedical scientists in a New Zealand university. Res. Policy 38(2), 306–317 (2009)
7. Ballantine, J., Levy, M., Powell, P.: Evaluating information systems in small and medium-sized enterprises: issues and evidence. Eur. J. Inform. Syst. 7, 241–251 (1998)
8. Serafeimidis, V., Smithson, S.: Information systems evaluation as an organizational institution - experience from a case study. Inform. Syst. J. 13, 251–274 (2003)
9. Thorleuchter, D., Van den Poel, D., Prinzie, A.: Analyzing existing customers' websites to improve the customer acquisition process as well as the profitability prediction in B-to-B marketing. Expert. Syst. Appl. 39(3), 2597–2605 (2012)
10. Thorleuchter, D., Van den Poel, D.: Using Webcrawling of Publicly-Available Websites to Assess E-Commerce Relationships. In: Service Research and Innovation Institute 2011 (SRII 2011), pp. 402–410. IEEE Press, New York (2012)
11. Lihui, C., Lian, C.W.: Using Web structure and summarisation techniques for Web content mining. Inform. Process. Manag. 41(5), 1225–1242 (2005)
12. Thorleuchter, D., Van den Poel, D.: Predicting e-commerce company success by mining the text of its publicly-accessible website. Expert. Syst. Appl. 39(17), 13026–13034 (2012)
13. Thorleuchter, D., Weck, G., Van den Poel, D.: Granular Deleting in Multi Level Security Models – An Electronic Engineering Approach. In: Jin, D., Lin, S. (eds.) Advances in Mechanical and Electronic Engineering. LNEE, vol. 177, pp. 609–614. Springer, Heidelberg (2012)
14. Thorleuchter, D., Weck, G., Van den Poel, D.: Usability Based Modeling for Advanced IT-Security – An Electronic Engineering Approach. In: Jin, D., Lin, S. (eds.) Advances in Mechanical and Electronic Engineering. LNEE, vol. 177, pp. 615–619. Springer, Heidelberg (2012)
15. Thorleuchter, D., Van den Poel, D.: High Granular Multi-Level-Security Model for Improved Usability. In: 2nd International Conference on System Science, Engineering Design and Manufacturing Informatization, pp. 191–194. IEEE Press, New York (2011)
16. Gericke, W., Thorleuchter, D., Weck, G., Reilaender, F., Loss, D.: Vertrauliche Verarbeitung staatlich eingestufter Information - die Informationstechnologie im Geheimschutz. Informatik Spektrum 32(2), 102–109 (2009)
17. Thorleuchter, D., Van den Poel, D.: Rapid Scenario Generation with Generic Systems. In: Management Sciences and Information Technology. Lecture Notes in Information Technology, IERI, Delaware (2012) (in press)
18. Zheng, H.T., Kang, B.Y., Kim, H.G.: Exploiting noun phrases and semantic relationships for text document clustering. Inform. Sciences 179(13), 2249–2262 (2009)

19. Thorleuchter, D., Schulze, J., Van den Poel, D.: Improved Emergency Management by a Loosely Coupled Logistic System. In: Aschenbruck, N., Martini, P., Meier, M., Tölle, J. (eds.) Future Security 2012. CCIS, vol. 318, pp. 5–8. Springer, Heidelberg (2012)
20. Thorleuchter, D., Van den Poel, D.: Companies Website Optimising concerning Consumer's searching for new Products. In: International Conference on Uncertainty Reasoning and Knowledge Engineering, pp. 40–43. IEEE Press, New York (2011)
21. Sparck Jones, K.: Collection properties influencing automatic term classification performance. Inform. Storage Retrieval 9(9), 499–513 (1973)
22. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Inform. Process. Manag. 24(5), 513–523 (1988)
23. Zhang, W., Yoshida, T., Tang, X.: A comparative study of TF*IDF, LSI and multi-words for text classification. Expert. Syst. Appl. 38(3), 2758–2765 (2011)
24. Thorleuchter, D., Van den Poel, D.: Extraction of Ideas from Microsystems Technology. In: Jin, D., Lin, S. (eds.) Advances in CSIE, Vol. 1. AISC, vol. 168, pp. 563–568. Springer, Heidelberg (2012)
25. Thorleuchter, D., Herberz, S., Van den Poel, D.: Mining Social Behavior Ideas of Przewalski Horses. In: Wu, Y. (ed.) Advances in Computer, Communication, Control and Automation. LNEE, vol. 121, pp. 649–656. Springer, Heidelberg (2011)
26. D'Haen, J., Van den Poel, D., Thorleuchter, D.: Predicting Customer Profitability During Acquisition: Finding the Optimal Combination of Data Source and Data Mining Technique. Expert. Syst. Appl. (in press, 2013), doi:10.1016/j.eswa.2012.10.023
27. Chen, M.Y., Chu, H.C., Chen, Y.M.: Developing a semantic-enable information retrieval mechanism. Expert. Syst. Appl. 37(1), 322–340 (2010)
28. DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L.: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44(3), 837–845 (1988)
29. Hanley, J.A., McNeil, B.J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 131(1), 29–36 (1982)
30. Thorleuchter, D., Van den Poel, D.: Using NMF for Analyzing War Logs. In: Aschenbruck, N., Martini, P., Meier, M., Tölle, J. (eds.) Future Security 2012. CCIS, vol. 318, pp. 73–76. Springer, Heidelberg (2012)
31. Thorleuchter, D., Van den Poel, D.: Protecting Research and Technology from Espionage. Expert. Syst. Appl. (in press, 2013), doi:10.1016/j.eswa.2012.12.051
32. Thorleuchter, D., Van den Poel, D., Prinzie, A.: Mining Ideas from Textual Information. Expert. Syst. Appl. 37(10), 7182–7188 (2010)
33. Thorleuchter, D., Van den Poel, D.: Improved Multilevel Security with Latent Semantic Indexing. Expert. Syst. Appl. 39(18), 13462–13471 (2012)
34. Thorleuchter, D., Van den Poel, D., Prinzie, A.: A compared R&D-based and patent-based cross impact analysis for identifying relationships between technologies. Technol. Forecast. Soc. Change 77(7), 1037–1050 (2010)
35. Thorleuchter, D., Van den Poel, D.: Web Mining based Extraction of Problem Solution Ideas. Expert. Syst. Appl. (in press, 2013)

Building Accountability for Decision-Making into Cognitive Systems

Victoria L. Lemieux and Thomas Dang

iSchool@UBC 4th Floor, I.K. Barber Learning Centre, University of British Columbia,
1961 East Mall, Vancouver, B.C., Canada, V6S 0B6
vlemieux@mail.ubc.ca

Abstract. This paper lays out a theoretical framework for engineering accountability for decision-making into cognitive systems, based on a combination of theories from the fields of archival and cognitive science, and demonstrates the application of this framework in a prototype cognitive system - an interactive visual dashboard for fixed income analytics.

Keywords: cognitive systems engineering, cognitive science, archival science, records, accountability.

1 Introduction and Motivation

John E. Kelly of IBM has observed that, today, we are at the dawn of another epochal era in the development of technology – the era of cognitive systems [1]. He argues that the changes coming in the next 10 to 20 years will transform how we live, work and learn, just as programmable computing did previously [1]. What are cognitive systems and how might this era of cognitive systems unfold? Cognitive systems encompass a range of technologies, including artificial intelligence, expert systems, and human machine interaction interfaces and are defined as systems that are capable of learning from their interactions with data and humans [1]. These systems afford mastery of the tasks on which they work, as well as extraction of contextual information from the environment in which these tasks are undertaken. They solve problems as they arise and plan for the future. They communicate appropriately with others about themselves and their activities in order to work effectively in close collaboration. They also adapt their understanding and skills as they and the world around them evolve [2]. Traditional computers are organized around microprocessors. With cognitive systems, it is much more about the data and drawing insights from it through analytics [1].

Though the ultimate goal may be to create intelligent machines, the next logical step in the evolution of cognitive systems is to augment human cognitive capabilities by developing machines capable of offloading human thought processes and actively supporting individuals in pursuing their goals. In this sense, the focus is less on creating artificial intelligence (AI) than on augmenting human intelligence (IA). Such cognitive systems serve the individual by reducing their cognitive supervisory burden. "They enhance the individual's cognitive abilities by supplementing memory and

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 575–586.
DOI: 10.1007/978-3-642-36981-0_53

problem-solving capabilities and by providing direct access to relevant data, expertise, guidance, and instruction. They work towards shared goals while understanding enough about the task, the individual, and each other to assist, mentor, cooperate, and monitor as needed. And they reduce the individual's performance degradation by offloading activities and by anticipating the kinds of errors that tend to occur in stressful situations" [2].

Visual Analytics (VA) can be classified as an IA type of cognitive system. VA is defined as "the science of analytical reasoning facilitated by interactive visual interfaces" [3]. VA harnesses the power of human visual perception and cognition, lending truth to the old adage that "a picture is worth a thousand words." Visualization has advantages over other modes of communication and thought because humans have evolved visual and spatial skills that include the ability to detect edges and discontinuities, things that stand out, variations in color and shape, and motion; to recognize patterns; and to retrieve information using visual cues [4] [5]. Each of these visual and spatial attributes can be transformed into a graphical image to provide a rich visual description of data. As these features can be observed with 'pre-attentive processing'; that is, they are perceived prior to conscious attention, they are understandable at a glance and much more rapidly than words [6]. Through encoding of data into graphical images, visualizations can act as a repository of data which allows individuals to offload cognition to the perceptual system, using visuals as a form of virtual memory [7]. This can enlarge problem-solving capabilities by enabling the processing of more data without overloading the decision maker [8]. Finally, because visual cues stand out to human perception more than words, a picture sometimes "forces us to notice what we never expected to see" thereby allowing for greater insight [9].

VA systems, and other cognitive systems of this type, often serve as tactical decision aids. Such systems are generally referred to as Command and Control Systems (C2) [10]. Initially limited to military operations and emergency response, cognitive systems that serve as decision aids are increasingly moving into other areas as well. Many futurists forecast the emergence of cognitive assistants that will allow for bundling of work and priority mechanisms and will act as a buffer between ever-increasing content and the needs of workers to arrange their knowledge and tasks [11]. In the future, virtual representatives – avatars – may become central to the way virtual working occurs. How will this affect accountability for decision-making? You can hear it now: it wasn't me officer, it was my avatar! In light of these changes it is critical that we consider how to build accountability for decision-making into cognitive systems. This paper considers the problem by first laying out a theoretical framework for building accountability into cognitive systems, and then considering the application of that framework in the context of design of a VA system for fixed income analytics.

2 Theoretical Framework

We begin by stating that we accept the fundamental hypothesis of cognitive science, that thinking is produced by computational procedures operating on mental processes [12]. Reasoning is the term used to describe the mental process by which agents

(people or machines) use information to derive new information [13]. Decision-making may be defined as the mental process concerned with how agents' beliefs and values determine decisions [13, 14], while a decision may be defined as the act of making up one's mind about something or a position reached after consideration [14]. While expected utility theory is widely accepted as a normative framework for how decisions ought to be made, there are many theories used to explain how decisions are actually made [13]. A full discussion of these is beyond the scope of this paper.

Decision-making and decisions, in the context of C2 systems, lead to actions, many with significant consequences for human life and safety. Traditionally, accountability for decision-making, decisions, and their resultant actions, occurs on an *ex post facto* basis. Generally, the operation of accountability relies on some evidence of the decision–making and of the actions that resulted from those decisions [15]. In highly formal accountability frameworks, such as those developed for financial accountability, such evidence is typically found in the form of accounts [16]. In the case of financial decision-making, for example, decisions, and their resulting financial transactions, will be represented in the form of financial accounts. Accounts provide summaries of transactions undertaken in a particular reporting period (e.g., a financial quarter or year). Such transactions are captured as records, with each individual record of a transaction being aggregated to report on accounting activity. Trustworthy accounts, that is, those that can be trusted as reliably accounting for decisions and their related actions depend upon 1) the creation and preservation of trustworthy records documenting each underlying transaction and 2) the reliable aggregation of those records into summary reports, or accounts. Where accountability is less formal or in situations where accountability is sought for a particular action (e.g., who made the decision not to replace the part that led to a catastrophic oil spill), accountability will rely less on summary accounts and more on records that provide evidence which directly relates to the action under scrutiny. Thus, the creation and preservation of *records*, and moreover, records that can be *trusted*, is a critical element of the operation of accountability in all of its forms.

To explore what is needed for the creation and preservation of trustworthy records, it is necessary to first consider the concept of 'record'. Many archival theorists have written on the nature of records, but three characteristics standout as consistent in the literature and discussions about records: 1) that records are linked to transactions, 2) that records are linked to the notion of evidence, 3) that context is important to interpreting the meaning of records.

Luciana Duranti and her colleagues on the InterPARES project place emphasis on the transactional nature of records, which they describe as documents made or received in the course of a practical activity as instruments or by-products of such activity, and set aside for action or reference [17]. In her early writings, which draw upon the medieval discipline of Diplomatics[1], Duranti observes that documents refer to acts, and that acts are subsets of facts. Facts are, according to Duranti, an event,

[1] Diplomatics is the study of the Wesen [being] and Werden [becoming] of documentation, the analysis of genesis, inner constitution and transmission of documents, and of their relationship with the facts represented in them and with their creators.

whether intentionally or unintentionally produced, the results of which are taken into consideration by the juridical system in which they take place, while acts are special types of facts which result from a will determined to produce them [18].[2] The will is expressed by the author of the record. The author of a record is the person(s) competent for the creation of the document, which is issued by that individual or by his or her command, or in his or her name (e.g., the testator in a will, the King in letters patent, the university in a degree diploma) [18].

Geoffrey Yeo offers another perspective on the nature of records. Yeo does not agree with Duranti that a record is a document made or received and set aside in the course of practical activity. Such definitions begin with emphasizing the format of the record carrier, but Yeo argues that many systems generate data that can be kept independent of any physical form and even when records are in documentary form, in some cases the record is not a single document, but a collectivity of documents (the file) [19].

Yeo proposes an alternative approach based on the notion that records are representations of activities (he expands this to occurrents, a collective noun encompassing aggregations of functions, processes, activities, transactions and events), created by participants or observers of those activities or by their authorized proxies. And that they are persistent representations in the sense that they last beyond the end of the activities to which they relate [19] [20]. In this sense, they are information systems for which the state tracking has been switched off in order to create a lasting version of an occurrent in space-time.

The link between evidence and records is another distinguishing feature of records. Duranti argues that records constitute written evidence of an activity [18]. Barbara Reed, representative of the 'Australian School' of thought on records observes that records have a dual role as evidence of actions and to "memorialize" or stand as a representation of action which can be recalled [21]. Yeo, on the other hand, argues that records are not evidence, nor do they constitute evidence; rather, records *afford* evidence. That is, they may provide it but they are not categorically equivalent to it. In any case, all three writers link records to evidence, which provides the basis for using records in accountability for decision-making [19] [20].

The final characteristic of records on which all archival theorists agree is on the link between records and context. To Reed and others in the Australian School, the highly contextual nature of records is critical, as for example, when she writes that if an individual wants to share the results of his or her actions, it is necessary to place the records in a context [21]. Within a small group of people, this can still be reasonably ad hoc, but if the individual wished to communicate beyond the domain of their immediate group, he or she would have to make more information explicit: it would be necessary to identify the individual, the organization the individual works for, what authority the individual has, what the action is, what the background to the action is, etc.. That said, not all of these relationships, networks, and linkages need to be defined at the point of capturing the record. Layers of additional data can be added at various times to extend the reach of the records as they pass from one domain of action to others. Where the possibility of reconstruction of context is not integral to

[2] Duranti further distinguishes between an act and a transaction. A transaction is a declaration of will directed towards obtaining effects recognized and guaranteed by the juridical system.

the management of information, objects are not being managed as records, according to Reed [21]. Duranti also emphasizes the importance of context, stating that for a record to be considered a record, it must possess explicit linkages to other records inside or outside the digital system, which she calls an "archival bond" [22]. Traditionally, the process known as classification formalizes the linking of records together and to their context [21].

Reflecting on the theory of the record in relation to cognitive systems raises a number of theoretical questions. The first is in relation to the transactional nature of records. Decision-making is a cognitive process. It may result in an act in the 'real world' but is not an act in itself. At best it is part of an act (i.e., the will that gives rise to the act). Nevertheless, when determining who may be at fault (e.g., in a legal dispute), it is necessary to determine who made the decision and whether the decision-making was somehow flawed. This raises two important questions. First, in the case of accountability for decision-making in cognitive systems, there is as much interest in the cognitive process, and evidence of that, as there is in the actions taken on the basis of that process. As such, the question arises as to whether the notion of records and recordkeeping needs to be extended to include the creation of representations that provide or afford evidence of the formation of will, not just the acts undertaken as a result of that will. In order to have a complete accounting for decision-making, the answer to this question must be 'yes'. This results in a novel extension of the traditional accountability framework, as depicted in Figure 1, which we argue is necessary to achieve full accountability for decision-making in the context of cognitive systems. This is so because, while a record of the final products generated from an analytical process is of great value (e.g., a record of an act), a record of the processes of the analyses themselves are just as important if not more so in the context of accountability for decision-making in cognitive systems, since these processes not only contain information on individual insights discovered, but also about how agents arrived at these insights.

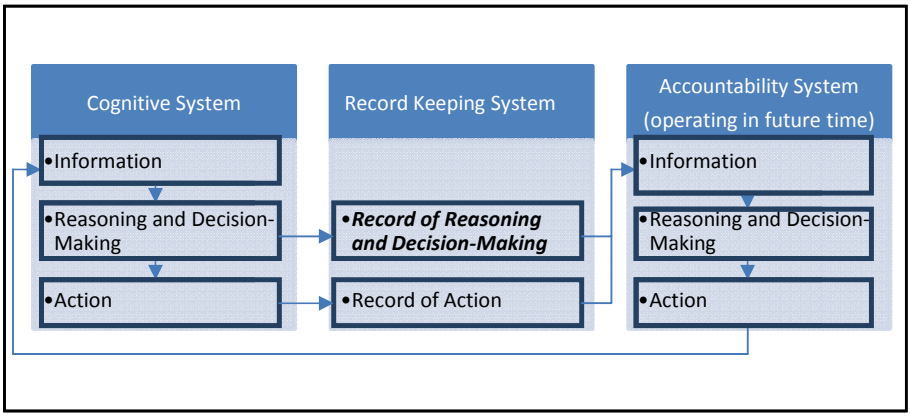


Fig. 1. A high-level representation of a theoretical framework for building accountability for decision-making into cognitive systems. The area in bold and italics captures the novel aspect of the framework: a requirement to capture a record of reasoning and decision-making.

It is possible to draw upon cognitive science to extend archival theory for the purposes of building accountability into cognitive systems. In making a decision, a human, a machine or some combination of the two will consider the data and draw insights from it. On the basis of these insights, a decision will be made that ultimately may lead to an action. Understanding this analytic process (i.e., the process by which insights are generated and by which they lead to decisions and actions) requires an understanding of human information processing, human-computer interaction, human factors, and computer processing. In the field of visual analytics, the term usually used to describe this understanding is "analytic provenance" (or also "insight provenance") [23] [24] [28]. Klein, Moon and Hoffman's and Pirolli and Card's theories of expert sense-making are among those typically used to explain the reasoning and decision-making processes of human agents using VA systems [25] [26]. These theories place emphasis on the abductive and inferencial reasoning processes used by expert human analysts as they make sense of heterogeneous data under uncertain conditions. Many AI system, however, are rule based and use deductive reasoning to arrive at new insights [13]. Thus, in distributed cognition, we may expect to find that agents, in making a decision, will use deductive, abductive and inductive reasoning processes in various combinations and relationship to one another. To acquire knowledge of analytic provenance, researchers require the use of methods that allow them to capture and conceptualize the analyst's reasoning processes, whatever type of reasoning is used. The concept of keeping track of and visualizing the analytic process has been implemented in several visual analytic tools such as Harvest [23] and CZSaw [24] [27], but these tools are not designed to capture records of decision-making for accountability purposes. Chen, Yang and Ribarsky [28] discuss the challenges of capturing the analytic process in VA systems. In many of these approaches, insights are captured from interactive visual exploration and used for supporting high level hypothesis generation and evaluation toward problem solving and decision making. A significant challenge faced by these approaches is that large amounts of insights are often involved in the sense-making process. As a consequence, insight management, such as insight recording, association, retrieval, and exchange, becomes essential for effective visual analytics approaches.

Cognitive science also helps in addressing a second question not adequately answered by archival theory, which is, who or what is the will behind the act (i.e., where does intentionality reside)? In a cognitive system, that will may be distributed, to varying degrees, between the human and the machine (i.e., there will be distributed cognition). In the case of AI, the will will be entirely the will of the machine, whereas in the case of IA, and by extension VA, the locus of decision-making may remain with the human, though augmented by the machine. If will is expressed as authorship, then in the case of AI, the author will be the machine, as the machine is competent to act, while in AI and VA, authorship is less clear in that cognition is much more distributed between machine and human. In this case, authorship should be partially attributed to systems. Duranti, who was writing before the era of cognitive systems, has disagreed with this position on the grounds that electronic systems do not possess more knowledge than a code of administrative procedure used in a complex records office: they both are embodiments of human choices, adopted because they satisfy the

needs and requirements of those whose will determines the creation of records [18]. In the era of cognitive systems, however, these arguments no longer hold because the machine (or the human and the machine working collaboratively as a unit) is able to use reasoning processes to generate new information and insights. Thus, it is necessary to reconsider the notions of will, authorship and the creation of records. Tracing the analytic provenance of decision-making makes it theoretically possible, though still technically difficult, to determine where will resides.

In addition, because the meaning of records is not to be found fully within the records themselves, but is found in the combination of their content (facts), their physical and intellectual form, and the procedures and other elements (context) that lead to the content in the record, it is necessary to link the inside, being the records, with the outside, being the context of their creation [18]. The act of assigning metadata tags to particular objects (records) can establish these critical links [21]. However, in a system wherein cognition is distributed between the human and the machine, it is only possible to tag, or assign semantically linking metadata to that portion of the cognitive process that takes place within the machine or that aspect of human cognition that is somehow captured and recorded (i.e. set aside as a record) within the machine. Moreover, for accountability purposes, the aim is to link the decision-making to some action, and that action will also take place, in many cases, outside of the machine in the real world. Thus, contextualizing records for the purposes of engendering accountability for decision-making in cognitive systems is not a straightforward proposition, and some means of linking the processes that take place within the machine to those that take place within the human and in the real world is necessary to achieve full accountability for decision-making in cognitive systems.

Having established some of the necessary conditions and actions for creating records, which may provide evidence needed as the basis of accountability for decision-making, it is necessary to consider the conditions and actions that would result in records that could be trusted as evidence for accountability purposes. Here again, archival theorists do not universally agree on the conditions that establish trustworthiness, but the characteristics that they generally identify as doing so are: 1) faithfulness of the representation, 2) reliability, particularly in respect to the author and the writer of the record, and 3) completeness of form.

Yeo acknowledges that a representation is never perfect, and is always constrained by the nature of whatever representational system is employed. Within the limits of any particular system, representational accuracy may be said to occur when the properties and relationships expressed in the representation match the properties and relationships expressed in the represented objects to the full extent that the system will allow such matching to take place [19] [20].

Duranti [18] rejects the notion that we determine the trustworthiness of records according to the extent to which they mirror facts. She notes that bureaucracy adopts two methods for assessing the record as a fact. The first method involves deciding whether record-writers are reliable. Record writers, as opposed to authors, are those who inscribe the record (e.g., a medieval monk or, today, a computer system). If the writer is reliable, the user can identify him or herself with the writer, that is, as a witness to the fact. To be able to rely on record-writers requires controlling them, by: 1) restricting the privilege of record-writing; 2) imposing sanctions on record-writers

who fail to report the fact (e.g., by requiring signatures); 3) giving responsibility to each writer for reporting only a portion of a fact, and/or increasing the number of those who report the same fact, so that what their records will have in common will be the true fact; and 4) making the same record serve a number of different users so that the writer cannot tailor the message to the audience.

The second method for assessing the record focuses on its completeness. If a record possesses all the various bureaucratically necessary forms and those forms are complete, it is possible to treat the record as an object which is showing what it purports to be. Duranti [18] argues that completeness is the bureaucrat's way to the real. For example, by requiring the indication of the place and time in which a record is written, bureaucracy transforms the record into the fact, because the mention of a date captures the relation between writer and fact, and this relationship becomes one of the things the record speaks about: a fact belonging to the past can be known by the record-user if the relationship between the person who writes about it and the fact itself is localized in space and time.

To summarize, then, building accountability for decision-making into cognitive systems depends on the making and preservation of records that represent and serve as evidence of the analytic process from the formation of insight, to the making of a decision based on insight, to action. This relies upon some means of tracing the analytic provenance of the reasoning and decision-making processes. Moreover, records, to be meaningful for the purposes of serving *ex post facto* as evidence for accountability need to be contextualized (i.e., tagged) in such a way that they can be interpreted and understood by not just the individual or machine conducting the analysis or making the decision, but by individuals (or machines) who may be scrutinizing the decision-making. Further, to be trustworthy, such records must faithfully represent, or mirror as closely as possible, the analytic process as it takes place and the resulting actions, and they must also be reliable (as to their authorship and writing) and complete in their form.

3 Application of the Theory

The above theoretical framework was used to guide the design of an interactive visual interface that builds accountability for decision-making into a VA type cognitive system. The interactive visual interface (FIVA) is a visual analytic tool that addresses two common denominator problems that often arise for all fixed income asset managers: 1) portfolio monitoring and 2) scenario analysis. With portfolio monitoring, an analyst often wishes to understand at a glance which client portfolio is performing to expectation, and which is lagging. The interactive visual interface is able to take advantage of pre-attentive processes in visual cognition and perception to reduce cognitive load in portfolio monitoring and to generate faster and more complete "situational awareness" regarding the status of client portfolios. Where, previously, fixed income analysts processed on the order of hundreds of spreadsheets to monitor client portfolios, FIVA allows analysts to visually assess portfolio status in a pre-attentive timeframe and to make decisions and take actions based on these insights.

In Scenario analysis, an analyst will hypothesize certain values in an account, or certain market or economic conditions and observe the results of those changes on the performance of the account. The major performance measure used in portfolio

monitoring and scenario analysis in the FIVA tool is the Cost of Outstanding Income, either in a given year or aggregated over a retirement plan. In other words, the performance in a given year is defined as whether the income goal can be met exactly by all the fixed income holdings that mature on that year. In the FIVA tool, the table of income goals, or streams, and the table of prices can be edited. Linking behaviours are built between the editable fields in the user interface and the two visualizations so that they can be used for scenario analysis. The analyst can edit a field to make hypotheses about various factors affecting the account, and the visualization will re-render instantly to reflect the result. The analyst can also edit multiple fields without the visualizations re-rendering automatically, and then manually command the visualization to re-render with an update button, so that the differences caused by a change of multiple fields can be seen more clearly.

Most importantly from the perspective of accountability for decision-making, FIVA provides features for an analyst to record and visualize analytic actions (for the purpose of recording the analytic process) and to restore any state of the data and resulting visualization within this process (see Figure 2). Relying on theories of visual perception and cognition, this feature is implemented via an "action ribbon" visualization that details the actions of the analyst over time along the critical path of the analytic process, rather than as a text-based log which would be more cognitively demanding for a human analyst to read, understand and draw insight from. This is akin to Harvest, which uses a similar widget to the "action ribbon" [23]. FIVA also employs an "action tree" visualization that lists, in less detail, all of the actions that the analyst has taken, including branching paths representing different lines of thought. The action tree widget in FIVA is similar to Kadivar and Chen et al`s action tree widget [24]. Thus, by means of an innovative combination of these features, FIVA captures a faithful representation of the analytic reasoning processes (whether deductive, abductive or inductive), including lines of reasoning that were considered but that led to no action.

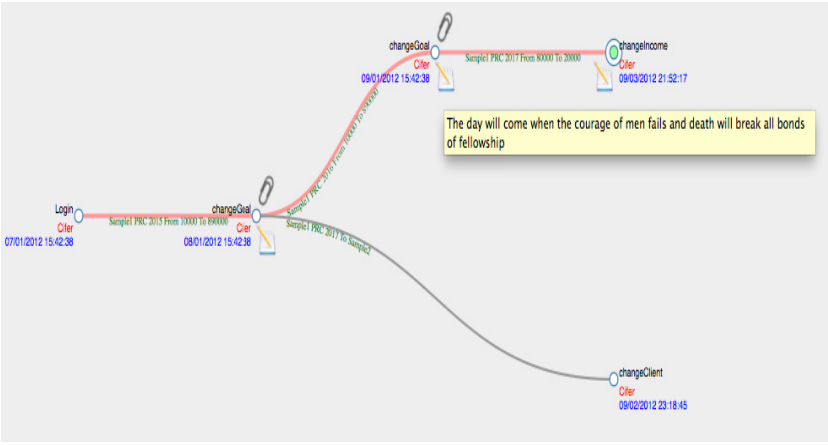


Fig. 2. FIVA Tool Activity captures the human agents' analytic reasoning and decision-making processes as they interact with the VA tool. This supports accountability for decision-making in cognitive systems.

The activity tracker is integrated into the back end of FIVA, so that every action on the FIVA interface is automatically saved, along with any change in the data. This activity history is preserved across sessions. The first step to provenance and accountability is identity, so a login system is crucial. The FIVA login is password protected to ensure authenticity. The analytic process is envisioned as a tree structure, with each action being a node, and the state change (the 'delta') of those actions being the links. At any given time, the tree always contains a critical path, highlighted in red, with the very current action being animated. The critical path is the sequence of actions leading to the current state of the data and treemap and bulletgraph visualizations seen on the FIVA dashboard. The activity tracker records every relevant and appropriate piece of information associated with each new action. All actions are time stamped and associated with a user, but a login action does not change the data or the interface in anyway. Meanwhile, for the nearby "change Goal" action, the tool records which client, and the specific account in that client portfolio, to which the action was applied, the retirement year, the old value of the data, and the new value. At any time in the analytic process it is possible to jump to another node and have the critical path leading to that node retraced and applied to the dashboard. This concept is simple but very powerful when applied to scenario analysis. A 'scenario' is defined as any critical path that can be traced on the activity tracker. Any scenario that was envisioned and tested is automatically recorded, and scenarios can be resumed and compared in one click. Changes made to the editable data columns will trigger the visualizations to recalculate the performance measure and re-render. The action recording is not limited to any type of action or any degree of granularity. Low level program actions, to configurations of the interface, to navigational actions that are relevant to most analytic processes, to high level, abstract data changes that are defined by the analyst, can all be recorded. Another important part of examining and comparing scenarios in the analytic process and providing accountability for decision-making is being able to download and compare recorded history; it is also possible to download the data associated with any state on the activity tracker. The activity tracker will automatically record actions, but it cannot automatically tell which node on the action tree represents an interesting insight. FIVA, therefore, provides functionality for the analyst to write an annotation or upload an attachment at any point of the analytic process. These can all be captured as records and linked, via classification metadata tags, to other records that document actions taken on the basis of these insights and decisions.

4 Conclusion

This paper has laid out a novel theoretical framework for engineering accountability for decision-making into cognitive systems, based on a combination of theories from the fields of archival and cognitive science, and demonstrated the application of this framework in a prototype of an innovative interactive visual dashboard for fixed income analytics that captures as a visual record the novel features of the theoretical framework (i.e., a visual record of the decision-making agents' analytic processes). As society moves into the era of cognitive systems in which more decisions are taken

by intelligent machines or humans working in collaboration with intelligent machines, it becomes increasingly pressing to consider how we will build accountability for decision-making into such systems. The rise of cognitive systems raises new theoretical and practical challenges to achieving accountability for decision-making, but without accountability society has no way to measure how much trust to place in the operation of AI and IA cognitive systems and, when problems occur, no means by which to assess the quality of decision-making in order to take corrective action or determine responsibility when harm or loss results. The hope is that this paper provides the beginnings of a foundation for engineering such functionality into cognitive systems.

Acknowledgments. The authors would like to thank Yao Shen and Qianqian Yu whose research assistance contributed to the development of the FIVA tool, and the Boeing Corporation for its generous financial support of this research.

References

1. Kelly, J.E.: Welcome to the era of cognitive systems (May 10, 2012), http://asmarterplanet.com/blog/2012/05/welcome-to-the-era-of-cognitive-systems.html
2. Computing Research Association. Grand Research Challenges in Information Systems. CRA, Washington, D.C. (2002)
3. Thomas, J.J., Cook, K.: Illuminating the path: the research and development agenda for visual analytics. IEEE Computer Society, Los Alamitos (2005)
4. Kosslyn, S.M.: Image and Mind. Harvard University Press, Cambridge (1980)
5. Lurie, N.H., Mason, C.: Visual representation: implications for decision making. J. Mark. 71, 160–177 (2007)
6. Ward, M., Grinstein, G., Keim, D.: Interactive data visualization: foundations, techniques, and applications. AK Peters/CRC Press, Massachusetts (2010)
7. Munzner, T.: Visualization. In: Ashikhminand, S.P., Marshner, S. (eds.) Fundamental of Graphics, 3rd edn. AK Peters/CRC Press, Massachusetts (2009)
8. Tegarden, D.P.: Business visualization. Commun. Assoc. Inf. Sys. 1, 2–38 (1999)
9. Tukey, J.W.: Exploratory Data Analysis. Addison-Wesley, N.Y. (1977)
10. Cooley, J.G., McKneely, J.A.B.: Command and Control Systems Engineering: Integrating Rapid Prototyping and Cognitive Engineering. Johns Hopkins APL Technical Digest 31(1), 31–42 (2012)
11. Gratton, L.: The shift: The future of work is already here. HarperCollins, London (2011)
12. Thagard, P.: Cognitive Architectures. In: Frankish, K., Ramsey, W.M. (eds.) The Cambridge Handbook of Cognitive Science. Cambridge University Press, Cambridge (2012)
13. Oaksford, L., Chater, N., Steward, N.: Reasoning and Decision-Making. In: Frankish, K., Ramsey, W.M. (eds.) The Cambridge Handbook of Cognitive Science. Cambridge University Press, Cambridge (2012)
14. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database, vol. 147, pp. 195–197. MIT Press, Cambridge (1981)
15. Hoskins, K.: The 'Awful' Idea of Accountability: Inscribing People into the Measurement of Objects. In: Munro, R., Mouritsen, J. (eds.) Accountability: Power, Ethos and Technologies of Managing. International Thompson Business Press, London (1996)

16. Stewart, J.D.: The Role of Information in Public Accountability. In: Hopwood, A., Tomkins, C. (eds.) Issues in Public Sector Accounting. Philip Allen, Oxford (1984)
17. InterPARES 2 Terminology Database, http://www.interpares.org/ip2/ip2_terminology_db.cfm
18. Duranti, L.: New uses for an old science. SAA, ACA, and Scarecrow Press, Chicago (1998)
19. Yeo, G.: Concepts of record (1): evidence, information, and persistent representations. American Archivist 70(2), 315–343 (2007)
20. Yeo, G.: Concepts of record (2): prototypes and boundary objects. American Archivist 71(1), 118–143 (2008)
21. Reed, B.: Records. In: McKemmish, S., Piggott, M., Reed, B., Upward, F. (eds.) Archives: Recordkeeping in Society, pp. 101–130. Charles Sturt University, Centre for Information Studies, WaggaWagga NSW (2005)
22. Duranti, L.: The archival bond. Arch. Mus. Informatics 11, 213–218
23. Gotz, D., Zhou, M.: Characterizing users' visual analytic activity for insight provenance. In: Proc. Visual Analytics Science and Technology, pp. 123–130 (2008)
24. Kadivar, N., Chen, V., Dunsmuir, D., Lee, E., Qian, C., Shaw, C., Woodbury, R.: Capturing and supporting the analytic process. IEEE Visual Analytics Science and Technology, 131–138 (2009)
25. Klein, G., Moon, B., Hoffman, R.R.: Making Sense of Sensemaking 1: Alternative Perspectives. IEEE Intelligent Systems 21(4), 7–73 (2006)
26. Pirolli, P., Card, S.K.: Information Foraging. Psychological Review 106(4), 643–675 (1999)
27. Chen, V., Alimadadi, S., Guenther, J., Kadivar, N., Lee, E., Dill, J., Woodbury, R.: CZSaw [computer software], http://siat.sfu.ca
28. Chen, Y., Yang, J., Ribarsky, W.: Toward Effective Insight Management in Visual Analytic Systems. In: Visualization Symposium, PacificVis 2009, pp. 49–56. IEEE Pacific (2009)

The Relationship between Management Decision Support and Business Intelligence: Developing Awareness

Rimvydas Skyrius, Gėlytė Kazakevičienė, and Vytautas Bujauskas

Department of Economic Informatics, Faculty of Economics, Vilnius University,
Sauletekio 9, LT-10222 Lithuania
{rimvydas.skyrius,gelyte.kazakeviciene,
vytautas.bujauskas}@ef.vu.lt

Abstract. A reduction of interest in management decision support systems and technologies did not reduce the importance of well-supported business decision making. Being attributed to the group of intelligent systems and technologies, decision support technologies have been largely supplemented by business intelligence technologies. Both types of technologies are supported by respective information technologies, which often appear to be quite closely related. The objective of this paper is to define relations between decision support and business intelligence technologies, based on common goals of sense-making and use of advanced analytical tools. A model of two interconnected cycles has been developed to relate the activities of decision support and business intelligence.

Keywords: management decision support, business intelligence, information needs.

1 Introduction

The once-prominent area of management decision support systems (DSS) apparently has settled to stable levels of both academic and practitioner activities [8]. However, a somewhat faded interest in decision support systems does not imply any reduction in importance of well-supported decision making, as well as general awareness of the state of internal and external business environment. On the contrary, the current economic situation in most settings demands an efficient and reliable, "military grade" management environment to support decisions, insights, recovery or mere survival.

Management decision making is an information-intensive activity, where the structuredness of the problem to be solved directly translates to the complexity of the information tasks to produce a well-supported decision. Such complexity creates difficult demands for support environment in terms of variety of information sources and incompatibility of information obtained from these sources; use of sophisticated and problem-specific analysis and modeling tools; use of communication and group support environments to fine-tune the interests of stakeholders, to name a few [8]. Computerized management decision support is intended to boost intelligent management activities, where the expected payoff is quite high. So it is no surprise

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 587–598.
DOI: 10.1007/978-3-642-36981-0_54

that for a prolonged time developers and producers of support technologies have designed support products ranging from a simple functional support (Microsoft Excel being the most prominent example [1]) to fully automated decision making functionality (expert systems).

Decision support alone, being reactive and activated only when a problem is encountered, eventually proved to be insufficient. The problem solving context received IT-based support mostly from the resources of a regular information system, therefore of a limited nature and in most cases complicated by time pressures. An alternative use of decision support, if coupled to a proactive monitoring of the environment, ensured better understanding of the problem context, leading to higher decision quality. A term "business intelligence" came into use, serving as an umbrella term for tools and technologies that let business information users stay aware of changes in internal and external environments.

The research problem of this paper is centered around how the current array of technologies and approaches provides support for business decision making functions. Currently there is a confusion in defining whether business intelligence is a part of decision support function, or vice versa; eventually this confusion spreads to business management community which at all times has expressed the need for reliable decision support which would justify substantial investments into support technologies. In this paper, the authors have decided to use the results of their earlier research to make an attempt in developing a model relating business intelligence and decision support functions.

2 Business Intelligence and Information Needs

Although business intelligence is regarded as a relatively new term, with authorship assigned to Howard Dressner of Gartner Group in 1989, we can have a retrospective look at the mission of management information systems (MIS), whose role of keeping management aware of the state of business has never been downplayed, and mission definitions for MIS sound very much like the mission definitions for business intelligence today. A few explanations of MIS role from earlier sources are presented below:

- "Two types of information for strategy implementation are in use. The first one is the external information, used for strategy development. The second type is internal information, used to monitor strategy execution" [13].
- "A management information system refers to many ways in which computers help managers to make better decisions and increase efficiency of an organization's operation" [6].
- "For information to be useful for managerial decision making, the right information (not too much and not too little) must be available at the right time, and it must be presented in the right format to facilitate the decision at hand" [3].
- "A management information system is a business system that provides past, present, and projected information about a company and its environment. MIS may also use other sources of data, particularly data about the environment outside of the company itself." [5].

- "The systems and procedures found in today's organizations are usually based upon a complex collection of facts, opinions and ideas concerning the organization's objectives. … For an organization to survive, it must learn to deal with a changing environment effectively and efficiently. To accomplish the making of decisions in an uncertain environment, the firm's framework of systems and procedures must be remodeled, refined, or tailored on an ongoing basis." [2].

There are definitions of business intelligence that do not differ much from the above definitions; e.g., Vuori [15] states that "… business intelligence is considered to be a process by which an organization systematically gathers, manages, and analyzes information essential for its functions". In order to have a more precise definition of business intelligence, we have to decide whether all informing functions are „intelligence" because they increase awareness, or does BI have a clear separation from other (lower level) informing functions. If so, the separation criteria between BI systems and any other management information systems have to be defined. For the purposes of this paper, we will use the division of management information needs along two dimensions – their simplicity or complexity, and common or specific focus, as presented in the Table 1 and based on earlier work by one of the authors [10]:

Table 1. Relation of simple-complex and common-special information needs

	Simple needs	Complex needs
Special needs (problem-specific)	Simple special needs	Complex special needs
Common needs (available permanently)	Simple common needs	Complex common needs

The mission of BI becomes clearer if weighted against the types of served information needs. Regarding the positioning of these needs against the axis of simple-complex information needs, they usually fall into the more sophisticated part of the information needs complexity spectrum. Same can be said about the process of decision making, which often requires sophisticated tools to support awareness, communication, sense-making and evaluation of risks. The dimension of common and special information needs separates decision making from the rest of business intelligence in a sense that while decision support activities are directed towards a certain problem which has been recognized and has created a task of its solving, business intelligence can be considered an activity which, apart from encompassing decision support, allows the discovery of problems and general awareness about the state of activities.

3 Decision Support and Business Intelligence Processes

3.1 Decision Support Process

A decision making process includes a number of stages, and if accumulation and subsequent use of experience is included, the process takes a cyclical nature (Fig. 2, from [10]):

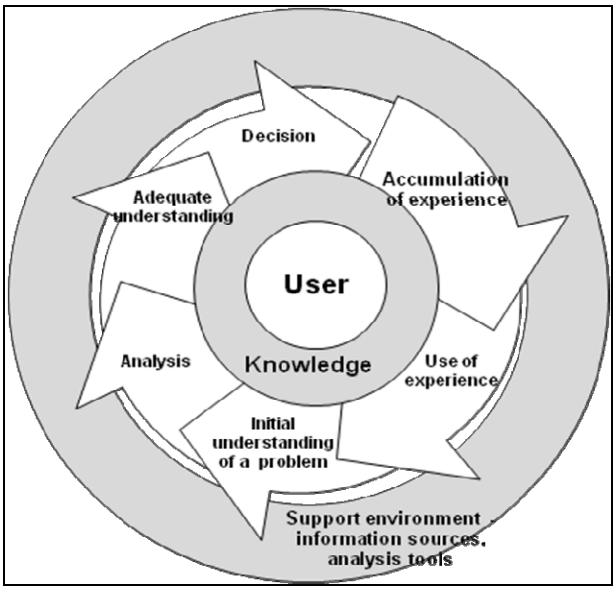


Fig. 1. The decision support process [10]

The structure of the decision support process can be related to relevant information needs:

1. Monitoring (using previous experience): the environment, both internal and external, is being watched to notice things worth attention; *simple* and *common* information needs prevail.
2. In the case of recognizing a situation of interest (initial understanding of a problem or opportunity) the situation is evaluated and given extra attention to achieve desired understanding. At this stage *special* information needs arise.
3. Additional analysis and decision development is required if the situation is complex enough (semi-structured or unstructured); *simple* needs are complemented by *complex* needs; more information is brought into decision making environment; specific problem-solving tools such as formal approaches and models are likely to be used to achieve an adequate understanding of a problem.
4. The decision-making stage involves formerly available as well as newly gained understanding of the situation, and the decision maker or makers will use all possessed knowledge to arrive at the best possible decision, time or other circumstances permitting. In this paper, the term "knowledge" is deliberately avoided most of the time, but here it serves to show that data or information alone are insufficient for decision making; all that is known will be used in its entirety, and new knowledge most likely will be gained.
5. The experience accumulation stage records the newly gained experience from both decision making and its implementation, and keeps it for possible reuse. *Special needs* become *common*, adding new material to the already available body of experience, and the need to capture the essential features of the recorded

case keeps this sort of information need in the complex segment. This phase should also include the practical experience in decision implementation, which can sometimes reveal additional circumstances of the problem.

6. The use of new experience, along with that formerly accumulated, brings the process back to stage 1 – monitoring.

Stage 1 of the above process is directly related to (or can be considered a part of) business intelligence, because that's where the actual monitoring of the business environment is being done. Stage 2 is a principal point of joining business intelligence and decision support.

As we can see, during the decision making process the focus of information needs moves around the quadrants of Table 1: stage 1 concentrates in the simple/common sector; stage 2 moves on to simple/special sector, stages 3 and 4 concentrate in the special/complex sector, stage 5 moves into complex common sector, and finally stage 6 brings the focus back to simple/common sector.

3.2 Business Intelligence Process

The business intelligence process, too, takes a cyclical nature (Fig.4., from [15]), and includes the stages of information needs definition, information collection, information processing, analysis, information dissemination, information utilization and feedback. The cycle structure is justified if the received feedback helps to reevaluate or redefine information needs.

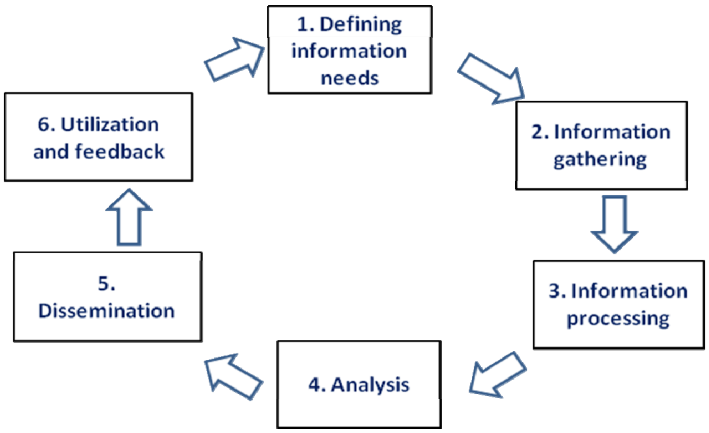


Fig. 2. A generic business intelligence process model [15]

In business intelligence process, there's usually no clear concentration on a specific topic or problem, and the resources of a BI system are used for constant monitoring of internal and external business environment. In other words, such systems serve *common* information needs to keep users informed about the state of business

environment, often combining a monitoring function with alerts, exception reports and other tools to draw attention to changes or inconsistencies. Therefore, an important feature of BI systems is their ability to produce a complete composite view that would help avoiding surprises.

The business intelligence cycle, as presented in Fig. 4., raises several questions. First of all, it does not disclose the difference between regular management information systems or their current incarnation, ERP systems, and business intelligence systems. It is unclear, for example, whether external information is used in the cycle, and if so, in what ways. Secondly, the cyclical feedback should invoke the re-evaluation of information needs, as business conditions change, or some needs have been incorrectly assessed from previous cycles (inclusion of irrelevant information or omission of important information).

From the above descriptions of technologies and processes for both decision support and business intelligence we can define two different but interrelated cycles: cycle 1 for business intelligence process, and cycle 2 for decision support process (Fig. 5).

As cycles 1 and 2 unfold, the focus moves around different types of information needs. In cycle 1, the steps of information gathering and processing can be attributed to the *common* and *simple* part of information needs. The analysis step uses processed information and produces derivative results that produce additional insight and move from *simple* to more *complex* needs. If a problem situation is recognized, *special* needs arise, and cycle 2 is activated. For a problem analysis, *special* needs may be both of *simple* and *complex* nature, depending upon the severity of a problem. A problem-specific model is developed for better understanding of the problem and evaluating the alternatives. Decision implementation brings in valuable experience that is saved for later reuse and, together with other experience, satisfies *common* information needs important both for future business intelligence and decision making.

4 Issues in IT Use for Decision Support and Business Intelligence

The opinions on IT role in supporting the sophisticated side of information needs can be roughly split into deterministic approaches and behavioural, human-centered approaches. The former assign prime importance to IT performance and ability to automate complex analytical procedures [1], while the latter assign prime importance to human skills and creative powers ([4], [12], [14]), at the same time stating that the majority of existing decision support and analytical tools are technology-centric rather than user-centric. The conflicting attitudes have initiated a survey, performed earlier by one of the authors [10], where issues like monitoring of internal and external environment, IT role in the monitoring process, and experience management have been researched. The survey had yielded 250 responses from a convenience sample of managers of small and medium businesses in Lithuania.

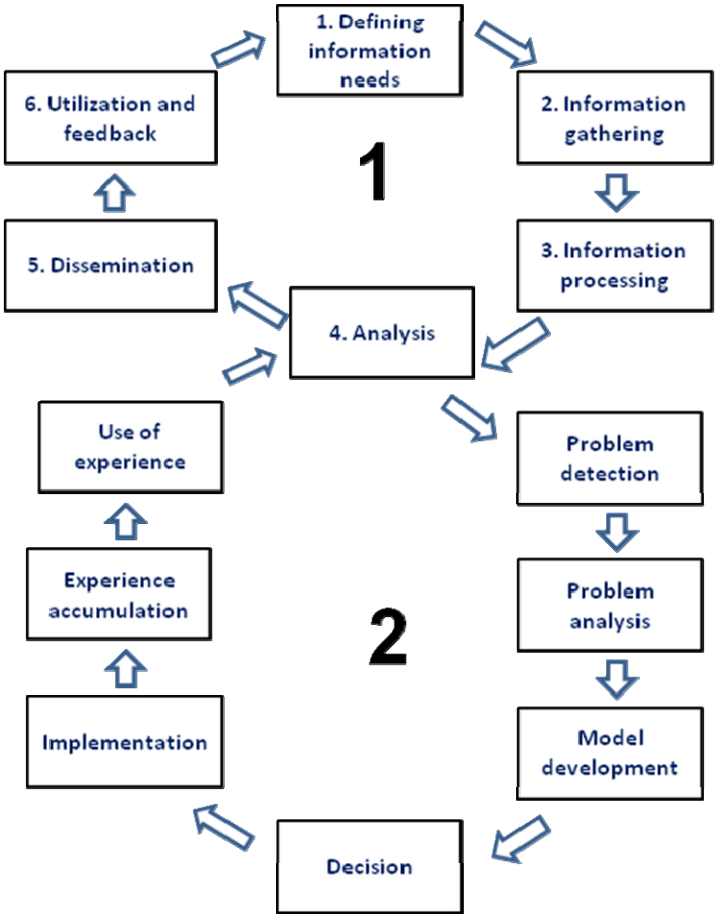


Fig. 3. Relation of business intelligence (1) and decision support (2) cycles

Regarding the *monitoring of internal organization environment*, the users appeared to be quite comfortable using IT for monitoring key data about their organization's activities. Such information is contained within their in-house information system that has been created to monitor these activities. The absolute majority of responders (161 or 64.4%) have indicated that IT is used to monitor all issues relating to an organization's internal information needs; such needs are attributed mostly to the simple common needs. The information system-based information tasks are largely routine, and satisfaction of this type of information needs does not pose any significant problems.

For *external monitoring* the use of IT is significantly lower; the number of responders having indicated that they use IT to monitor all external issues has been 125, or 50%; 122 responders, or 48.8%, had stated that they use IT for some of the external monitoring issues. The lower numbers of use do not point to second-rate

importance of external monitoring; rather, they indicate that the sources of external information are not under the control of a single own information system, as it is in the case of internal information sources. The external environment, being an important source of changes, opportunities and risks, is much more turbulent, and there is a greater variety of issues to be monitored, information sources, formats, and access modes; this variety significantly complicates the use of IT for external monitoring.

Supporting the detection of important changes, IT had been considered a helpful aid in monitoring and detecting changes, but rather limited in supporting information needs for sense-making. The absolute majority of responses (105 out of 207 responders having indicated that IT has some role in detecting important changes, or about 51%) stressed the role of IT as a principal technical support tool. No responses stated that IT had significantly supported the function of sense-making (revealing important changes in the environment).

The reuse of experience and competence information is one of the most important functions in the process chains of BI and DS; this statement can be supported by a seemingly growing number of published work on experience management systems. The results of the survey have indicated that the reuse of important problem-solving and decision making experience is of mixed success; recorded practice is reused – in most cases conditionally, as situations change and information needs have to be constantly re-evaluated. The survey had also shown that experience records are recorded in all convenient ways: free text format in digital media, structured format (with some standardized features and values) in digital media, and same on paper. IT role can be seen mostly in arranging, managing structures, imposing standards, and allowing easy filtering and retrieval. Level of reuse is limited due to changing context, although the reuse of templates, structures, models and other procedural issues is commonplace.

Decision-making information needs are hard to plan because of their variety and unstructuredness. Regarding this issue, the respondees have been asked about:

- decision making infomation needs that are known beforehand, and the principal types of such information;
- decision making information needs that are not known beforehand and emerge in the process of developing a decision, and the principal types of such information.

The *known* information needs relate to information whose content and location are known and accessible because of earlier experience, or this information is already available. This information or tools for its access can be placed in close proximity to the decision makers. The distribution of responses between the different types of this information is given in Table 2.

The *unexpected* information needs emerge mostly because of turbulent business nature, are hard to plan, and the use of programmed solutions is rather limited. The distribution of responses between the different types of this information is given in Table 3.

Table 2. Known information needs for decision making

Market information (customers, sales, needs, opportunities)	49	19.6%
Competition information (competitors' status, strength, intentions, actions)	29	11.6%
Internal information (financials, capacity, inventory)	27	10.8%
Legal information (laws, regulations, standards)	26	10.4%
No such cases	26	10.4%
Technical information	2	0.8%
Did not specify	91	36.4%
Total:	**250**	100.0%

Table 3. Unexpected information needs for decision making

No such cases	86	34.4%
Yes, there have (without specifying the information)	46	18.4%
Market information	23	9.2%
Internal information	15	6.0%
Competition information	14	5.6%
Legal information	14	5.6%
Technical information	14	5.6%
Informal, "soft" information (e.g., opinions, foresights)	12	4.8%
Confidential information (e.g., customer reliability checks)	5	2.0%
Did not specify	21	8.4%
Total:	**250**	100.0%

The distribution of both responses is not much different, and suggests that often decision makers have to look deeper into existing issues ("more of the same"). However, the significant presence of unexpected information needs might require a set of support tools that would allow tailored approaches using assorted decision support techniques – e.g., modeling, data mining, text mining, information integration and others.

The above separation of information needs into known and unexpected roughly corresponds to the related cycles pictured in Fig.5, where the business intelligence

cycle is performed mostly against known information needs. If a specific problem is detected, the known needs together with readily available information move to the decision support cycle, where additional information needs of unexpected nature are likely to emerge. This approach can be useful in designing business intelligence environments incorporating a sub-level for decision support, with generic functionality contained mostly in the 1st cycle, and the problem-specific tools an techniques in the 2nd cycle.

5 Discussion and Conclusions

There is no doubt that the need for well-informed business decisions, as well as for general awareness of developments in the business environment, will remain acute. The current state of management decision support gets more complicated as rapidly changing conditions often require swift reaction, information overload is commonplace, and additional issues arise regarding information quality [7]. Under these conditions, a need for right information at the right time and in the right place remains essential, and the well-aimed and reasonable use of support technology can increase decision making quality and efficiency, regardless of whatever name this technology is bearing at the moment.

We suggest here to use here the arguments presented in this paper, regarding the development of an efficient information environment for decision makers. It has been proposed that such environment should be split into two tiers:

- the first tier containing a simple set of support tools that are close and easy to use;
- the second tier containing more distant and more complicated information sources and processing techniques that are required much less often;
- manageable support environment that allows easy switching of items between tiers, similar to the form of managerial dashboards with interchangeable items on display.

The items contained in the first tier would be required most of the time, simple to use and able to be configured to the users' needs:

- basic data on internal and external environment: sales, market share, cash-at-hand, order or project portfolio, comparative figures by time/place/product etc.;
- information access tools: simple search in own sources – databases and data warehouses, simple search in public sources, tools for arranging search results (e.g., by relevance or size), easy classification and annotation;
- tools for simple calculations: templates, financial models, other simple models.

The second tier might include:

- access to more distant and complex information sources with advanced search tools;
- modelling tools for forecasting, simulation, scenario development;
- data analysis and presentation technologies – drill-down tools, OLAP queries, data and text mining facilities, graphing and visualization tools.

Such split of functionality would roughly reflect required functions for generic business intelligence and decision support cycles respectively. It would also allow for required cross-functionality in the cases when simple decision support needs would be well-served by first tier functions alone, or when business intelligence needs would required more advanced tools. The more defined set of features for both tiers of the support environment could lead to a possible set of requirements for the interface design of an information environment for decision makers.

The above presented survey results show, among other findings, that a rational balance between functionality of IT and flexibility of user activities is a goal that should be addressed against the criteria of better awareness of business environment. Some possible risks to avoid can be presented here:

- Business intelligence sometimes seems tempted to squeeze any possible meaning out of available data, which complicates information activities significantly – the ultimate goal of intelligence is action, not awareness *per se*;
- The definitions of important information to look at in a constantly changing environment resemble "shooting at a moving target" and have to combine tried-and-tested checkpoints with required flexibility for adjustment; this is especially true for early warning systems and detection of anomalies;
- Real-time business intelligence seems a tempting option; however, it is expensive and of limited demand: in another survey [11], most of the permanently monitored data has a latency of 1 day, which appears to be quite sufficient for most business activities;
- Excessive betting on information technology, when the primary success factors, as a number of sources indicate, are user motivation, creation of intelligence culture and horizontal intelligence networks inside the organization.

References

1. Bhargava, H., Sridhar, S., Herrick, C.: Beyond Spreadsheets: Tools for Building Decision Support Systems. IEEE Computer 32(3) (March 1999)
2. Davenport, T.: Competing on analytics. Harvard Business Review, 99–107 (January 2006)
3. FitzGerald, J., FitzGerald, A., Stallings, W.: Fundamentals of systems analysis. John Wiley & Sons, New York (1981)
4. Gremillion, L., Pyburn, P.: Computers and information systems in business. McGraw-Hill, New York (1988)
5. Johnstone, D., Bonner, M., Tate, M.: Bringing human information behaviour into information systems research: an application of systems modelling. Information Research 9(4), paper 191 (2004), http://InformationR.net/ir/9-4/paper191.html
6. Kroenke, D., McElroy, M., Shuman, J., Williams, M.: Business computer systems. Mitchell Publishing, Santa Cruz (1986)
7. Pick, J.: Computer systems in business. PWS Publishers, Boston (1986)
8. Power, D.: Decision Support Basics. Business Expert Press (2009)
9. Shim, J.P., Warkentin, M., Courtney, J., Power, D., Sharda, R., Carlsson, C.: Past, present, and future of decision support technology. Decision Support Systems 33, 111–126 (2002)

10. Skyrius, R.: Satisfying Complex End of User Information Needs: User Experiences. The Business Review 6(2), 132–138 (2006)
11. Skyrius, R.: The current state of decision support in Lithuanian business. Information Research 13(2), paper 345 (2008), http://InformationR.net/ir/31-2/paper345.html
12. Skyrius, R., Bujauskas, V.: Business intelligence and competitive intelligence: separate activities or parts of integrated process? In: The Global Business, Economics and Finance Research Conference, London (2011)
13. Sjoberg, L.: The distortion of beliefs in the face of uncertainty. SSE/EFI Working Paper Series in Business Administration, 2002:9 (2002)
14. Tom, P.: Managing information as a corporate resource. Scott, Foresman and Company, Glenview (1987)
15. Turpin, M., du Plooy, N.: Decision-making biases and information systems. In: The 2004 IFIP International Conference on Decision Support Systems (DSS 2004), Prato, Italy, pp. 782–792 (2004)
16. Vuori, V.: Methods of Defining Business Information Needs. In: Frontiers of e-Business Research Conference, Tampere University of Technology, Tampere (2006)

Multi-Agent System for Teaching Service Distribution with Coalition Formation

José Joaquim Moreira[1,3] and Luís Paulo Reis[2,3]

[1] ISLA Gaia, Rua do Cabo Borges, 55, 4430-646 V. N. Gaia, Portugal
[2] EEUM/DSI - School of Engineering of the University of Minho, Information Systems Dep., Campus de Azurém 4800-058 Guimarães, Portugal
[3] LIACC - Artificial Intelligence and Computer Science Laboratory of the University of Porto
jose.moreira@unisla.pt, lpreis@dsi.uminho.pt

Abstract. In University Management, one of the recurring problems that each department has to solve, each year, is the Teaching Service Distribution (TSD) or Teaching Assignment Problem (TAP). The problem of TSD consist to assign teachers to courses classes - lectures, tutorials, practical or laboratory - taking into account these preferences and qualifications for teaching. This is a crucial stage, since it is almost imperative that the TSD is fully defined before the process of schedules generating. However, most institutions of higher education, don't have a specific software tool to support the process of TSD. In this paper we propose a new approach for solving the TSD consisting on the formulation of the problem as a distributed scheduling problem with the formation with coalitions formation. The problem is solved in the context of a multi-agent system where the real agents are modeled by computational agents, with their interests, but may cooperate in alliance groups.

Keywords: University management, TSD, TAP, Scheduling, Multi-agent systems, Coalitions formation.

1 Introduction

In University Management, one of the recurring problems that each department has to solve is Teaching Service Distribution. This is a crucial phase since it is almost imperative that the Teaching Service Distribution is fully defined before the process of generating timetables.

The problems of scheduling in University Management are often modeled as combinatorial optimization problems and solved automatically using methodologies typical of these problems.

In this paper we propose a new approach to solving the problem of Teaching Service Distribution consisting in the formulation of the problem as a scheduling problem, combining multiple scheduling and optimization algorithms with coalition formation in a multi-agent system where the real agents are modeled by computational agents, with their interests, but may cooperate in alliance groups.

The main aim of the work was to formulate the problem of Teaching Service Distribution in Portuguese universities as a problem of coalition formation in a

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 599–609.
DOI: 10.1007/978-3-642-36981-0_55

multi-agent system. Teachers are modeled as autonomous agents that associate in coalitions to execute tasks, we mean teaching courses, seeking to maximize their own evaluation functions.

The problem of the Teaching Service Distribution, which in specialty papers is referred as Teacher Assignment Problem (TAP) [1], [2] consists of assigning teachers to classes courses – theoretical, theoretical and practical, laboratory or practice – taking into regard the preferences and qualifications of these to the lectures.

A coalition can be defined as an alliance between individuals, during of a given process, in which they cooperate in common objective, but each with its own interests. In the area of Informatics and computer science, the entities that make up the coalition are typically autonomous software agents.

An agent can be defined as a computational entity, commonly software, located in a particular environment, which has the perception that environment via sensors, is capable of reasoning and acts autonomously in that environment through actuators in order to perform a given function for which it was designed. Some researchers say an agent to be considered as such, must also have communication capabilities with other high-level agents and / or humans. However, the definition of agent is not consensual and is so problematic for the community of Distributed Artificial Intelligence as the definition of intelligence to the Artificial Intelligence community.

Multi-Agent Systems are systems composed of multiple agents that exhibit an autonomous behavior but at the same time interact with the other agents present in the system. These agents exhibit two fundamental characteristics: being able to act autonomously, making decisions, leading to the satisfaction of their goals, being able to interact with other agents using protocols inspired by social interaction of humans and requiring coordination, cooperation and negotiation.

2 Related Work

In this section, we provide a brief overview about the various scheduling problems existing in University Management, within this highlighting the problem of Teaching Service Distribution, making an approach to its solution. It is also the subject of this section the problem of coalition formation, identifying the main classes of algorithms to generate coalition's structures.

The most significant scheduling problems that are commonly found in University Management are as follows [3]:

- Teaching Service Distribution or Teacher Assignment Problem;
- School Timetabling**;**
- University Timetabling;
- Examination Timetabling;
- Allocation of rooms.

This subject has received special attention of the scientific community in the last five decades. This great interest, causes in 1995, the creation of series of conferences PATAT (Practice and Theory of Automated Timetabling) with editions every two

years [4] and the establishment of EURO (Association of European Operational Research Societies) WATT (Working Group on Automated Timetabling). In 2002 emerged with the support of PATAT, the International Competition of Timetabling [5].

The problem of production of a factory described by Thompson [6], the problem of traveling salesman approached by Wren [7] and the problem of school timetabling, with a solution proposed by Queirós [8], for example, can be seen in perspective problems of sequential scheduling.

Oliveira [9] presents a language for representation of the timetabling problem, the UniLang. UniLang intends to be a standard suitable as input language for any timetabling system. It enables a clear and natural representation of data, constraints, quality measures and solutions for different timetabling (as well as related) problems, such as school timetabling, university timetabling and examination scheduling.

Gröbner [10] presents an approach to generalize all the timetabling problems, describing the basic structure of this problem. Gröbner proposes a generic language that can be used to describe timetabling problems and its constraints.

Chan [11] discusses the implementation of two genetic algorithms used to solve class-teacher timetabling problem for small schools.

Fang [12], in his doctoral thesis, investigates the use of genetic algorithms to solve a group of timetabling problems. Presents a framework for the utilization of genetic algorithms in solving of timetabling problems in the context of learning institutions. This framework has the following important points, which give you considerable flexibility: a declaration of the specific constraints of the problem and use of a function for evaluation of the solutions, advising the use of a genetic algorithm, since it is independent of the problem, for its resolution.

Fernandes [13] classified the constraints of class-teacher timetabling problem in constraints strong and weak. Violations to strong constraints (such as schedule a teacher in two classes at the same time) result in an invalid timetable. Violations to weak constraints result in valid timetable, but affect the quality of the solution (for example, the preference of teachers for certain hours). The proposed algorithm, evolutionary, has been tested in a university comprising 109 teachers, 37 rooms, 1131 a time interval of one hour each and 472 classes. The algorithm proposed in resolving the scheduling without violating the strong constraints in 30% of executions.

Eley [14] in PATAT'06 presents a solution to the exam timetable problem, formulating it as a problem of combinatorial optimization, using algorithms Ant, to solve.

Moreira [15] presents a solution method to the problem of automatic construction timetables for the exams. The method of solution is a meta-heuristics that includes a genetic algorithm.

Although several problems have been solved using optimization and learning algorithms [16-18], reduced attention has been given to the application of this type of algorithms to the TAP problem.

Gunawan et al. [1] and Gunawan and Ng [19] consider the teacher assignment problem in two different ways, where both perform the work in two stages. In first stage allocates courses to teachers and in the stage divides the course into classes and

allocates among teachers in order to balance the workload assigned to each teacher. In [1] solves the problem using Genetic Algorithms, while in [19] Gunawan and Ng uses the metaheuristics Simulated Annealing and Tabu Search.

Some works shows that when compared with the manuals schedules in real learning institutions, the times obtained by the algorithms for solving the timetabling problem are better quality, since, uses some function of evaluation.

The Teacher Assignment Problem (TAP), due to its complexity, most of the studies in the literature solves the problem using heuristic methods. However, when the instance of the problem has reduced size, the exact methods of resolution are an alternative. An example of application for this form of resolution is the work of Ferreira et al. [20]. This work also seeks maximize the satisfaction of the teacher by having groups of courses associated with him, however the teacher may receive courses that were not requested by him.

The TAP can be modeled using different purposes: minimizing the number of classes per teacher [21]; balancing the workload of teachers [22]; attendance of preferences of teachers [23], [24], [25], [26] and [2]. This last aspect is required in our work.

On the other hand, literature shows that several problems may benefit a lot from a multi-agent systems base approach [27-37].

3 Teaching Service Distribution

The TSD or TAP consists of assigning teachers to classes courses - theoretical, theoretical and practical, practical or laboratory - taking into account the preferences and qualifications of these to teach.

In the process of TSD or TAP occur several steps:

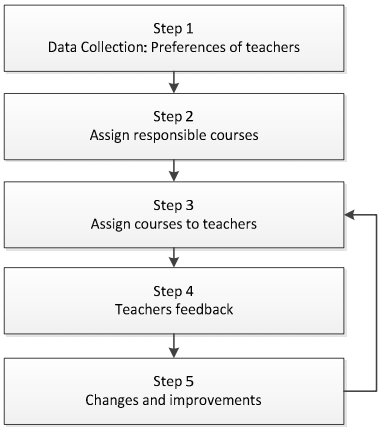


Fig. 1. Main steps in TAP

3.1 Problem Solution

The solution of TAP can be defined as:

> *For each course teaching at university, set teachers, conductors and for each class of each type to teach each course, set the teacher responsible for teaching, satisfying a given set of constraints.*

The TAP consist in define the teachers responsible for each of the courses - teachers qualified to teach, taking into account various criteria and constraints. The constraints include respect for the capabilities of teachers, the preferences expressed by them and respect for workloads that can be assigned in each period for each teacher.

In the process of Teaching Service Distribution there is a large number of rigid constraints that must be respected in order to obtain an adequate solution. There are also other restrictions more flexible whose non-infringement, although not is essential, contributes to increase the overall quality of solution.
The rigid constraints include:

- For all courses must be allocated a regent and teachers for all their classes;
- Teachers should teach an annual number of hours with minimum and maximum limits set by their categories, types of contracts and possible reductions in service due to other academic positions;
- Maximum limit of different courses that a given teacher can teach;
  Limit maximum weekly hours per period, usually semester, that teachers can teach;
- A teacher cannot teach a course for which he is not qualified;
- Impossibility of certain teachers teach in specific periods of time, trimesters or semesters.

The flexible constraints include:

- Preferences of teachers in relation to courses to teach and to be regent in each period;
- Teachers should not have a number of different courses to teach lower / higher than limits set;
- Teachers should not have a number of hours per week to teach in each period, usually semester, lower / higher than limits set;
- Preferences regarding the distribution of the workload of teachers for various periods of the year;
- Preferences regarding the formation of groups of teachers to teach specific courses;
- Incompatibilities of teachers to teach courses together with other teachers;
- Preferences course coordinator about the teacher teach each course;
- Respect for professional category and internal hierarchy of teachers;
- Preferences or other directives of the board or the department rectory of the University;
- Preference for a particular teacher teach in the afternoon / morning / evening, which may affect the allocation of some courses;
- Preferences of teachers to teach courses to specific classes.

Formalization of the Problem

Due to the large number of constraints of hard implementation and forms of evaluation of the solution, very subjective, is not easy to formalize the concept of Teaching Service Distribution. So, simplifying some constraints, we formalize the concept of problem Teaching Service Distribution as a combinatorial optimization problem.

Suppose the existence of:

- A set of Teachers = $\{T_1, T_2, ..., T_n\}$ – teachers who teach at the University;
- A set of Courses = $\{C_1, C_2, ..., C_n\}$ – courses taught at the University;
- A set of Sections = $\{S_1, S_2, ..., S_n\}$ – sections of different courses taught at the University;
- A set of Periods = $\{P_1, P_2, ..., P_n\}$ – different teaching periods (semesters, trimesters, etc.).

It is also known:

- $Preferences_i = \{Pref_1, Pref_2, ..., Pref_n\}$, $\forall$ i $\in$ Teachers, corresponding to the preferences of teachers regarding the teaching of each Course;
- $Capabilities_i = \{Cap_1, Cap_2, ..., Cap_n\}$, $\forall$ i $\in$ Teachers, corresponding to the capacities of teachers regarding the teaching of each Course;
- $Importance_i$, $\forall$ i $\in$ Teachers, representing the importance of each teacher at the University;
- $MaxHours_i$, $\forall$ i $\in$ Teachers, assigning each teacher a maximum annual hours;
- $MinHours_i$, $\forall$ i $\in$ Teachers, assigning each teacher a minimum annual hours;

The decision variables are:

- Regent = $\{RC_1, RC_2, ..., RC_n\}$ where $RC_i \in$ Teachers, corresponding to the regents of the various courses taught at the University;
- Service = $\{SS_1, SS_2, ..., SS_n\}$ where $SS_i \in$ Teachers, corresponding to the teachers assigned to each of the sections of several courses taught at the University;

Thus, Teaching Service Distribution consists of defining the values assigned to the variables of Regents and Service in order to:

$$\text{Minimize } f = \Sigma w_{load} * f_{teaching\_load} + w_{adequacy} * f_{teaching\_adequacy}$$

Where:

- w_{load} and $w_{adequacy}$ the relative weights given to respect for desired teaching loads for teachers and the adequacy of the teachers to teaching the courses;
- $f_{teaching\_load}$ a function that measures the respect for the teaching load of teachers.
- $f_{teaching\_adequacy}$ a function that measures the respect for the preferences and adequacy of the teachers to teach the courses.

4 Methodologies

In figure 2, we can see the main modules of the computer system for teacher assignment.

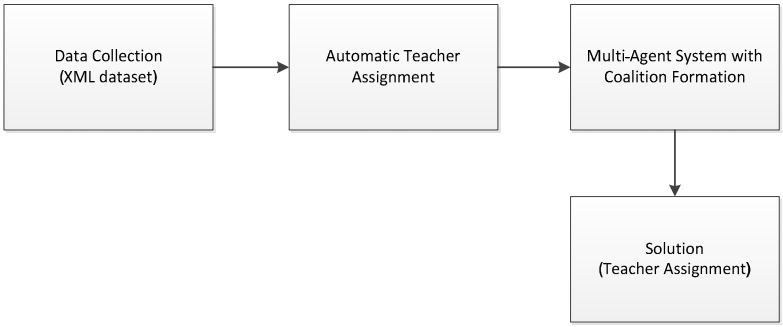


Fig. 2. Modules of the computer system for teacher assignment

4.1 Methods of Solving Problems of University Management

Some methods for the automatic resolution of problems of University Management:

- Simulated Annealing – Meta-heuristic proposed by Kirkpatrick et al. in 1983 [38]. Local search method that simulates the technique used for tempering metals by cooling.
- Tabu Search – The Tabu Search algorithm was designed in 1989 by Glover [39]. Uses the concept of neighborhood of a solution, using a concept of selective memory.
- GRASP (Greedy randomized adaptive search procedures) – Algorithm proposed by Resende in 1989 [40]. It consists of a probabilistic iterative method, where each iteration is obtained in a solution of the problem.
- Genetic Algorithm - Conceived by John Holland [41] at the end of years fifties, and uses a structure similar to that enunciated by Charles Darwin in "The Origin of Species'.
- VNS – In 1997 Pierre Hansen and Nenad Mladenovic [42], have proposed the metaheuristic Variable Neighborhood Search (VNS).
- Ant Colonies – Algorithms that are based on the social behavior of ants, and begins with the work of Colorni, Dorigo and Maniezzo and Goss in 1991 [43] [44].
- A* – That was first described in 1968 by Peter Hart, Nilsson, and Raphael [45]. In the article, was called A algorithm, however using this algorithm together with an appropriate heuristic is reached optimum performance, and became known as A*.
- Particle Swarm Optimization – Introduced by Kennedy and Eberhart in 1995 [46] emerged from experiments with algorithms that model the social behavior observed in many species of birds.

Rasmussen [47] proposes the classification shown in figure 3 for the optimization algorithms:

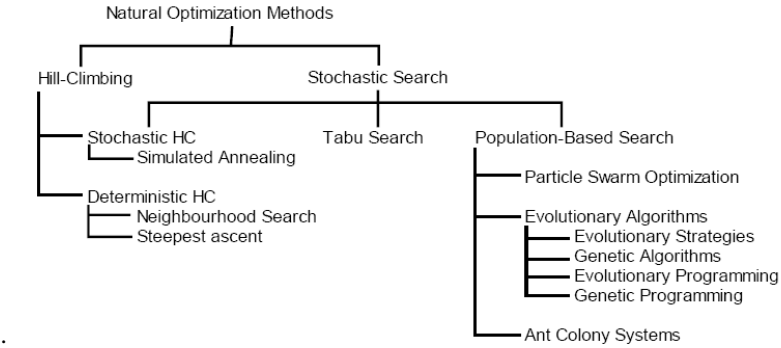


Fig. 3. Classification of natural methods for optimization

4.2 Methods for Coalition Formation

In this section, we present a classification of different algorithms to solve the problem of generating coalition structures. We can identify the following classes [48]:

- Low complexity algorithms that returns an optimal solution – This class of algorithms is designed to return an optimal solution while minimizing computational complexity. We point to that this kind of algorithms is preferable because, after all, the aim is to find an optimal solution and minimize the computational complexity.
- Fast algorithms that provide no guarantee on their solutions – These algorithms do not provide any guarantees to find an optimal solution. Instead, simply return solutions "good", but quickly. In this context, heuristic methods, such as simulated annealing, neuronal networks and genetic algorithms, constitute good methods to find a solution "good" [49].
- Anytime algorithms that return solutions within a bound from the optimal solution. The reason for the existence of this class of algorithms is that if the search space is too large to be completely analyzed, then the other option does not necessarily have to be the application of an algorithm of the class of fast algorithms that return solutions "good" . Between these two extremes have a class of algorithms that generate anytime solutions which, whilst not optimal, we are guaranteed that are within a limit of optimal solutions. Sandholm et al. [50] were the first to introduce an anytime algorithm for coalition structure generation that sets limits on the quality of solution found.

4.3 Teacher Assignment with Coalition Formation

The teacher assignment can be formalized with a coalition formation possibility, defining:

- Individual utilities for each agent belonging to the coalition that teaches each course.
- Utilities for each agent belong to a coalition with determined teachers.
- Utilities for total coalitions that each teacher belongs, e.g., number of courses that he teaches.

In computation, the entities that make up the coalition are typically autonomous software agents.

5 Conclusions

The area of multi-agent systems has maintained increasing attention from the scientific community with a growing number of works applying MAS to distinct problems. On the other hand although the scheduling problems and coalition formation are not new area, the scientific community has been paying close attention to them in the last years, with several papers about coalition formation for solving scheduling problems appearing in recent international conferences. The originality of this work is mainly concerned with the fact that we propose solving the teacher assignment allowing coalitions between the various agents involved. This model enables solving the TAP as a scheduling problem but using the typical and natural coalition formation that arises in the human solutions of this problem.

As future work, we intend to develop a fully automatic system for teacher assignment that allows solving this kind of problem using the proposed model in any type of higher education institution.

References

1. Gunawan, A., Ng, K.M., Ong, H.L.: A Genetic Algorithm for the Teacher Assignment Problem for a University in Indonesia. Information and Management Sciences 19(1), 1–16 (2008)
2. Wang, Y.Z.: An application of genetic algorithm methods for teacher assignment problems. Expert Systems with Applications, 295–302 (2002)
3. Reis, L.P., Oliveira, E.C.: A language for specifying complete timetabling problems. In: Burke, E., Erben, W. (eds.) PATAT 2000. LNCS, vol. 2079, pp. 322–341. Springer, Heidelberg (2001)
4. PATAT, Conferences: The International Series of Conferences on the Practice and Theory of Automated Timetabling (PATAT), http://www.asap.cs.nott.ac.uk/patat/patat-index.shtml
5. International timetabling competition, http://www.cs.qub.ac.uk/itc2007/index_files/overview.htm
6. Thompson, J., Dowland, K., et al.: General Cooling Schedules for a Simulated Annealing Based Timetabling System. In: Proceedings of the 1st International Conference on the Practice and Theory of Automated Timetabling (1995)
7. Wren, A.: Scheduling, Timetabling and Rostering — a special relationship? In: Proceedings of the 1st International Conference on the Practice and Theory of Automated Timetabling, pp. 474–495 (1995)
8. Queirós, F.H.: Construção automática de Horários de Aulas. Tese de Mestrado, Universidade Portucalense (1995)
9. Reis, L.P., Oliveira, E.C.: A Language for Specifying Complete Timetabling Problems. In: Burke, E., Erben, W. (eds.) PATAT 2000. LNCS, vol. 2079, p. 322. Springer, Heidelberg (2001)
10. Gröbner, M., Wilke, P.: A General View on Timetabling Problems. In: 4th International Conference on the Practice and Theory of Automated Timetabling, PATAT 2002 (2002)

11. Chan, H.W.: School Timetabling Using Genetic Search. In: 2nd International Conference on the Practice and Theory of Automated Timetabling, PATAT 1997 (1997)
12. Fang, H.L.: Genetic Algorithms in Tametabling Problems. PhD Thesis, University of Edinburgh (1994)
13. Fernandes, C.: Infected Genes Evolutionary Algorithm for School Timetabling. In: WSES International Conference (2002)
14. Eley, M.: Ant Algorithms for the Exam Timetabling Problem. In: Burke, E.K., Rudová, H. (eds.) PATAT 2007. LNCS, vol. 3867, pp. 364–382. Springer, Heidelberg (2007)
15. Moreira, J.J.: A System for Automatic Construction of Exam Timetable Using Genetic Algorithms. Polytechnical Studies Review VI(9) (2008)
16. Moreira, P.M., Reis, L.P., De Sousa, A.A.: I-om: Intelligent optimization for computer graphics and visualization. In: Proceedings of the 5th Iberian Conference on Information Systems and Technologies, CISTI 2010 (2010)
17. Faria, B.M., Reis, L.P., Lau, N., Castillo, G.: Machine Learning algorithms applied to the classification of robotic soccer formations and opponent teams. In: IEEE Conference on Cybernetics and Intelligent Systems, CIS 2010, pp. 344–349 (2010)
18. Ramos, J.A., Reis, L.P., Pedrosa, D.: Solving Heterogeneous Fleet Multiple Depot Vehicle Scheduling Problem as an Asymmetric Traveling Salesman Problem. In: Antunes, L., Pinto, H.S. (eds.) EPIA 2011. LNCS (LNAI), vol. 7026, pp. 98–109. Springer, Heidelberg (2011)
19. Gunawan, A., Ng, K.M.: Solving the teacher assignment problem by two metaheuristics. International Journal of Information and Management Sciences 22, 73–86 (2011)
20. Ferreira, P.S., Karas, E.W., Palucoski, F.L., Ribeiro, A.A., Silva, A.L.: Aplicação deprogramação inteira na distribuição de encargos didáticos em instituições de ensino. Tendências em Matemática Aplicada e Computacional 12, 135–144 (2011)
21. Hultberg, T.H., Cardoso, D.M.: The teacher assignment problem: A special case of the fixed charge transportation problem. European Journal of Operational Research 101, 463–473 (1997)
22. Uney-Yuksektepe, F., Karabulut, I.: Mathematical programming approach to course-teaching assistant assignment problem. In: Proceedings of the 41st International Conference on Computers & Industrial Engineering, pp. 878–883 (2011)
23. Tilleti, P.I.: An operations research approach to the assignment of teachers to courses. Socio-Economic Planning Sciences 9, 101–104 (1975)
24. Schniederjans, M.J., Kim, G.C.: A goal programming model to optimize departmental preference in course assignments. Computers & Operations Research 14, 87–96 (1987)
25. Hmer, A., Mouhoub, M.: Teaching Assignment Problem Solver. In: García-Pedrajas, N., Herrera, F., Fyfe, C., Benítez, J.M., Ali, M. (eds.) IEA/AIE 2010, Part II. LNCS, vol. 6097, pp. 298–307. Springer, Heidelberg (2010)
26. Maria, L.: Teacher Assignment Program. Master of Computer Science, University of Colorado (2010)
27. Silva, D.C., Braga, R.A.M., Reis, L.P., Oliveira, E.: Designing a meta-model for a generic robotic agent system using Gaia methodology. Information Sciences 195, 190–210 (2012)
28. Mota, L., Reis, L.P., Lau, N.: Multi-robot coordination using Setplays in the middle-size and simulation leagues. Mechatronics 21(2), 434–444 (2011)
29. Pereira, A., Reis, L.P., Duarte, P.: EcoSimNet: A Multi-Agent System for Ecological Simulation and Optimization. In: Lopes, L.S., Lau, N., Mariano, P., Rocha, L.M. (eds.) EPIA 2009. LNCS (LNAI), vol. 5816, pp. 473–484. Springer, Heidelberg (2009)
30. Pereira, A., Reis, L.P., Duarte, P.: Ecosimnet: A framework for ecological simulations. In: Proceedings - 23rd European Conference on Modelling and Simulation, ECMS 2009, pp. 219–225 (2009)

31. Azevedo, J., Oliveira, M., Pacheco, P., Reis, L.P.: A Cooperative CiberMouse@RTSS08 Team. In: Lopes, L.S., Lau, N., Mariano, P., Rocha, L.M. (eds.) EPIA 2009. LNCS (LNAI), vol. 5816, pp. 251–262. Springer, Heidelberg (2009)
32. Mota, A., Castro, A.J.M., Reis, L.P.: Recovering from Airline Operational Problems with a Multi-Agent System: A Case Study. In: Lopes, L.S., Lau, N., Mariano, P., Rocha, L.M. (eds.) EPIA 2009. LNCS (LNAI), vol. 5816, pp. 461–472. Springer, Heidelberg (2009)
33. Gimenes, R., Silva, D.C., Reis, L.P., Oliveira, E.: Flight simulation environments applied to agent-based autonomous UAVS. In: ICEIS 2008 - 10th International Conference on Enterprise Information Systems, Proceedings SAIC, pp. 243–246 (2008)
34. Certo, J., Lau, N., Reis, L.P.: A generic strategic layer for collaborative networks. In: Camarinha-Matos, L., Afsarmanesh, H., Novais, P., Analide, C. (eds.) Establishing the Foundation of Collaborative Networks. IFIP, vol. 243, pp. 273–282. Springer, Boston (2007)
35. Pereira, A., Duarte, P., Reis, L.P.: Ecolang - A communication language for simulations of complex ecological systems. In: Simulation in Wider Europe - 19th European Conference on Modelling and Simulation, ECMS 2005, pp. 493–500 (2005)
36. Reis, L.P., Lau, N.: COACH UNILANG - A Standard Language for Coaching a (Robo)Soccer Team. In: Birk, A., Coradeschi, S., Tadokoro, S. (eds.) RoboCup 2001. LNCS (LNAI), vol. 2377, pp. 183–192. Springer, Heidelberg (2002)
37. Reis, L.P., Lau, N.: FC Portugal Team Description: RoboCup 2000 Simulation League Champion. In: Stone, P., Balch, T., Kraetzschmar, G.K. (eds.) RoboCup 2000. LNCS (LNAI), vol. 2019, pp. 29–40. Springer, Heidelberg (2001)
38. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by Simulated Annealing Science. New Series 220(4598), 671–680 (1983)
39. Glover, F.: Tabu Search. ORSA Journal of Computing 1, 190–206 (1989)
40. Resende, T.A., Feo, M.G.C.: Greedy Randomized Adaptive Search Procedure, GRASP (2007), http://mauricio.resende.info/MiniCursoGRASP.pdf
41. Holland, J.: Scheduling, Adaptation in Natural and Artificial Systems. The University of Michigan Press (1975)
42. Hansen, P., Mladenović, N.: Variable neighborhood search, vol. 24, pp. 1097–1100. Elsevier Science Ltd. (1997)
43. Goss, S.: Self-Organized Shortcuts in the Argentine Ant. Naturwissenschaften 76, 579–581 (1989)
44. Colorni, A., Dorigo, M., Maniezzo, V.: Distributed Optimization by Ant Colonies. In: Proceedings of the First European Conference on Artificial Life, pp. 134–142. Elsevier Publishing, Amsterdam (1991)
45. Hart, P., Nilsson, N., Raphael, B.: A formal basis for the Heuristic Determination of Minimun Cost Paths. IEEE Trans. Syst. Cybernetics 4(2), 100–107 (1968)
46. Eberhart, R.C., Kennedy, J.: A new optimizer using particle swarm theory. In: Proceedings of the Sixth International Symposium on Micromachine and Human Science, Nagoya, Japan, pp. 39–43 (1995)
47. Rasmussen, T.K.: Improving Particle Swarm Optimization by hybridization of stochastic search heuristics and Self-Organized Criticality. Doctoral thesis, University of Aarhus, Department of Computer Science, Aarhus C, Denmark (2002)
48. Rahwan, T.: Algorithms for Coalition Formation in Multi-Agent Systems. PhD thesis, University of Southampton (2007)
49. Skiena, S.S.: The Algorithm Design Manual. Springer, New York (1998)
50. Sandholm, T.W., Larson, K., Andersson, M., Shehory, O., Tohme, F.: Coalition structure generation with worst case guarantees. Artificial Intelligence 111(1-2), 209–238 (1998)

A Comprehensive Study of Crime Detection with PCA and Different Neural Network Approach

Ahmad Kadri Junoh[1] and Muhammad Naufal Mansor[2]

[1] Institute of Engineering Mathematics (IMK), University Malaysia perlis, Malaysia
[2] Intelligent Signal Processing Group (ISP), University Malaysia perlis, Malaysia
ahmadkadrijunoh@unimap.edu.my, apairia@yahoo.com

Abstract. Crime rate in Malaysia is almost in awareness stage. The centre for Public Policy Studies Malaysia reports that the ratio of police to population is 3.6 officers to 1,000 citizens in Malaysia. This lack of manpower sources ratios alone are not a comprehensive afford of crime fighting capabilities. Thus, dealing with these circumstances, we present a comprehensive study to determine bandit behavior with PCA and different neural network algorithm such as Elman Neural Network (ELMNN), Feed Forward Neural Network (FFNN) and Cascade-Forward Neural Network (CFNN). This system provided a good justification as a monitoring supplementary tool for the Malaysian police arm forced.

Keywords: Crime rate, Principal Component analysis, Elman Neural Network, Feed Forward Neural Network and Probabilistic Neural Network.

1 Introduction

Based on the 2012 Global Piece Index (GPI), Malaysia rates very well both regionally and worldwide with 1.485 GPI. Looking at crime categories, Malaysia is assigned the same, or better, ratings than these countries with rank 20 of 158 in around the world (lower scores being optimal) [1].

Nevertheless, the Royal Malaysian Police (RMP) reports that they employ 93,000 officers [2],[3], and the Centre for Public Policy Studies Malaysia reports that the ratio of police to population is 3.6 officers to 1,000 citizens and the ethnic composition of the police force is 78% Malay, 4% Indian, 2.5% Chinese, and 15% others [4]. For comparison purposes, Australia's ratio is 2.1 officers per 1,000 citizens [5]. Many sources caution that police to population ratios alone are not a comprehensive measure of crime fighting capabilities [6].

Thus within this drawbacks, we present a comprehensive study to determine crime behavior with PCA and different neural network algorithm such as Elman Neural Network, Feed Forward Neural Network and Probabilistic Neural Network. This system provided a good justification as a monitoring supplementary tool for the Malaysian police arm forced.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 611–618.
DOI: 10.1007/978-3-642-36981-0_56

1.1 Face Detection

Principal component analysis is quiet common in face recognition [7]-[9]. One of researcher as Jin Zhang [9], proposed a novel approach using PCA and SVM for face recognition. PCA is used to reduce dimension of sample data of CMU face database. PCA transform the feature vectors, which are used for training SVM classifier. The experimental results demonstrate that the proposed method is encouraging with a successful detection rate. However, some of researches such as [10]-[14] adopted different approach of neural network algorithm with PCA in face recognition. Furthermore, most of the researchers compromised a well known database in their study. In contrast, we refine a new database consists a crime scene. In order to justify the robustness, images were taken in different angle. In addition we compare our experiment with determine bandit behavior with PCA and different neural network algorithm such as Elman Neural Network, Feed Forward Neural Network and Cascade-Forward Neural Network.

1.2 System Overview

The proposed face recognition system consists recognition phases as depicted in Fig. 1. It consists of several modules which are, face detection, recognition, feature extraction and neural network classifier are performed.

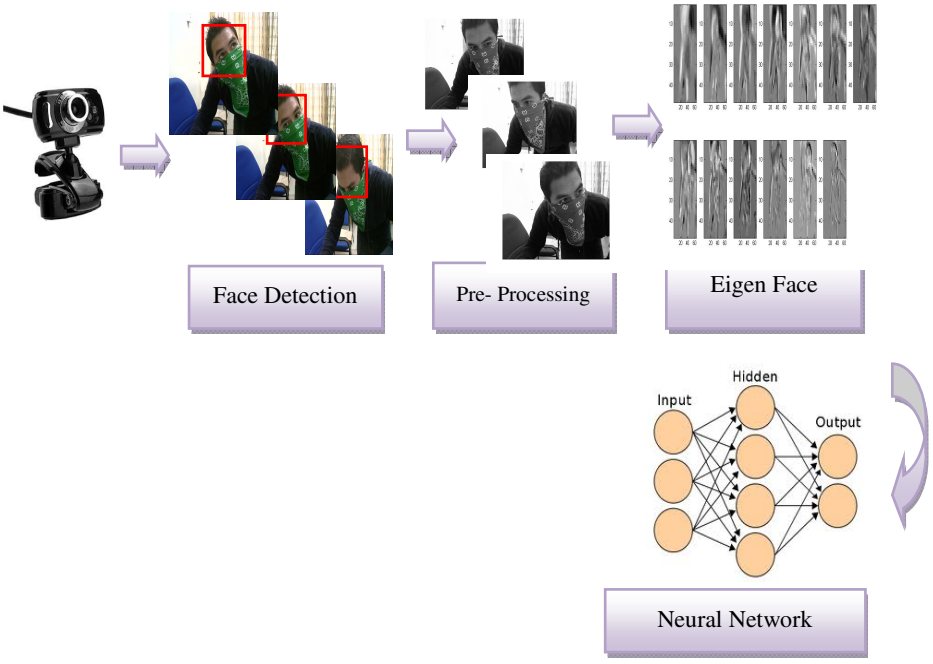


Fig. 1. Block Diagram of Face Recognition System

1.3 Recognition Phase

Recognition involves comparison with only those templates corresponding to claimed identity. The recognition phase involved several modules such as image acquisition, face detection, and face recognition. Some of bandit images are shown in Fig. 2.

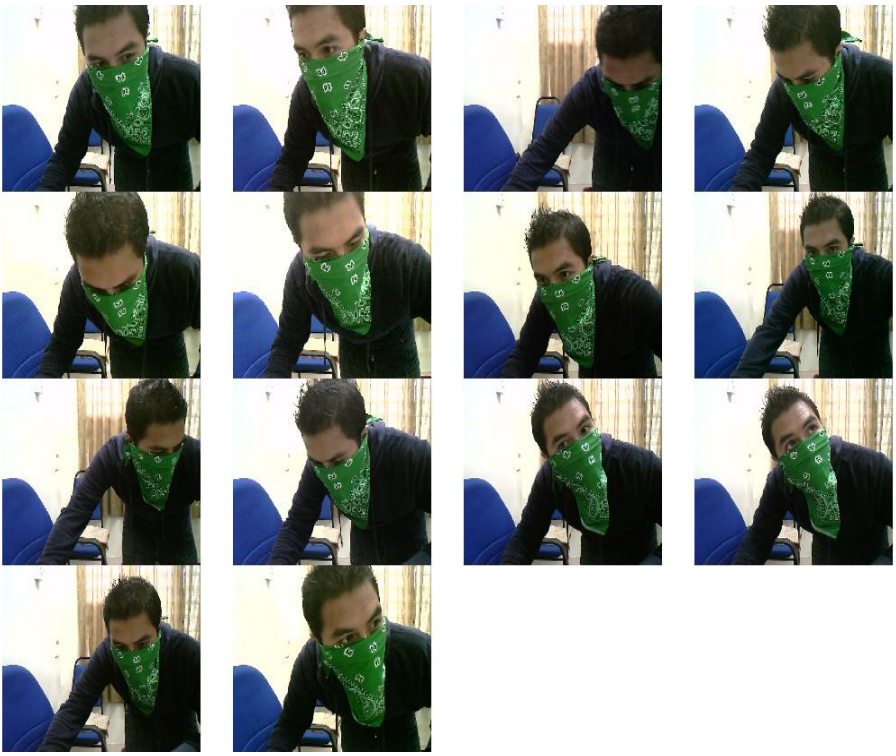


Fig. 2. Bandit Images

1.4 Face Recognition

The face recognition module contains of preprocessing, feature extraction, and classification sub-modules. The input to the face recognition/verification module is the face image, which is derived from two sources from the camera or from the database. During feature extraction, the normalized image is represented as feature vectors. The result of the classification for the recognition purpose is determined by matching the client index with the client identity in the database.

1.5 Feature Extraction

The feature extraction algorithms used is Principal Component Analysis (PCA) suggest by [15]-[18]. Principal component analysis (PCA) for face recognition is

based on the information theory approach. It extracted the relevant information in a face image and encoded as efficiently as possible. It identifies the subspace of the image space spanned by the training face image data and de correlates the pixel values. The classical representation of a face image is obtained by projecting it to the coordinate system defined by the principal components as shown in Fig. 3.

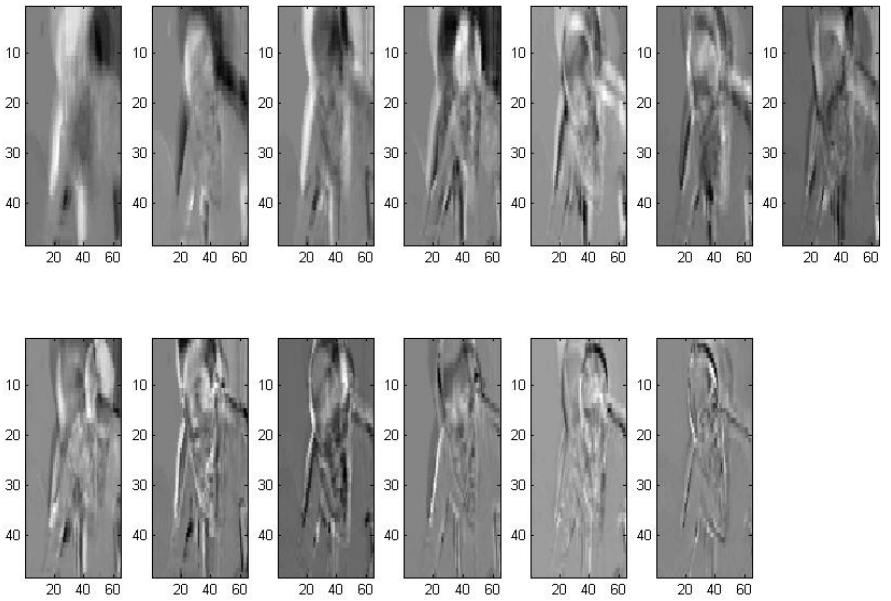


Fig. 3. Eigen Face of Bandit images

1.6 Classifier

There are many types of artificial neural networks (ANN). An artificial neural network is a computational simulation of a biological neural network [19]-[23]. To study the performance, we compare some neural network approach as Elman Neural Network, Feed Forward Neural Network and Probabilistic Neural Network.

2 Results and Analysis

The proposed algorithm was evaluated on a hundred and twenty subjects with different race, gender and age. The average size of each image is 400-500 pixels. The entire subjects were tested for ten trials with different neural network algorithm such as Elman Neural Network, Feed Forward Neural Network and Probabilistic Neural Network. Ten images were taken for each subject. Average accuracy different neural network algorithm for all subjects was shown in Table1 to Table 3.

Table 1. Average accuracy for Elman Neural Network

Number of Input neurons: 10		Number of hidden neurons:16
Number of output neurons:2		Momentum Factor: 0.9
Activation Function: Bipolar sigmoid		Training Tolerance: 0.01
Learning Rate: 0.5		Testing Tolerance: 0.1
Number of samples used for training: 96		
No. samples used for testing:24		
Total Samples:120		
Trail	Epoch	Classification Rate
	Mean Epoch for Training	Mean Classification Rate (%)
1	3459	90.00
2	2982	65.00
3	3278	77.50
4	2779	90.00
5	2912	83.75
6	4173	90.00
7	3008	90.00
8	2722	71.25
9	4264	96.25
10	3835	90.00
Average	3341	84.37

Table 2. Average accuracy for Feed Forward Neural Network

Number of Input neurons: 20		Number of hidden neurons:16
Number of output neurons:1		Momentum Factor: 0.9
Activation Function: Binary sigmoid		Training Tolerance: 0.01
Learning Rate: 0.5		Testing Tolerance: 0.1
Number of samples used for training: 96		
No. samples used for testing:24		
Total Samples:120		
Trail	Epoch	Classification Rate
	Mean Epoch for Training	Mean Classification Rate (%)
1	3718	83.75
2	3534	90.00
3	3031	83.75
4	3573	96.25
5	2159	87.50
6	3142	83.75
7	3593	96.25
8	1840	83.75
9	2007	77.50
10	3196	71.25
Average	2979	85.37

Table 3. Average accuracy for Cascade-Forward Neural Network

Number of Input neurons: 20		Number of hidden neurons:16
Number of output neurons:1		Momentum Factor: 0.9
Activation Function: Binary sigmoid		Training Tolerance: 0.01
Learning Rate: 0.5		Testing Tolerance: 0.1
Number of samples used for training: 96		
No. samples used for testing:24		
Total Samples:120		
Trail	Epoch	Classification Rate
	Mean Epoch for Training	Mean Classification Rate (%)
1	3474	90.00
2	3425	98.75
3	3344	77.50
4	3592	71.25
5	5000	92.50
6	3545	83.75
7	2626	71.25
8	3447	90.00
9	2161	52.50
10	3246	90.00
Average	3386	81.75

To evaluate the performance of the proposed method, we compare our bandit database with the Essex database. Based, from the table above it shows that PCA component analysis is still relevant for different neural network algorithm. It proved that FFNN gives the highest average detection accuracy at 85.37% compare to ELMNN and CFNN at 84.37% and 81.75% respectively.

3 Conclusion

In this paper, we proudly had proven that PCA and Neural network concept suitable in bandit recognition. Though, framework for a whole system has been developed in this paper, we concentrated only on database search mechanism. The proposed mechanism can be applied not only in crime monitoring but also in other surveillance system. This successfully provided a good justification as a monitoring supplementary tool for the Malaysian police arm forced.

Acknowledgments. This research was conducted under Fundamental Research Grant Scheme (FRGS) which is contributed by Ministry of Higher Education Malaysia.

References

1. Global Peace Index Vision of Humanity website, `http://www.Visionofhumanity.org/gpi/results/rankings.php` (accessed December 6, 2012)

2. RRT Country Advice 2010, Country Advice MYS35845 (January 11, 2010)
3. Royal Malaysian Police website, http://polismalaysia.brinkster.net/Royal%20Malaysian%20Police%20Force%20-%20About.asp (accessed December 6, 2012)
4. Center for Public Policy Studies Malaysia website, Policy Fact Sheet Polic, http://www.cpps.org.my/downloads/factsheets/Police%20factsheet.pdf (accessed January 4, 2010)
5. NationMaster.com website, Statistics Tab, Crime statistic – police per capita by country, http://www.nationmaster.com/graph/cri_pol_percap-crime-police-per-capita (accessed December 17, 2009)
6. International Association of Chiefs of Police website, Research Center Directorate Perspectives Police Officer to Population Ratios Bureau of Justice Statistics Data, http://www.theiacp.org/LinkClick.aspx?fileticket=LF7xdWl1tPk%3D&tabid=87(accessed December 21, 2009)
7. Cheriyadat, A.: Why principal component analysis is not an appropriate feature extraction method for hyperspectral data, vol. 6, pp. 3420–3422 (2003)
8. Park, M.S.: PCA-based Feature Extraction using Class Information, vol. 1, pp. 341–345 (2005)
9. Zhang, J., Zhang, X.-D., Ha, S.-W.: A Novel Approach Using PCA and SVM for Face Detection, vol. 3, pp. 29–33 (2008)
10. Jasmina, N., Milomir, M., Alempije, V.: Classification Accuracy of Neural Networks with PCA in Emotion Recognition. Theory and Applications of Mathematics & Computer Science, 11–161 (2011)
11. Eleyan, A., Demirel, H.O.: Face Recognition System Based on PCA and Feedforward Neural Networks. In: Cabestany, J., Prieto, A.G., Sandoval, F. (eds.) IWANN 2005. LNCS, vol. 3512, pp. 935–942. Springer, Heidelberg (2005)
12. Chitaliya, N.G.: Feature Extraction Using Wavelet-PCA and Neural Network for Application of Object Classification & Face Recognition, vol. 1, pp. 510–514 (2010)
13. Franco, L., Treves, A.: A neural network facial expression recognition system using unsupervised local processing, pp. 628–632 (2001)
14. Latha, P., Ganesan, L., Annadurai, S.: Face Recognition using Neural Networks. An International Journal (SPIJ) 3, 153–160 (2008)
15. Mansor, M.N., Yaacob, S., Hariharan, M., Basah, S.N., bin Ahmad Jamil, S.H.-F.S., Khidir, M.L.M., Rejab, M.N., Ibrahim, K.M.Y.K., bin Ahmad Jamil, A.H.-F.S., Junoh, A.K.: PCA - Based Feature Extraction for Preterm Birth Monitoring. Progress in process tomography & instrumentation system Series 8. Universiti Teknologi Malaysia (2012) ISBN 978-967-354-186-7
16. Mansor, M.N., Yaacob, S., Hariharan, M., Basah, S.N., bin Ahmad Jamil, S.H.-F.S., Khidir, M.L.M., Rejab, M.N., Ibrahim, K.M.Y.K., bin Ahmad Jamil, A.H.-F.S., Junoh, A.K.: PCA- Based Feature Extraction and k-NN algorithm for Early Jaundice Detection. Journal of Soft Computing and Software Engineering (2012) ISSN: 2251-7545
17. Mansor, M.N., Yaacob, S., Hariharan, M., Basah, S.N., bin Ahmad Jamil, S.H.-F.S., Khidir, M.L.M., Rejab, M.N., Ibrahim, K.M.Y.K., bin Ahmad Jamil, A.H.-F.S., Junoh, A.K.: PCA- Based Feature Extraction and LDA algorithm for Preterm Birth Monitoring. Journal of Soft Computing and Software Engineering (2012) ISSN: 2251-7545
18. Junoh, A.K., Mansor, M.N., Yaacob, S., Mukhtar, A.Z., Fauzi, S.F., Abu, M.S., Ahmad, W.Z.A.W., Jaafar, W.N.W.: Smart Surveillance System Detection Using PCA- Based Feature Extraction. Applied Mechanics and Materials Journal (2012) ISSN: 1660-9336

19. Werbos, P.J.: Generalization of backpropagation with application to a recurrent gas market model. Neural Networks 1 (1988)
20. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation 9(8), 1735–1780 (1997)
21. Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing 45, 2673–2681 (1997)
22. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks 18, 602–610 (2005)
23. Schmidhuber, J.: Learning complex, extended sequences using the principle of history compression. Neural Computation 4(2), 234–242 (1992)

A Conceptual Model of Layered Adjustable Autonomy

Salama A. Mostafa[1], Mohd Sharifuddin Ahmad[2], Muthukkaruppan Annamalai[3], Azhana Ahmad[2], and Saraswathy Shamini Gunasekaran[1]

[1] College of Graduate Studies, Universiti Tenaga Nasional, Putrajaya Campus 43000, Selangor Darul Ehsan, Malaysia
[2] College of Information Technology, Universiti Tenaga Nasional, Putrajaya Campus 43000, Selangor Darul Ehsan, Malaysia
[3] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor Darul Ehsan, Malaysia
semnah@yahoo.com, {sharif,azhana,sshamini}@uniten.edu.my, mk@tmsk.uitm.edu.my

Abstract. Autonomy and autonomous agents are currently the most researched topics in autonomous systems. Issues like autonomy adjustment, autonomy level, and the required degree of autonomy to be performed are investigated. Abstracting an autonomy model poses the problem of identifying specific aspects that merit an autonomous system. In this paper, we propose another model of autonomy that conceptualizes autonomy as a spectrum, which is constructed in a layered structure of a multi-agent environment called Layered Adjustable Autonomy (LAA). The autonomy spectrum of the LAA is divided into adjustable-leveled layers. Each of which has distinct attributes and properties that assist an agent in managing the influences of the environment during its decision-making process. The LAA structure is designed to endorse an agent's qualification to make a decision by setting the degree of autonomy to the agent's choice of decision-making. An Autonomy Analysis Module (AAM) is also proposed to control and delegate the agent's actions at specific autonomy levels. Hence, the AAM determines the threshold of the agent autonomy level to act in its qualified layer. Ultimately, the proposed LAA model will be implemented on an air drone for the purpose of testing and refinement.

Keywords: Software agent, Multi-agent system (MAS), adjustable autonomy, autonomous systems, Layered Adjustable Autonomy (LAA), decision-making.

1 Introduction

Many researches in agent-based systems have attempted to design intelligent and autonomous software that decides on a course of actions to achieve its goals without being explicitly programmed to do so (e.g. [1] and [2]). Others have repeatedly acknowledged the limiting capabilities of classical algorithms for automated planning (e.g. [3] and [4]). The implementation of autonomy in most agent-based systems is merely confined to simple behaviors that limit agents to perform non-critical tasks [5], [6].

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 619–630.
DOI: 10.1007/978-3-642-36981-0_57

Brainov and Hexmoor [7] emphasised that within social networks there exist a measurable degree of autonomy between an agent and its surrounding environment. "Adjustable autonomy implies that the level of autonomy in the decision-making process can be adjusted." [8]. Therefore, we define *an adjustable autonomous agent as an agent that can show different degrees of autonomy during its decision-making process cycle as a result of external influences.*

The degree of autonomy for an agent can be signified by the degree of intervention in its decision-making process by interveners [7]. The intervener can be humans or other agents and the intervention can either be within the system or by other system (e.g. in the case of distributed systems). The main cause of the intervention (or the adjustment) in the agent decision-making process is the lack of its local knowledge.

In this paper, we propose a Layered Adjustable Autonomy (LAA) model, which provides a coordinated mechanism that works in a Multi-agent System (MAS) environment. It enables an agent to perform at different layers of autonomy levels to achieve its goal. Layer selection condition is fulfilled based on the selected action type. If the action is critical and/or complex, then the system categorizes it as a high level action. This type of action is assigned to the layer with the highest level of autonomy, where the qualified agent must have the knowledge and the authority to make a decision.

2 Research Objectives

The adjustable approach to autonomy can help to increase system's reliability and reduce system's complexity [5]. Scerri and Reed [6] claimed that the key problem of building adjustable autonomy is "*to determine an appropriate distribution of autonomy and provide mechanisms to realize the autonomy changes.*" Consequently, the aim of this research is to construct a flexible and usable autonomous model by proposing the LAA model mechanism. It enhances agent decision-making quality and robustness by analysing the required level of autonomy for an agent to be accordingly updated. To achieve the aim of the research, the following objectives are proposed:

1. To introduce the concept of layers in adjustable autonomy.
2. To develop a model for autonomous system using the proposed LAA.
3. To refine the related concepts of autonomy to satisfy the proposed model.

3 Issues in Adjustable Autonomy

Building desirable and trusted autonomous systems is a prime concern for both researchers and implementers [9]. A fully autonomous agent *is an agent that can make its own decision based on its observations without any type of intervention.* In some circumstances, such agent behaviour is unacceptable, which might results in serious consequences [6]. This leads to the necessities of external intervention to such circumstances mainly by humans, hence, a need for adjustable autonomy. Some examples of recently developed systems that utilized the approaches of adjustable autonomy are [9], [10] and [11].

3.1 The Approaches of Adjustable Autonomy

The changes made to a system by an agent manifest it as an autonomous intelligent entity. As a result, some system activities which are not fixed at design time are handled by the agent to best configure the system's functionality [12]. The adjustable autonomy gives the system flexibility to work in different levels of autonomy.

The teamwork-centred adjustable autonomy is an approach that structures and draws the margins of the actors of a system to work in a cohesive unit as teamwork [12]. Basically, the system gives the operating controls for specific activities of the system to the agent and other specific activities to the human so as to work in a dynamic manner.

In another approach to adjustable autonomy, an agent control most of the system's activities but humans (or any third party systems) monitor and maintain (or 'adjust') the system's performance whenever is necessary to avoid its failure [13]. For instance, the decision of critical, uncertain, or unseen situation is transferred from the agent to a human based on certain conditions [14]. Furthermore, an adjustment can be made on the agent itself by the human either directly by adjusting its parameters through a GUI or indirectly by adjusting some activities of the system [2].

3.2 Challenges in Modelling Adjustable Autonomy

While adjustable autonomy is considered as a successful model, it shows some defects (complex and costly) that are crucial especially in systems where many players are involved in its control [15]. Apparently, the ensuing continuous interrupts make such systems dependent and slow [16], especially in systems where a dialogue is utilized in the sequential decision problem e.g. mixed-initiative approach [9]. The dependency has positive impact by increasing system initiative level and negative impact in how the decision is accurately figured.

Researchers have contributed some enhancements to the design of adjustable autonomy. For example, Cohen and Fleming [15] suggested two scenarios to switch control to humans: the first scenario is that the agent internally determines when to switch control to humans i.e. mixed-initiative. The second scenario, the decision of switching the mode is fixed externally by humans. Dorais et al. [14] set teamwork policies to manage control transference among system's team members that are certified to do the adjustment. Nevertheless, most of the enhancements are either in a high complexity [16], unstructured [4] and/or fixed domains [10].

3.3 Related Works in Adjustable Autonomy

As deliberated in the previous sections, there are many opinions and diverse understanding of what autonomous agent and adjustable autonomy are. As a result, different approaches are proposed to resolve some of the adjustable autonomy problems. In this section, we review some of related works in autonomy and adjustable autonomy. As a part of this research, we come up with some conclusion that explains the meaning of the word 'adjustability' in the autonomy paradigm.

Brooks [17] is one of the leading researchers who conducted research in autonomy while drawing up the roadmap of intelligent mobile robots in the early 80's. He proposed the layers of control to the mobile robots in the well-known reactive architecture. The reactive architecture consists of parallel and independent modules that are responsible for specific tasks. Each module is independently interfaced to the world rather than to each other. Communication between the modules is minimized and of a low-level nature. Reactive agents act in a stimulus response manner to the present state of the environment in which they are embedded.

A group of researchers from the National Aeronautic and Space Administration (NASA), USA ([5], [12], [13]) used adjustable autonomy for human-agent groups involving a mix of humans and autonomous agents (i.e. Mixed-Initiative adjustable autonomy). They used human-centred perspective to human-agent interaction by integrating the Brahms and KaoS agent frameworks. The aim is to model and simulate realistic work situations in space and to support the design of robotic and software agents for human-agent teamwork. They include in their model some considerations such as scheduled and unscheduled activities, work practices emergence and resource availability. In addition, they modelled a mechanism of interaction between human and machine that enabled the system to work in a collaborative manner.

Situation awareness is a type of reasoning methodology that is used in autonomous systems for assessments during adjustable autonomy planning step [10]. It is divided into three levels, situation perception, situation comprehension and situation projection. Situation awareness reasons the given situation information and derive decisions to counter the situation. However, in the recent research made by Yin Lili et al. [10] in the situation awareness of adjustable autonomy, they presented an adjustable autonomy module named Situation Reasoning Module (SRM). SRM is tested using Unmanned Surface Vehicles (USV) in a navigation mission. The results show that SRM implementation helps in calculating the autonomy level regarding the situation which leads to increased system efficiency.

In a recent research by NASA, Dumond et al. [11] proposed a system based on the adjustable autonomy for humans-agents' coordination and cooperation. They developed the CHAMP model (Coordinating with Humans by Adjustable-autonomy for Multirobot Pursuit) that is simulated for military purposes. The challenge of the CHAMP model is to operate in an environment that incorporates small-unit tactical team and semi-autonomous robots team effectively. Both teams are dynamically controlled via distributed optimization strategies to perform the coordination and operated via the adjustable autonomy. The aim of the CHAMP model is to dynamically adjust the coordination of the robots team towards goal achievement through the dynamics of the situations.

Based on what we reviewed, agent autonomy adjustment can be internally made based on the agent's desire or given by a third party. We can then conclude that autonomy adjustment of an agent indicates the need of a mechanism that specifies and controls the adjustment process. Consequently, the adjustment as a process is strongly related to an action type while the action type is determined based on the characteristics of the event. Thus, *the adjustment, based on specific mechanism,*

changes the parameters of the decision-making process related to the given event so as to influence the decision and the final outcome.

4 The LAA Model

We propose the Layered Adjustable Autonomy (LAA) to manage some of the autonomy adjustment issues. Those issues include software agent and human-agent autonomy management, the corresponding decision-making management and the dynamic autonomy adjustment in a MAS setting.

Adjustable autonomy means that an autonomous agent has a certain variable level of ability to act on its own [18]. Consequently, the LAA separates the autonomy into number of layers that are used to encapsulate an agent's decision process adjustment. A layer is defined as *a virtual location that accepts specific set of actions and assists the agent to implement the actions toward an optimized decision*. Each layer is attributed to deal with actions that correspond to its autonomy level.

4.1 Autonomy Layers

In the LAA model, the agents must follow certain rules regarding the responses to an event e. The agents' knowledge, authorities, responsibilities and roles have different attributes and properties. Therefore, in the model, the agent A_i is considered as a qualified agent (Aq_i), $Aq_i \in A_i$, to make a decision d, if and only if A_i satisfies two conditions:

1. The agent A_i has the knowledge to make the decision (i.e., *know* is true), and
2. The agent A_i has the required level of authority to make the decision (i.e., *can* is true).

We submit that, in general, when these two conditions are satisfied, then the agent A_i is a qualified agent Aq_i and has the autonomy to decide.

$$If\ (A_i\ know{=}{=}\text{true})\ \text{and}\ (A_i\ can{=}{=}\text{true}) \rightarrow A_i : Aq_i$$

For the purpose of *know* and *can* measurements, we suggest that *know* represents the agent's ability to plan an action and successfully executing it and *can* represent the consistency of the agent to successfully execute the action repeatedly. While *can* denotes the mandate for an agent to perform the action imposed by an authorized agent, there are open actions, which can be autonomously performed by any agents (see Section 4.4). Correspondingly, the *know* and *can* preconditions of the layers' truth satisfaction are graded differently as different actions have different autonomic constraints. However, the details of the measurements are beyond the scope of this paper as it is still in progress.

The degree of adjustment that acts in a layer is fixed based on the layer's status. The autonomy of the qualified agent that is active in the specific layer is adjustable based on the attributes of the layer. The layer contains two modules that determine its attributes, which are Data Management Module (DMM) and Process Management Module (PMM) as shown in Figure 1.

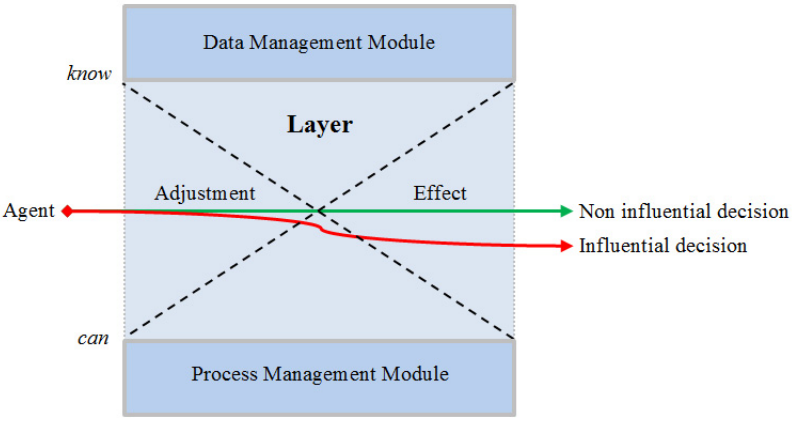


Fig. 1. The concept of the layer in the LAA model

The primary task of the DMM is to manage the flow of input and output data during the agent's decision making process while the primary task of the PMM is to assist the decision making process by performing processes related to action's implementation. The numbers and the types of processes that can be performed in a layer to execute an action represent the layer's compliance to the qualified agent's decision-making ability. While some of the layers' processes overlap, others have distinct properties.

4.2 The Structure of the LAA Model

We conceive the novel concept of layered autonomy as follows: Some of the decision-making abilities regarding a specific event can be switched from one agent to another qualified agent depending on whether the autonomy level of the first agent fulfil the higher layer's threshold. The qualified agent is the one that has the knowledge and authority to handle an event. Otherwise, if there are no other qualified agents to make the decision, the system requests human interventions [19]. The structure of the LAA model follows the typical steps of agent-based systems from sensing, deliberating through reasoning and action selection to action's implementation. It extends the layers as channels of actions' processing constraints and the Autonomy Analysis Module (AAM) as a controller to layers' selection as shown in Figure 2.

The AAM is a management module that handles dynamic autonomy distribution. In a given situation, it is responsible for ensuring that the combination of the agent, the autonomy level and the layer are compatible with the selected action. This module updates the agent's mental states to use the corresponding autonomy components as required, which enables critical decisions-making. Preconditions and postconditions of the combination must be fulfilled during the autonomy analysis period. When the agent fulfils its requirements of the layer, it is entertained by the DMM and PMM services.

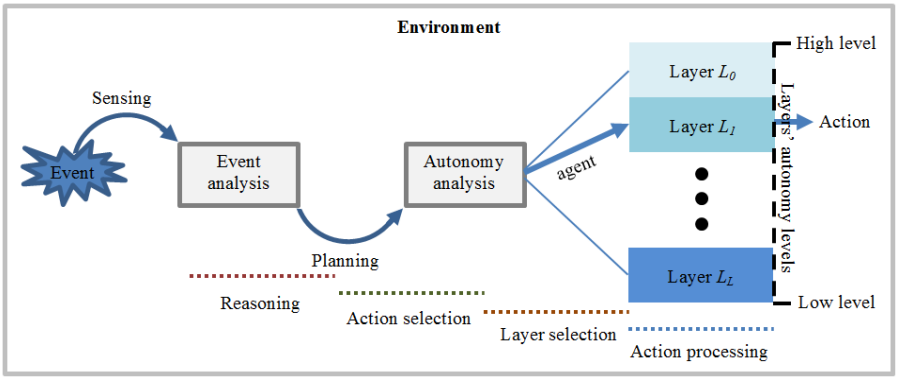


Fig. 2. The proposed LAA model structure

The agent at the high autonomy level has the ability to control the external influences such as other agents' messages. It utilizes its reasoning rules based on its own observation of the external influences before making its decisions (this agent can autonomously act in the Layer L_0 and L_1). On the contrary, the agent with a lower autonomy level accepts some of the external influences as facts that directly affect its decision-making process (this agent can only autonomously act in the lower Layers) [8]. In addition, the agent at the higher autonomy level needs and contributions are different from those of the agent at the lower level (which the corresponding layers of each ensure to provide). Furthermore, it has more responsibilities and higher authority than the agent at the lower autonomy level (that is why it can act at the top layers) [6]. Thus, such agent is more likely situated at the top layers of the LAA model and able to autonomously perform critical decisions.

4.3 LAA Actions and Layers

In this scenario, we classify the actions of a system based on its functionalities into three levels: high, intermediate and low level actions [10]. Consequently, in this setting, we assume that each level of the actions is assigned to its specific autonomy layer (see Figure 3) and each actor of the system (e.g. human or agent) has a set of possible actions along with their autonomy properties [8]. Particular resources and autonomy properties of a layer are different from other layers such as the adjustment mechanism of actions' parameters.

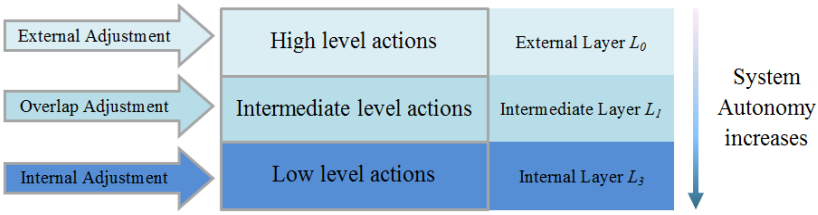


Fig. 3. The actions and the autonomy layers

The high level actions represent the achievement of a goal that is externally assigned (e.g. human command to track an object by a drone). The easiest way to implement the interaction with the system is by utilizing some GUI if the agent is human [19]. Otherwise, in the case of non-human agent, intermediate software is required to give the commands and to perform parameters' adjustments [20]. High level actions and their adjustment parameters must be judiciously assigned due to the state of external authorization. This type of actions corresponds to the external layer and promotes external adjustments. The only intervention the system can make in this layer is due to force majeure event (e.g. the drone changes its direction to avoid an attack by activating an override plan).

The intermediate actions are those that assist in consistently performing both the high level and the low level ones (e.g. minding actions, such as agent activates the override plan to avoid the attack). The intermediate layer of autonomy handles the adjustment of the intermediate level actions. The adjustment decision is made by the consideration and assistance of an external agent. Thus, intermediate level actions involve analysing the assigned high level actions and their related parameters and linking them with the low level actions during the system's planning stage.

The low level actions are the basic actions that are responsible for the system's self-control and management (e.g. manoeuvring the drone after the human gives the command to track an object). These actions are constrained to system change. Low level actions are in the internal layer zone and subjected to internal adjustments. Thus, the responsible agent makes less influential decisions and with extreme control.

4.4 An Example Scenario

This section represents an attempt to clarify and evaluate the LAA model by providing pragmatic approach for specifying and verifying the model's efficiency. We acknowledge here that the intelligent agent has the characteristics of autonomy, reactivity, pro-activity, goal-directedness, embodiments, sociality and rationality [21].

There exists a company, C, that consists of n actors, Act_n: Manager, Engineer, Technician, Security Guard and Other-employees. Assume that the Security Guard discovered a defective automatic door at the main entrance of the company's building. A decision, *d*, is required for the defective door event, *e*. Based on the event *e*, some actors are directly involved in making the decision *d*. Some of them are responsible for implementing the required action, *a*, which is fixing the door, while others are affected by the event, e.g. Other-employees cannot use the door.

In LAA, for the qualified actor, Act_i, to make a decision, *d*, he/she must satisfy two conditions:

- The actor, Act_i has the knowledge to make the decision (*know* is true), and
- The actor, Act_i has the authority to make the decision (*can* is true)

If these conditions are true, then the actor, Act_i has the autonomy to decide (i.e. $know \wedge can \rightarrow decide$). In practice, however, to handle an event usually requires corresponding decisions (D_m) and corresponding actions (A_m). Consequently, the four main players in this scenario are Manager, Engineer, Technician and Security Guard. The possible corresponding decisions and actions to handle *e* are suggested in Table 1.

Table 1. The possible decisions and actions to the scenario

Actors	Time	Decisions	Actions
Security Guard	T_1	D_1: Inform all actors Act_n about the problem.	A_1: Put an announcement near the door.
Manager	T_2	D_2: Inform the Engineer to diagnose the problem.	A_2: Send a message to the Engineer.
Engineer	T_3	D_3: Replace door sensors.	A_3: Send a message to the Technician.
Technician	T_4	D_4: Acknowledge the Engineer's decision.	A_4: Replace door sensors.
Security Guard	T_5	D_5: Inform all actors Act_n that the problem is solved.	A_5: Remove the announcement.

However, the diagnosis process may contain many related actions to be performed (i.e. pre-actions). For instance, The Security Guard needs to inspect and confirm that the door is not functioning and the Engineer needs to know where the door is and inspect the door before diagnosing it. Therefore, to satisfy a decision, one or more actions represent the agent intentions while other actions assist to achieve the intentions and the sequence of the overall actions represent the plan [3].

The event handling cycle can be shortened to specific sequence based on the actors' proactive behaviour [21]. Some decisions such as informing the actors can be made by any actor because it is low level action and the two conditions of qualifications (*know* and *can*) are easily satisfied. For instance, if the Engineer him/herself detects the door is not functioning and he/she autonomously decides to fix the door and the Technician is influenced by the Engineer's decision and respond accordingly, then only T_3, T_4 and T_5 are required to handle the defective door event. Hence, the LAA efficiency is clearly visible as it is concerned with adjusting the autonomy level of the actors to manifest a fixable and reliable autonomy model.

The following algorithm illustrates Table 1 implementation based on the LAA:

((Define-Actors: (Act_n: Manager, Engineer, Technician,
  Security Guard, Other-employees)
(Define actors' attributes: (a, b, c,...))
(Define attributes' properties: (1, 2, 3,...))
(T_0: (Act_i fetch an event *e*))
(T_1: (Act_i inform all
(T_2: Act_n sequence acting pattern
(Engineer: *know* $\wedge$ *can* $\rightarrow$ True)
(T_3: (Engineer make D_3, process A_3 of *e*)
(Technician: *know* $\wedge$ *can* $\rightarrow$ True)
(T_4: (Technician make D_4, process A_4 of *e*)
(T_5: (Act_i Inform all)))))))

5 Discussions

To construct an adjustable autonomous system, two major issues are involved: the agent and the adjustable autonomy. There are many attempts to consolidate the agent and the adjustable autonomy to manifest an autonomous system. However, adjustable autonomy as a mechanism is not clearly illustrated in the literature. The concept is strewn between the agent and the system design to fit the exact case rather than being generalized. Some attempts in modelling autonomy adjustment either have static nature or fixable configuration mechanism (e.g. a slider).

In the LAA model, a layer is like a workshop for the agent to find the required tools and materials that enables it to perform an action. The agent that controls specific action processing in specific autonomy layer must *know* about the action processing and *can* work in the layer that meets the action's level. Therefore, fewer interruptions are guaranteed and indecisive cases are more likely reduced as illustrated in Section 4.4. Thus, by determining the layer of autonomy for each action performance, the overall system efficiency is enhanced.

From the given example in Section 4.4, the Manager in many cases has fulfilled the autonomy level condition *can* that enables him/her to make the decisions. However, in the MAS environment, agent reasons on its beliefs and then builds its desire to respond. As a result, the agent with a minimum qualification is able to participate in the decision-making process. Furthermore, to accomplish the agent's intentions, some agent behavioural responses must be proactive in nature. In certain circumstances, however, the actors are not able to act on their decision even though they are competent to do so due to some constraints. For example, the Technician cannot act because the sensors are not available (i.e. an issue of situation awareness).

We argue that the layer's attributes and properties are domain-specific. In addition, the *can* condition fulfilment requires some authority and special skills attributes. As a result, the layered concept can be extended to be used for security purposes in the autonomous agent domain. For instance, accessing a secured layer requires highly attributed authority and some special properties such as a password key or a certain level of trust. However, such requirements are beyond the scope of this work.

6 Conclusion and Future Work

The Layered Adjustable Autonomy (LAA) as a model attempts to provide clues for managing adjustable autonomy in the MAS using autonomy analysis module. LAA separates the autonomy into layers. Reaching a specific layer by an agent is determined by the layer's and the agent's attributes to achieve an efficient fixable and reliable autonomous system.

In our future work, actions structure, autonomy measurements along with the actions processing mechanism that suit the LAA model will be proposed. Additionally, studying the possibility of merging LAA architecture with the Belief-Desire-Intention (BDI) architecture is one of the key aspects of this research. Ultimately, a prototype of the proposed model will be developed using an air drone for testing purposes.

Acknowledgments. This project is sponsored by the Malaysian Ministry of Higher Education (MoHE) under the Exploratory Research Grant Scheme (ERGS) No. ERGS/1/2012/STG07 /UNITEN/02/5.

References

1. Cesta, A., D'Aloisi, D., Collia, M.: Adjusting Autonomy of Agent Systems. AAAI Technical Report SS-99-06 (1999)
2. Myers, K.L., Morley, D.N.: Human Directability of Agents. In: K-CAP 2001. ACM 1-58113-380 (2001)
3. Ghallab, M., Nau, D.S., Traverso, P.: Automated planning: Theory and practice. Morgan Kaufmann, San Mateo (2004)
4. Nau, D.S.: Current trends in automated planning. AI Magazine 28(4), 43–58 (2007)
5. Bradshaw, J.M., Feltovich, P.J., Jung, H., Kulkarni, S., Taysom, W., Uszok, A.: Dimensions of Adjustable Autonomy and Mixed-Initiative Interaction. In: Nickles, M., Rovatsos, M., Weiss, G. (eds.) AUTONOMY 2003. LNCS (LNAI), vol. 2969, pp. 17–39. Springer, Heidelberg (2004)
6. Scerri, P., Reed, N.: Designing agents for systems with adjustable autonomy. In: The IJCAI 2001 Workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents (2001)
7. Brainov, S., Hexmoor, H.: Quantifying Relative Autonomy in Multiagent Interaction. In: Hexmoor, H., Castelfranchi, C., Falcone, R. (eds.) Agent Autonomy, pp. 55–74. Kluwer, Dordrecht (2002)
8. Van der Vecht, B., Dignum, F., Meyer, J.-J.C., Neef, M.: A dynamic coordination mechanism using adjustable autonomy. In: Proc. COIN@MALLOW 2007, Durham, UK (2007)
9. Rogers, A., Ramchurn, S.D., Jennings, N.R.: Delivering the smart grid: Challenges for autonomous agents and multi-agent systems research. In: Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI 2012), Toronto, CA, July 22-26, pp. 2166–2172 (2012)
10. Lili, Y., Rubo, Z., Hengwen, G.: Situation reasoning for an adjustable autonomy system. International Journal of Intelligent Computing and Cybernetics 5(2), 226–238 (2012)
11. Dumond, D., Ayers, J., Schurr, N., Carlin, A., Burke, D., Rousseau, J.: Coordinating with Humans by Adjustable-Autonomy for Multirobot Pursuit (CHAMP). Unmanned Systems Technology XIV. In: Karlsen, R.E., Gage, D.W., Shoemaker, C.M., Gerhart, G.R. (eds.) Proceedings of the SPIE, The Smithsonian/NASA Astrophysics Data System, vol. 8387, pp. 838703–838703-15 (2012)
12. Bradshaw, J.M., Sierhuis, M., Acquisti, A., Feltovich, P., Hoffman, R., Jeffers, R., Prescott, D., Suri, N., Uszok, A., Van Hoof, R.: Adjustable Autonomy and Human-Agent Teamwork in Practice: An Interim Report on Space Applications. In: Hexmoor, H., Falcone, R., Castelfranchi, C. (eds.) Agent Autonomy, pp. 243–280. Kluwer (2003)
13. Sierhuis, M., Bradshaw, J.M., Acquisti, A., Van Hoof, R., Jeffers, R., Uszok, A.: Human-Agent Teamwork and Adjustable Autonomy in Practice. In: Proceeding of the 7th International Symposium on Artificial Intelligence, Robotics and Automation in Space: i-SAIRAS, NARA, Japan, May 19-23 (2003)
14. Dorais, G., Bonasso, R.P., Kortenkamp, D., Pell, B., Schrekenghost, D.: Adjustable Autonomy for Human-centered Autonomous Systems on Mars. In: Proceedings of the AAAI Spring Symposium on Agents with Adjustable Autonomy, AAAI Technical Report SS-99-06. AAAI Press, Menlo Park (1999)

15. Cohen, R., Fleming, M.: Adjusting the Autonomy in Mixed-initiative Systems by Reasoning about Interaction. Multiagent Systems, Artificial Societies, and Simulated Organizations 7, 137–158 (2003)
16. Schurr, N., Marecki, J., Tambe, M.: Improving Adjustable Autonomy Strategies for Time-Critical Domains. In: The 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009). International Foundation for Autonomous Agents and Multiagent Systems, Budapest, Hungary, pp. 353–360 (May 2009)
17. Brooks, R.A.: A Robust Layered Control System for a Mobile Robot. IEEE Journal of Robotics and Automation 2(1), 14–23 (1986)
18. Ball, M., Callaghan, V.: Explorations of Autonomy: An Investigation of Adjustable Autonomy in Intelligent Environments. In: The Eighth International Conference on Intelligent Environments, pp. 114–121. IEEE Press, Guanajuato (2012)
19. Mohammed, K.A., Ahmad, M.S., Mostafa, S.A., Sharifuddin, F.M.A.M.: A Nodal Approach to Modeling Human-Agents Collaboration. International Journal of Computer Applications, Foundation of Computer Science 43(12), 33–40 (2012)
20. Bradshaw, J.M., Jung, H., Kulkarni, S., Johnson, M., Feltovich, P., Allen, J., Bunch, L., Chambers, N., Galescu, L., Jeffers, R., Suri, N., Taysom, W., Uszok, A.: Kaa: Policy-based Explorations of a Richer Model for Adjustable Autonomy. In: AAMAS 2005, Utrecht, Netherlands, July 25-29 (2005)
21. Wooldridge, M.: An Introduction to Multiagent System, 2nd edn. A John Wiley and Sons, Ltd. (2009)

A Dynamically Adjustable Autonomic Agent Framework

Salama A. Mostafa[1], Mohd Sharifuddin Ahmad[2], Muthukkaruppan Annamalai[3], Azhana Ahmad[2], and Saraswathy Shamini Gunasekaran[1]

[1] College of Graduate Studies, Universiti Tenaga Nasional, Putrajaya Campus 43000, Selangor Darul Ehsan, Malaysia
[2] College of Information Technology, Universiti Tenaga Nasional, Putrajaya Campus 43000, Selangor Darul Ehsan, Malaysia
[3] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor Darul Ehsan, Malaysia
semnah@yahoo.com, {sharif,azhana,sshamini}@uniten.edu.my, mk@tmsk.uitm.edu.my

Abstract. The design and development of autonomous software agents is still a challenging task and needs further investigation. Giving an agent the maximum autonomous capabilities may not necessarily produce satisfactory agent behavior. Consequently, adjustable autonomy has become the hallmark of autonomous systems development that influences an agent to exhibit satisfactory behavior. To perform such influences, however, a dynamic adjustment mechanism is needed to be configured. The influences are costly in time and implementation especially for systems with time-critical domain. They might negatively influence agent decisions and cause system disturbance. In this paper, we propose a framework to govern an agent autonomy adjustment and minimize system disturbance. The main components of the proposed framework are the Planner, Scheduler and Controller (PSC) that conform to the current trends in automated systems. Two modules are also suggested which are Autonomy Analysis Module (AAM) and Situation Awareness Module (SAM). They are accordingly used to distribute the autonomy and provide balance to the system so that it's local and global desires do not conflict.

Keywords: Software agent, Multi-agent system (MAS), autonomous systems, adjustable autonomy, autonomy analysis, situation awareness, decision-making.

1 Introduction

Developing an intelligent automated system to do some tasks on behalf of humans is a crucial issue [1]. Autonomous systems have been developed for environments that are inaccessible or unsafe for humans to work (e.g. [2]). Such systems are commonly utilized for surveillance and exploration purposes [3]. The software agent is invariably the de facto technology that is utilized to develop autonomous systems. An agent is an artifact that has the characteristics of intelligent entity i.e. autonomy, goal-directedness, reactivity, pro-activity, rationality, embodiments and sociality [4].

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 631–642.
DOI: 10.1007/978-3-642-36981-0_58

Maes [3] stated that "*autonomous agents are computational systems that inhabit some complex dynamic environment, sense and act autonomously in this environment, and by doing so realize a set of goals or tasks for which they are designed*." Autonomy is a characteristic of an intelligent agent that enables it to perform actions based on the current states of its environment without the intervention of humans in the pursuit of its goal [4]. It is useful in human cognitive load reduction for tasks like analysis, control and reasoning [5].

Based on [6], we define a fully autonomous agent as *an agent that can make its own decision based on its observations without any type of intervention to its decisions.* In ambiguous and uncertain circumstances, agent behaviour needs guidance to make successful decisions. Otherwise, undesirable decisions might cause serious unintended consequences [3]. Consequently, external interventions to such circumstances, mainly by humans, are necessary, hence, the case for adjustable autonomy.

Autonomous agents incorporate flexibility to system development in a generic way. However, in this technology there are still many deficiencies that need to be resolved [7]. System's new options, features, and modes create new difficulties. Issues like automation surprises and new error types and causes of failure are resulting challenges that need to be investigated [8], [9]. Subsequently, it is endorsed that agent autonomy is restricted especially for end-user applications [1].

In this paper, we propose the PSC framework that provides autonomy adjustment mechanism that works in a Multi-agent System (MAS) environment. The main parts of the PSC are Planner, Scheduler and Controller that handle system activities. The PSC consists of two modules: Situation Awareness Module (SAM) and Autonomy Analysis Module (AAM) that work on the system's autonomy distribution and self-control. The modules enable an agent to perform at different levels of autonomy to achieve its goal. Autonomy selection condition is fulfilled based on the selected action type. The proposed PSC manifests the specifications of autonomy self-management that enables it to work in an intelligent environment.

2 Research Objectives

This paper represents an attempt to construct an achievable and usable framework that can be adopted in developing a dynamic and robust autonomous system. The following outline the objectives that represent the aim of the research:

1. To review some of the existing agent models that are proposed in agent-based system development.
2. To analyze the related issues in modeling autonomy and adjustable autonomy.
3. To propose a framework and algorithm that can be utilized in developing an intelligent agent-based autonomous system.

3 Agent-Based Autonomy

Autonomous models vary across initiatives as a result of their continuous development by researchers and based on the essential needs of the technology [9]. This development helps to transfer the technologies from the manual mode to the sophisticated autonomous mode (i.e. smart autonomous systems). Initiative system means the ability of the system to perform some activities autonomously based on its reasoning of the environment [1]. An autonomous agent is seen to be one of the core players that contribute in developing system autonomy [10]. Figure 1 shows the autonomy spectrum based on some of the developed agent-based autonomous systems.

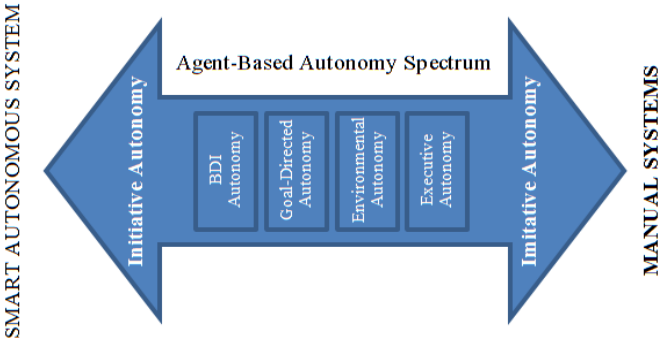


Fig. 1. The spectrum of the agent-based autonomy

3.1 Executive Autonomy

At the right side of the spectrum is the Executive Autonomy which is a type of autonomous systems that can only follow directly what it has been told to do (e.g. [10] and [11]). The agent is considered primitive and acts without reasoning the causes and effects of the actions.

3.2 Environmental Autonomy

Environmental autonomy (e.g. [12] and [13]) is configured based on sensing the changes that occur in an environment that the autonomous entity resides in and generates responses. In this type, the environmental inputs determine the number of possible actions, which might limit the autonomous behaviour of the agent. The environment dependency stands against the agent potentiality especially in uncertain and less intelligent environments [1].

3.3 Goal-Directed Autonomy

In Goal-Directed Autonomy (GDA) (e.g. [14] and [15]) the system can plan, generate and implement its own goals based on specific circumstances (see Figure 2). In GDA, however, goal reasoning and revision mechanism are still not matured [16]. Foremost, there is insufficiency in responding to unanticipated events and dealing with discrepancies of the environment as a result of the absence of retrospective process that associates two (or more) of the generated plans.

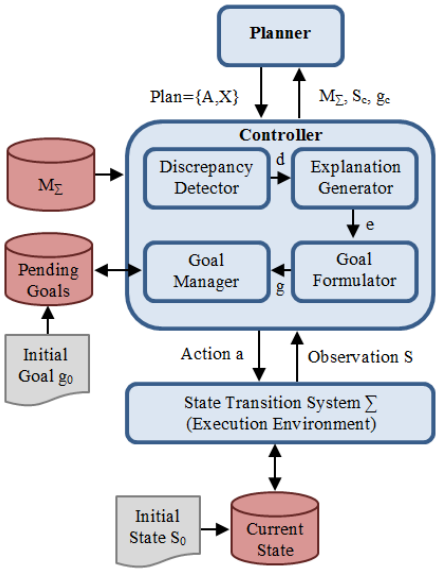


Fig. 2. The GDA architecture [17]

3.4 Belief-Desire-Intention Autonomy

Alternatively, the Belief-Desire-Intention (BDI) model that is introduced by Rao and Georgeff [18] gives the autonomy spectrum another dimension, in the sense that, the agent can autonomously choose the goal that needs to be accomplished based on its beliefs and decide what sub-goals are related to the goal to achieve. The BDI enables an agent to make the most human-like decisions autonomously (e.g. [19] and [20]). Foremost, adopting the BDI architecture in modelling autonomy is due to its synonymy to human's practical reasoning. The four important roles in the BDI practical reasoning are summarized as follows [4]:

1. Observe the world and update beliefs;
2. Deliberate what intentions to achieve by determining the available options and then by filtering;
3. Use means-end reasoning to select the plans to achieve these intentions;
4. Execute the selected plan;

The agent needs to have a repertoire of possible actions (i.e. precompiled actions) that can be performed in its environment, which state its ability in that environment. The agent's role is to select the best sequence of actions from the set of possible actions to achieve its objective [4]. A rational agent would perform the most promising actions based on its beliefs of the environment. Figure 3 illustrates the logical sequence of the BDI model.

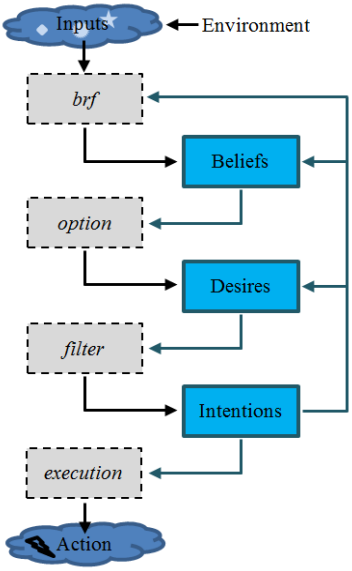


Fig. 3. The BDI architecture

Agent mental attitudes of the BDI represent respectively, the informational, the motivational, and the deliberation states [18]. These mental attitudes determine the system's behaviour and are critical for achieving adequate or optimal performance when deliberation is subject to resource bounds. The agent's beliefs are its knowledge about the world, its internal state, and other agents and their mental states. Its desires correspond to the tasks allocated to it. Since it will not be able to achieve all its desires, it must decide on some subset of its desires and commit resources to achieving them. The chosen desires represent its intentions to achieve the goal. Consequently, Thangarajah et al. ([21], [22]) subscribe to the strategy of aborting tasks in BDI agents and a mechanism to allow agents to detect and avoid negative interactions in which the effect of one goal undo conditions that are potential for another goal. To simplify the BDI algorithm, the following steps are performed:

1. Sense the environment to generate beliefs;
2. If there is no plan;
   - choose a desire to pursue;
   - find a plan to achieve that goal;
3. Decide on the next act to perform from the plan;
4. Perform that act;
5. Every now and then check the plan is still valid;

4 Related Works in Adjustable Autonomy

The changes made to a system by an agent manifest it as an autonomous intelligent entity. As a result, some system activities which are not fixed at design time are handled by the agent to best configure the system's functionality. Thus, Scerri and Reed [7] defined an adjustable autonomous system as "*an intelligent system where the distribution of autonomy is changed dynamically to optimize overall system performance.*" Therefore, autonomy and resource distribution of a system are consistently directed to meet the current situation needs (local desires) and to ensure that the current situation is aligned with the goal situation (global desires) [23].

In the early 80's Brooks [24] was considered as one of the leading researchers that conducted research in autonomy. His research introduces milestones to the intelligent mobile robots field such as the layers of control to the mobile robots in the so called reactive architecture. The reactive architecture consists of parallel and independent modules that are responsible for specific tasks. Reactive agents act in a stimulus response manner to the present state of the environment in which they are embedded [25]. A good example of a reactive architecture is the subsumption architecture of a robot he developed. The architecture is "*a collection of task accomplishing behaviours, each of which is a finite state machine that maps perceptual input to action output*". The behaviours are organized into a layered hierarchy, with lower layers representing less abstract behaviours.

Bonasso [26] applied the 3T layered software architecture in controlling a robot that is tasked for elderly care and support. The 3T architecture is divided into three layers: planner, sequencer and skill manager. It is used to plan for controlling the actions of smart home and linking them to fit elderly peoples' needs. Bonasso argued that fully autonomous system in such domains is likely unachievable. Thus, adjusted semi-autonomous architecture is proposed. Autonomy level selection is adjusted based on the elderly person's conditions. For instance, for the elderly person with less capability, the autonomy level of the system is increased.

Myers and Morley [23] proposed a framework for human-agent system using the adjustable autonomy approach. The research aim is to achieve reliable autonomous system through human guidance. The Intelligence Gathering and Emergency Response (TRAC) framework basically discusses human directability of agents in which policies are defined by human supervisor to influence agent actions during run time. It considers the agent capability of deciding at different degrees of autonomy and studies agent tasks' accomplishment. Human intervention is customized in TRAC to ensure its preferences and to overcome unexpected situations. The TRAC framework is implemented in a MAS environment to support Procedural Reasoning System (PRS). The PRS is used for security purposes and guided by human officer where autonomous agents collect information and perform emergency response.

Van der Vecht et al. [6] proposed a dynamic mechanism of coordination for adjustable autonomy in a multi-agent system to balance the global control with the local autonomy of the system. They experimentally prove that coordination type selection must consider the conditions of the given situation. They argue that, the autonomy adjustment is not a separated mechanism that is performed by the system. Instead, it is an inner process that the agent itself reasons and assigns based on local observations.

Agent-based autonomous systems respond based on some delegated goals, which makes it somehow less initiative [4]. A key aspect of advanced autonomous system is its ability to concurrently communicate and cooperate with other systems [6]. The cooperation between the systems and their entities in the autonomous environment may operate at different level of intelligence and with different degree of autonomy [27]. Human and agent cooperation can be a good example of entities interaction and cooperation. For simplicity, we can illustrate the factors that are responsible for adjusting the autonomy of a system into three main issues [7]:

1. Adjustability information: The information that causes the influence which is related to the decision-making adjustment.
2. Adjustability reasoner: Reasons about the possible and the required adjustment.
3. Adjustability actuator: Translating the reasoner's decisions to actions.

5 The Proposed PSC Framework

The proposed Planner-Scheduler-Controller (PSC) framework conforms to the current trends in automated planning [17]. It consists of five core components which are the Planner, the Scheduler, the Controller, the Situation Awareness Module (SAM) and the Autonomy Analysis Module (AAM) as shown in Figure 4. Each performs its function to achieve the adjustable autonomic actions.

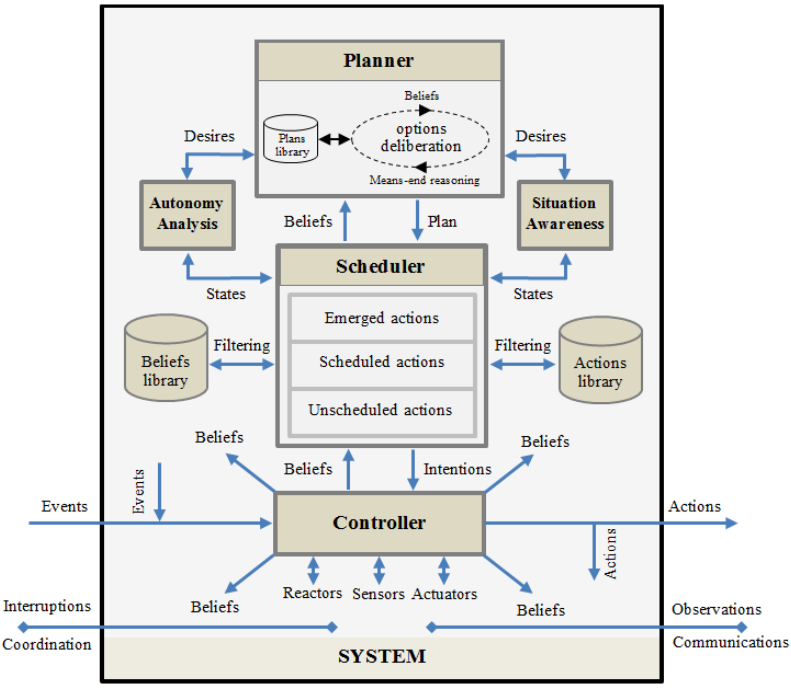


Fig. 4. The PSC framework

The Planner's purpose is to formulate a plan based on the given beliefs and environment states to achieve system's desire. This process is accompanied by selecting the required sequenced sets of actions from the actions library that represent the corresponding intentions to be executed. The former process of deciding the relevant actions to be selected is the deliberation practice and the later process of deciding actions achievements represents the means-ends reasoning practice [4]. The Planner is assisted by two modules, the SAM and AAM.

Situation awareness of autonomous systems is a type of reasoning methodology that is utilized for assessments during adjustable autonomy planning step [28]. The SAM provides information about the system's current situation along with the external influences by analysing the updated beliefs and formulating new desires. It then adjusts the current situation desires to meet the desired situation. The desired situation must be formulated to meet the system's objective which is translated to a plan by the Planner.

The AAM analyses and manages the selected actions for the autonomic execution and deals with its autonomy adjustment. System autonomic responsibilities and authorities are changed based on the actions' types and the system's privileges. If the system selects an action for handling an event, the AAM checks if the system has the knowledge and the authority to handle such event. Then it decides the intervention on the action execution if it is required.

Consequently, actions' execution sequence is implemented according to action categories which are emerged, scheduled and unscheduled actions of the Scheduler as shown in Figure 4. This categorization helps in managing actions implementation and the required adjustment process to be more organized. It also manages the possible events that might happen during system run time such as new situations, coordination, interruptions and impasses [9]. The Scheduler concerns with actions' coordination and provisions to be controlled and executed as organized [26]. These concerns include objectives constraints, timing constraints, resource constraints and adjustments constraints which are provided within the plan by the Planner [17].

Although the emerged actions (also known as contingencies) are considered as a subset of the unscheduled actions, we separate them to prioritize differently for better control of the emerged events. For instance, an autonomous air drone avoids an attack during its surveillance mission is an urgent event that requires high priority to the system. Thus, emerged actions have higher priority followed by the scheduled actions and the lowest are the unscheduled actions. This priority distribution is considered during actions planning and implementation to be adjusted accordingly. Initial plan is needed for system first run (booting) while default plans are required to avoid some causes of system failure such as handling undecidable situation. Plans are formulated based on internal and external events, states and actions that are provided by the system to be controlled.

The Controller of the system implements the translated plan to actions (i.e. intentions) by the Scheduler based on the event situations to achieve the objectives. In order to manipulate the actions, however, the control part utilizes the controlling tools which are the reactors, the sensors and the actuators. Consequently, the Controller is responsible for converting the data provided by the controlling tools to beliefs to be

processed. The control process requires concurrent provisions of the Scheduler's parameters to perform the execution of the actions.

5.1 The PSC Execution Cycle

Assume that there is an event e that occurs in environment E, $e \in E$. There exist an agent Ag_k that functions based on PSC and observes e. The following algorithm represents Ag_k practical reasoning execution cycle:

1. Controller: sense the E to generate beliefs Agb about e;
2. SAM: generate desires Agd_m based on Agb;
3. Scheduler: deliberate the Agb to be filtered to $Agbf$;
4. Planner: deliberate the options for $Agbf$ based on Agd_m;
5. Planner: retrieves and checks If there is a plan that meet the Agd_m to be used:
   if (plan == nil) {
   a. SAM: select a desire Agd_j to pursue;
   b. Planner: find a plan to achieve Agd_j;
   c. if (plan == nil)&&(Agd_j != nil)
      go to a;
   d. else
      Planner: select default plan;
   }

   Configuring a plan
6. Scheduler: select an Ac_i from the plan based on action categories:
   if (action != nil){
   a. if (emerged action != nil)
      switch to emerged action;
   b. else if (scheduled action != nil)
      switch to scheduled action;
   c. else if (unscheduled action != nil)
      switch to unscheduled action;
   }
   else exit;

   Action categories selection
7. AAM: set *autonomy update function (£)* parameters of the Ac_i;
8. Controller: execute the Ac_i and update Agb;
9. SAM: update the Agd_m;
10. Planner:
    if(plan valid==True)
    go to 6;
    else go to 3;

5.2 PSC Potential Advantages

The PSC is a result of the cohesive integration of both GDA and BDI models. We believe that the two models are landmarks that are generally integrated to form a potential operational framework. The PSC accommodates the advantage of the BDI

deliberation and maturity of reasoning with the GDA foundations. The potential advantages of the proposed framework are; directing agent's autonomic decision to minimizing system disturbances, and self-organizing and sequencing system activities to deal with ill-structured conditions as demonstrated in Section 5.1.

5.3 Action's Format

The action consists of a sequence of functions that represent the action's performance. The first function within each action is the *autonomy update function*, the parameters of which are set during scheduling time by the AAM.

$$Ag_k : Ac_i \rightarrow (run\ (£_{i,0}, f_{i,1},\ f_{i,2},\ \ldots, f_{i,\,m}))$$

where Ac_i is the selected action by agent Ag_k, *run* is the *execution function*, £ is the *autonomy update function* for Ac_i, $f_{i,m}$ are the functions that implement Ac_i where m is the total number of the required functions to implement Ac_i.

6 Conclusion and Future Work

Progressively, autonomous systems assume responsibility in tasks that are complicated and critical. Consequent attempts to develop a system that is capable of successful autonomous actions are visible in the literature. However, autonomous systems improvements are found to be aligned with investigating the capabilities of improving systems' autonomy. In this paper, we propose another framework that can deliver autonomic systems' improvement. The Planner-Scheduler-Controller (PSC) framework has the potential to be adopted to work in a practical environment. It details the autonomous system components to manifest an operational autonomy function.

In our future work, a prototype to the proposed PSC framework will be developed using an air drone for testing purposes. The prototype will then be tested in a collaborative environment to demonstrate its usefulness in collective intelligence environments.

Acknowledgments. This project is sponsored by the Malaysian Ministry of Higher Education (MoHE) under the Exploratory Research Grant Scheme (ERGS) No. ERGS/1/2012/STG07 /UNITEN/02/5.

References

1. Ball, M., Callaghan, V.: Explorations of Autonomy: An Investigation of Adjustable Autonomy in Intelligent Environments. In: Eighth International Conference on Intelligent Environments, pp. 114–121. IEEE Press, Guanajuato (2012)
2. Dorais, G., Bonasso, R.P., Kortenkamp, D., Pell, B., Schrekenghost, D.: Adjustable Autonomy for Human-centered Autonomous Systems on Mars. In: Proceedings of the AAAI Spring Symposium on Agents with Adjustable Autonomy, AAAI Technical Report SS-99-06. AAAI Press, Menlo Park (1999)

3. Maes, P.: Artificial Life Meets Entertainment: Lifelike Autonomous Agents. Communications of the ACM 38(11), 108–114 (1995)
4. Wooldridge, M.: An Introduction to Multiagent System, 2nd edn. John Wiley and Sons, Ltd. (2009)
5. Ferguson, G., Allen, J., Miller, B.: TRAINS-95: Towards a mixed-initiative planning assistant. In: Proceedings of the Third Conference on Artificial Intelligence Planning Systems (AIPS 1996), Edinburgh, Scotland, pp. 70–77 (1996)
6. Van der Vecht, B., Dignum, F., Meyer, J.-J.C., Neef, M.: A dynamic coordination mechanism using adjustable autonomy. In: Proc. COIN@MALLOW 2007, Durham, UK (2007)
7. Scerri, P., Reed, N.: Designing agents for systems with adjustable autonomy. In: The IJCAI 2001 Workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents (2001)
8. Rogers, A., Ramchurn, S.D., Jennings, N.R.: Delivering the smart grid: Challenges for autonomous agents and multi-agent systems research. In: Twenty-Sixth AAAI Conferences on Artificial Intelligence (AAAI 2012), Toronto, CA, pp. 2166–2172 (2012)
9. Bradshaw, J.M., Sierhuis, M., Acquisti, A., Feltovich, P., Hoffman, R., Jeffers, R., Prescott, D., Suri, N., Uszok, A., Van Hoof, R.: Adjustable Autonomy and Human-Agent Teamwork in Practice: An Interim Report on Space Applications. In: Hexmoor, H., Falcone, R., Castelfranchi, C. (eds.) Agent Autonomy, pp. 243–280. Kluwer (2003)
10. Barber, K.S., Martin, C.E.: Agent Autonomy: Specification, Measurement, and Dynamic Adjustment. In: Proceedings of the Workshop on Autonomy Control Software, International Conference on Autonomous Agents, Seattle, WA (1999)
11. Ferguson, G., Allen, J., Miller, B.: TRAINS-95: Towards a Mixed-initiative Planning Assistant. In: Proceedings of the Third Conference on Artificial Intelligence Planning Systems (AIPS 1996), Edinburgh, Scotland, pp. 70–77 (1996)
12. Falcone, R., Castelfranchi, C.: From Automaticity to Autonomy: The Frontier of Artificial Agents. In: Hexmoor, H., Castelfranchi, C., Falcone, R. (eds.) Agent Autonomy, pp. 79–103. Kluwer, Dordrecht (2002)
13. Braynov, S., Hexmoor, H.: Quantifying Autonomy. In: Hexmoor, H., Falcone, R. (eds.) Agent Autonomy, pp. 43–56. Kluwer (2003)
14. Molineaux, M.: Design and Evaluation of a Goal-Directed Autonomous Agent. Association for the Advancement of Artificial Intelligence, AAAI (2010)
15. Ashok, K.G.: Reflection in Action: Meta-Reasoning for Goal-Directed Autonomy. Association for the Advancement of Artificial Intelligence, AAAI (2010)
16. Muñoz-Avila, H., Aha, D.W., Jaidee, U., Klenk, M., Molineaux, M.: Applying Goal Driven Autonomy to a Team Shooter Game. In: Proceedings of the Twenty-Third International Florida Artificial Intelligence Research Society Conference (FLAIRS 2010). Association for the Advancement of Artificial Intelligence, AAAI (2010)
17. Nau, D.S.: Current Trends in Automated Planning. AI Magazine 28(4), 43–58 (2007)
18. Rao, A.S., Georgeff, M.P.: BDI-Agents: From Theory to Practice. In: First International Conference on Multi-Agents System (ICMAS 1995), San Francisco, USA (1995)
19. Veres, S.M., Luo, J.: A Class of BDI Agent Architectures for Autonomous Control. In: 43rd IEEE Conference on Decision and Control, CDC, pp. 4746–4751. IEEE Press, Paradise Islands (2004)
20. da Silva, M.R., Borges, A.P., Dordal, O.B., Sato, D.M.V., Avila, B.C., Enembreck, F., Scalabrin, E.E.: An Architecture of BDI Agent for Autonomous Locomotives Controller. In: Proceedings of the 2012 IEEE 16th International Conference on Computer Supported Cooperative Work in Design, pp. 2–29. IEEE Press (2012)

21. Thangarajah, J., Padgham, L., Winikoff, M.: Detecting and Avoiding Interference Between Goals in Intelligent Agents. In: Proceedings of the 18th International Joint Conference on Artificial intelligence (IJCAI 2003), pp. 721–726. Morgan Kaufmann Publishers Inc. (2003)
22. Thangarajah, J., Harland, J., Morley, D., Yorke-Smith, N.: Aborting Tasks in BDI Agents. In: AAMAS 2007, Hawaii, USA (2007)
23. Myers, K.L., Morley, D.N.: Human Directability of Agents. In: K-CAP 2001. ACM 1-58113-380 (2001)
24. Brooks, R.A.: A Robust Layered Control System for a Mobile Robot. IEEE Journal of Robotics and Automation 2(1), 14–23 (1986)
25. Brooks, R.A.: Intelligence without Representation. Artificial Intelligence 47, 139–159 (1991)
26. Bonasso, R.P.: Using Adjustable Autonomy to Extend the Independence of the Elderly. Developing Assistive Technology for People with Disabilities. AAAI, Fall Symposium (1996)
27. Schurr, N., Marecki, J., Tambe, T.: Improving Adjustable Autonomy Strategies for Time-Critical Domains. In: 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009). International Foundation for Autonomous Agents and Multiagent Systems, Budapest, Hungary, pp. 353–360 (2009)
28. Lili, Y., Rubo, Z., Hengwen, G.: Situation Reasoning for an Adjustable Autonomy System. International Journal of Intelligent Computing and Cybernetics 5(2), 226–238 (2012)

High-Level Language to Build Poker Agents

Luís Paulo Reis[1,3], Pedro Mendes[2], Luís Filipe Teófilo[1,2], and Henrique Lopes Cardoso[1,2]

[1] LIACC – Artificial Intelligence and Computer Science Lab., University of Porto, Portugal
[2] FEUP – Faculty of Engineering, University of Porto – DEI, Portugal
[3] EEUM – School of Engineering, University of Minho – DSI, Portugal
lpreis@dsi.uminho.pt, {ei01108,luis.teofilo,hlc}@fe.up.pt

Abstract. On the last decade Poker has been one of the most interesting subjects for artificial intelligence, because it is a game that requires game playing agents to deal with an incomplete information and stochastic scenario. The development of Poker agents has seen significant advances but it is still hard to evaluate agents' performance against human players. This is either because it is illicit to use agents in online games, or because human players cannot create agents that play like themselves due to lack of knowledge on computer science and/or AI. The purpose of this work is to fill the gap between poker players and AI in Poker by allowing players without programming skills to build their own agents. To meet this goal, a high-level language of poker concepts – PokerLang – was created, whose structure is easy to read and interpret for domain experts. This language allows for the quick definition of an agent strategy. A graphical application was also created to support the writing of PokerLang strategies. To validate this approach, some Poker players created their agents using the graphical application. Results validated the usability of the application and the language that supports it. Moreover, the created agents showed very good results against agents developed by other experts.

Keywords: Knowledge Representation, Decision Support Systems, Artificial Intelligence, Computer Games, Poker.

1 Introduction

Poker is the most popular betting game in the world. Played by millions around the world, poker has become a very profitable business. Given its growing popularity and the amounts of money involved (millions of dollars), Poker became a research subject in very different areas such as Mathematics, Artificial Intelligence and Sociology, among others. Key features such as incomplete knowledge, risk management, need for opponent modeling and dealing with unreliable information, have turned Poker into an important topic in Computer Science, especially for artificial intelligence.

Since the number of online Poker players still continues to grow, a large number of tools have been created to assist players in their game. Most tools are statistics-based programs that save information about the played games, creating statistical knowledge about opponents in order to help the user to make the right decision in future games.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 643–654.
DOI: 10.1007/978-3-642-36981-0_59

The main goal of this work is to provide a tool capable of creating Poker Agents, through the definition of high level rules. This way, anyone with interest and knowledge about Poker can easily create a Poker agent, even without any computer programming skills. This work has been divided into the following sub goals:

- Create a language of concepts, which includes the main ideas behind poker moves and agent behavior.
- Build a graphical user interface for this language, which allows the user to create rules in a more simple way.
- Develop a Poker agent that follows the language specification.
- Evaluate the interface usability and the performance of the developed agent.

The rest of the paper is as follows. Section 2 presents recent methods to create Poker agents and information representation in Poker. Section 3 presents the specification for PokerLang. Section 4 presents a graphical application that was built to aid the creation of PokerLang files. Section 5 describes the agent that was built to follow PokerLang strategies. Section 6 presents some experiments and results. Finally, section 7 concludes and points directions for future research.

2 Related Work

First approaches to build Poker agents were rule-based, which involves specifying the action that should be taken for a given information set [1]. The next approaches were based on simulation techniques like [2], i.e. generating game random instances in order to obtain a statistical average and decide the action. These approaches led to the creation of agents that were able to defeat weak human opponents.

The great breakthrough in Computer Poker research was the discovery of the Counter Factual Regret Minimization Algorithm (CFR) in [3]. The CFR algorithm allows for the computation of a Nash Equilibrium strategy in large games such as Poker through self-play[1]. This could be done before through linear programming methods (e.g. Simplex) but CFR is much faster because the processing time is proportional to the number of information sets instead of to the number of game states (about 6 orders of magnitude less). Several approaches based on CFR, like Restricted Nash Response [4] and Data-biased response [5] backed up the first victories against Poker experts.

Other recent methodologies were based on pattern matching [6, 7], Monte Carlo Search Tree algorithm [8], reinforcement learning [9] and case based reasoning [10]. More recent works are described in the reviews [11, 12].

Another possible approach consists on the defining of the agent's strategy through a high level specification language. One example is the Poker Programming Language (PPL) [13], which is the most similar work to the one described in this paper. The main issue about PPL is that it only considers low level features of Poker which means that it takes a long time to specify a complete strategy. Moreover, the

[1] Self-play – an agent playing against itself or against an agent with an equal strategy.

absence of advanced game concepts (such as pot odds, implied odds and others) makes it only possible to create very basic and static strategies which are easily beaten by a medium level opponent.

3 PokerLang

Due to its stochastic nature, Poker players use rather different tactics in each game situation. A tactic is used under certain conditions that are represented by specific game features such as current stack, number of opponents, position at the table and others. A set of tactics compose the player's strategy. In order to specify these concepts and determine when to use each tactic, a high-level language was created – PokerLang – whose syntax and grammar was based on COACH UNILANG [14, 15] and similar languages [15–17]. COACH UNILANG was successfully used in the robotics soccer domain [18–21]. The generic approach of this language allows for its easy adaption to other domains.

The language root starts by defining the concept of strategy: a strategy is a set of tuples which one composed by a tactic and an activation condition for that tactic. Each activation condition corresponds to a set of verifications of the visible game features (through evaluators) or predictions about uncertain information (through predictors). The activation condition consists of comparing those features with parameterized values. The tactic is the procedure followed by the player when the activation condition is met. The tactic could be either user-defined or language predefined (based on common expert tactics). A top level specification of the language can be found below. In the following subsections, each language concept will be presented in depth.

```
<STRATEGY>::= {<ACTIVATION_CONDITION> <TACTIC>}
<ACTIVATION_CONDITION>::= {<EVALUATOR>}
<TACTIC>::= <PREDEFINED_TACTIC>|<TACTIC_NAME><TACTIC_DEFINITION>
<PREDEFINED_TACTIC>::= loose_agressive | loose_passive |
                       tight_agressive | tight_passive
<TACTIC_NAME>::= [string]
<TACTIC_DEFINITION>::= {<BEHAVIOUR> <VALUE>}
<BEHAVIOUR>::= {<RULE>}
<RULE>::= {<EVALUATOR> | <PREDICTOR>} <ACTION>
<ACTION>::=  {<PREDEFINED_ACTION><PERC> | <DEFINED_ACTION><PERC>}
```

3.1 Evaluators

Evaluators compare the game's visible features with given values. Since Poker is an incomplete information game, evaluators make use of only certain measures to assess how the player is standing in the game.

```
<EVALUATOR>::= <NUMBER_OF_PLAYERS> | <STACK> | <POT_ODDS> |
            <HAND_STRENGTH> | <HAND_REGION> | <POSITION_AT_TABLE>
```

Number of Players. This evaluates how many players one is competing against. The number of players is an important measure because the higher it is, the lower is the probability of success of a given hand[2].

Stack. The stack is the relative amount of chips that a player currently owns given by formula M (Equation 1). The value has to be relative since there is a plethora of possibilities of a player's amount of chips. $M = \frac{Stack}{SB+BB+Antes}$. The stack evaluator is defined by levels. They can be predefined (see Table 1) or customized as follows.

```
<STACK>::= <PREDEFINED_STACK_REGION> | <STACK_REGION_DEFINITION>
<PREDEFINED_STACK_REGION>::= green_zone | yellow_zone | orange_zone |
                          red_zone | dead_zone
<STACK_REGION_DEFINITION>::= <STACK_REGION_NAME> <STACK_INTERVAL>
<STACK_REGION_NAME>::= [string]
<STACK_INTERVAL>::= <MIN_STACK> <COMP> <STACK_VALUE> <COMP>
<MAX_STACK>
<MIN_STACK>::= <STACK_VALUE>      <MAX_STACK>::= <STACK_VALUE>
```

Table 1. User defined Stack Regions

Name	*Stack/M*
Green Zone	M>20
Yellow Zone	20>M>10
Orange Zone	10>M>5
Red Zone	5>M>1
Dead Zone	M<1

Pot Odds. Pot Odds is the ratio between the size of the pot and the cost calling[3] the opponent's bet. The pot odds are usually compared with the hand odds. When the pot odds are higher than the hand odds, the player should call the hand.

Hand Region. The probability of winning a game in Poker depends on the player's starting cards. There are $\frac{52!}{(52-2)!\times 2!} = 1326$ possible combinations of starting hands. This poses a problem because if the user were to define a tactic for every starting hand, the number of possible combinations would be enormous. To solve this problem, the language uses *bucketing*. Bucketing is an abstraction technique that consists of grouping different hands that should be played in a similar way [5]. PokerLang allows the users either to define their own groups or to use Dan Harrington's groups (see Table 2) [22].

[2] Hand – set of a player's cards that determine his/her score in the game.
[3] Call – match the current highest bet.

```
<POT_ODDS>::= <REAL>
<HAND_REGION>::= <PREDEFINED_HAND_REGION> | <HAND_REGION_DEFINITION>
<PREDEFINED_HAND_REGION>::= a | b | c | d | e
<HAND_REGION_DEFINITION>::= <HAND_REGION_NAME> {<HAND>}
<HAND_REGION_NAME>::= [string]
```

Table 2. Dan Harrington's Groups

Group	*Hands*
A	AA, KK, AKs
B	QQ, AK, JJ, TT
C	AQs, 99, AQ, 88, AJs
D	77, KQs, 66, ATs, 55, AJ
E	KQ, 44, KJs, 33, 22, AT, QJs

Hand Strength. This defines the minimum hand strength to activate the evaluator. The hand strength is given by the ratio between the number of hands that has lower score than the player's hand and the total number of possible hands [23].

Position at Table. The position at table is the player's relative position to the current Big Blind position[4]. The later the position is, the better chance the player has to observe his or her opponents' moves. Since games have a variable number of players, in order to better abstract the strategies, the position value is defined through the position quality PQ (Equation 2), which also depends on the type of the opponents.

$$\text{PQ} = \text{Position} - (\text{Number of aggressive players} + \text{Number of tight players}) \quad (1)$$

The range of position quality depends on the number of players in the following proportion: Range = [-(Number of players-2), (Number of players-2)]. For instance, in a 10 player table, the range would be [-8, 8].

```
<POSITION_AT_TABLE>::= <PREDEF_POSITION_REGION>|<POSITION_REGION_DEF>
<PREDEF_POSITION_REGION>::= bad_pos | normal_pos | good_pos
<POSITION_REGION_DEF>::= <POSITION_REGION_NAME>{POSITION}
<POSITION_REGION_NAME>::= [string]
<POSITION>::= <MIN_POS> <COMP> <POS_VALUE> <COMP> <MAX_POS>
<POS_VALUE>::= <INTEGER>
```

There are 3 predefined regions but the user is allowed to defined custom regions. Being n the number of players, the regions are calculated as depicted in Equation 3.

$$Min = -n + 2, Max = n - 2, TR = (n - 2) \times 2 \quad (2)$$

$$Bad = [Min, Min + {}^{TR}/_{3}\,[Normal =] Min + {}^{TR}/_{3}, Max - {}^{TR}/_{3}$$
$$[Good =]\, Max - {}^{TR}/_{3}, Max]$$

[4] Big blind position – the position of the last player to act.

3.2 Predictors

The predictors are game features that are estimated. Since the hidden information in Poker (opponents' cards) is crucial to the game's outcome, to be competitive a player must make predictions about what is the actual game state. Predictions are based on the opponents' moves on previous games.

```
<PREDICTOR>::= <IMPLIED_ODDS> | <OPPONENT_HAND> | <TYPE_OPPONENT > |
               <STEAL_BET> | <IMAGE_AT_TABLE>
```

Implied Odds. Implied Odds corresponds to the pot odds but taking into account the evolution of the player's hand.

```
<IMPLIED_ODDS>::= <REAL>
```

Opponent Hand. A possible opponent hand taking into account the player's cards and the community cards[5]. For instance, if the opponent hand predictor is "Flush", this should be read as "If the opponent is able to reach a flush".

```
<OPPONENT_HAND>::= <HAND>
```

Type of Player. The type of the last opponent in the table taking into account his/her past behavior in the game. There are 4 predefined types of opponents based on [24].

```
<TYPE_OPPONENT>::= loose_agressive | loose_passive | tight_agressive |
                   tight_passive
```

Steal Bet. The steal bet is the amount of chips you need to get the pot with no hand at all. It depends on the type of opponents that one is facing.

```
<STEAL_BET>::= <BET_VALUE>
```

Image at Table. The type of player that one's opponents see in him / her. This is rather important because if, for instance, the player is seen as a tight player, his/her bluffs will have higher probability of succeeding.

```
<IMAGE_AT_TABLE>::= <TYPE_OF_PLAYER>
```

3.3 Actions

There are several poker plays that one can use in a game. These moves are specific ways of handling a hand to achieve a goal. In this definition, the user can choose predefined moves (based on common expert moves) or custom moves.

[5] Community card – table card that every player can score with.

```
<ACTION>::=  {<PREDEFINED_ACTION><PERC> | <DEFINED_ACTION><PERC>}
<PREDEFINED_ACTION>::= <STEAL_THE_POT> | <SEMI_BLUFF> |
      <CHECK_RAISE_BLUFF> | <SQUEEZE_PLAY> | <CHECK_CALL_TRAP> |
      <CHECK_RAISE_TRAP> | <POST_OAK_BLUFF>
```

Moves can be customized by defining the bet amount on each round.

```
<DEFINED_ACTION>::= <ACTION_NAME>{<PRE_FLOP_ACTION> | <FLOP_ACTION> |
                    <TURN_ACTION> | <RIVER_ACTION>}
<PRE_FLOP_ACTION>::= {<BET_VALUE><PERC>}
<FLOP_ACTION>::= {<BET_VALUE><PERC>}
<TURN_ACTION>::={<BET_VALUE><PERC>}
<RIVER_ACTION>::={<BET_VALUE><PERC>}
```

4 Poker Builder

After defining the high-level language, the next phase of this work was concerned with building a simple graphical application to allow users to easily create PokerLang files based on the group previous work on the area [25, 26]. Poker Builder is a Flex application that allows the user to create rules of concepts using the language previously introduced, and set the behavior of a poker agent. With a smooth interface and simple features, Poker Builder is accessible to any user that understands the main concepts of poker. One of the purposes of this work was to make a very practical application, even usable to users only familiarized with the most basic computer usage.

For the implementation of the language of concepts, Poker Builder is divided in four major classes: Strategy, Tactic, Rule and Property (Fig. 1). The interface begins with an instance of the Strategy Class that creates instances of all other classes depending on what the user is creating. Poker Builder gives the user two different views to create rules: Strategy View and Tactic View.

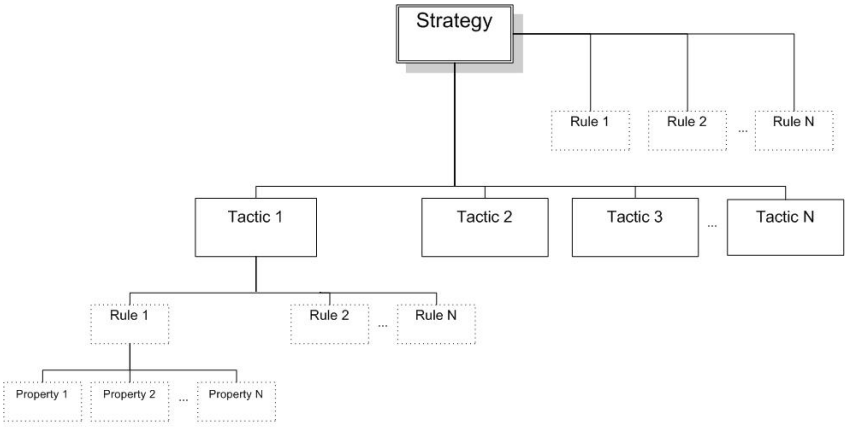


Fig. 1. Main Classes Diagram

The software includes a Strategy View to allow the user to create decision rules in what tactic the poker agent should use on the defined circumstances. It is the most high-leveled definition that the user can use from the language of concepts. The main distinction from the Tactic View is that despite the fact that the evaluators and predictors are the same, the actions are not. Instead of the list of poker moves that are available to the user, the Strategy View presents the list of tactics already defined by the user. The software also includes a Tactic View as the main view of Poker Builder. It is presented when the program starts and is where the user defines the lower level specifications of the agent. It is presented with a list of the evaluators, the predictors and some common poker moves that professional players use in their game (actions). The menus are only available in this view, which includes the possibility of saving and loading strategies or tactics.

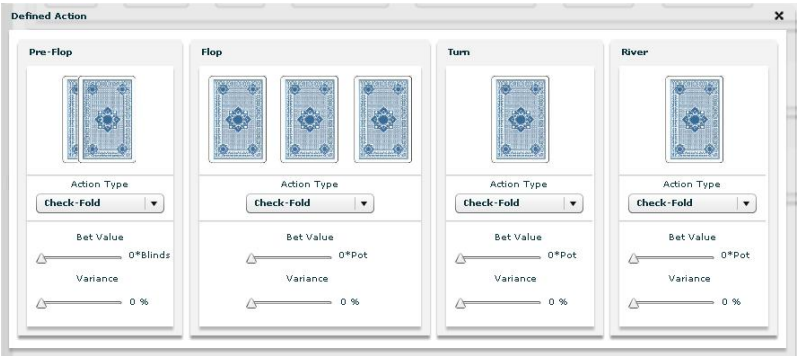


Fig. 2. Action Interface in Poker Builder

5 Poker Builder Agent

The final step of this work was to build the poker agent that uses the strategies previously created. To be able to follow the strategies, the agent needs some reading features of the information gathered at a poker table. Obtaining the evaluators' features is trivial because they comprise perfect information (data obtained just by looking at the table), as opposed to predictors. Predictors require a statistical study of the played hands in order to get reliable information. Another feature required by the agent is an algorithm to select which rule to apply. An agent with these features will be an agent capable of strictly following the strategy defined previously.

The agent's action sequence starts by reading the strategy to use from the respective file. In each of the states, the agent will follow sequentially three major steps: reading all the information of the table, which includes setting the values of the evaluators, and trying to suit the imperfect information of the predictors, searching the most suitable rules for the table circumstances and choosing the rule to follow. At the end of each hand, the agent will save all the hand's information: bets from the opponents, each opponent hand (if shown), the position of the opponent and more.

The agent was built to work on the LIACC Simulator described in [27]. This simulator has features that ease the construction, test and validation of the agent. Moreover, due to compatibility with AAAI simulator, it also allows the developed agent to directly participate in the annual computer poker competition [28].

6 Tests and Results

Poker is a game with elements of chance thus complicating player rating. The purpose of this work is not to build a poker agent to win against every opponent but to enable the user to define behaviors in a simple way.

All the tests were conducted in the Pre-Flop version of No Limit Texas Hold'em in head's up games. Two distinct agents were built:

- Agent PokerTron - This agent has a simple strategy (with only one tactic and five rules) but yet capable of trapping and bluffing opponents along the game. The behavior of this agent with all hands has a good variety of moves making it very difficult to read.
- Agent Hansen - This agent has a much more complex strategy than PokerTron. It contains three different tactics, used in specific circumstances, being the choice of what tactic to use based on the current stack. With a large stack, the agent will play a very loose game, practically never folding any hand pre-flop and trying to get their opponents out of the game with large bets. With a normal stack it will play more specific hands (group A and B, see Table 2) more carefully, avoiding making bluffs. With a very small stack, the agent will wait for a hand A or B and goes all-in[6].

Two simulations were run: one to test the PokerLang agents' behavior and another to test their performance against two previously developed agents.

6.1 Behavior Test

In Table 3 we can see the percentage of rule activation for each agent, during the 10 games played. This represents the number of times each agent makes a decision based on its strategy. The fact that a strategy is defined does not imply that it will be followed every single hand. This happens because the strategy does not cover all possible circumstances that can occur in a poker game. In Table 3, we can see that agent Hansen has a higher percentage of rule activation. This means that the full area of possible circumstances is more covered in agent Hansen than it is in agent PokerTron.

Table 3. Rule Activation of Hansen and PokerTron Agent

	Hansen	*PokerTron*
Rule Activation	64%	48%

[6] All-in – betting the total amount of chips.

Another important statistic is the tactic activation (Table 4). In the case of PokerTron, there is only one tactic defined, but in Hansen there are three. The "aggressive" tactic has a higher percentage (the agent won most of the simulated games), which means it had a high stack most of the times. The low stack tactic was less used because this tactic is only activated for low stack and for hands of group A/B, which did not happen often since Hansen was almost always leading the tournament.

Table 4. Tactic Activation of Hansen Agent

	HighStack	*NormalStack*	*LowStack*
Tactical Activation	56%	39%	5%

6.2 Performance Test

Hansen and PokerTron were put up against the two observing agents created by Dinis Ferreira [12] in a tournament (limited resources). Figure 3 shows that the PokerLang agents ended up competing against themselves with a final victory for Hansen (the agent with a more complex strategy).

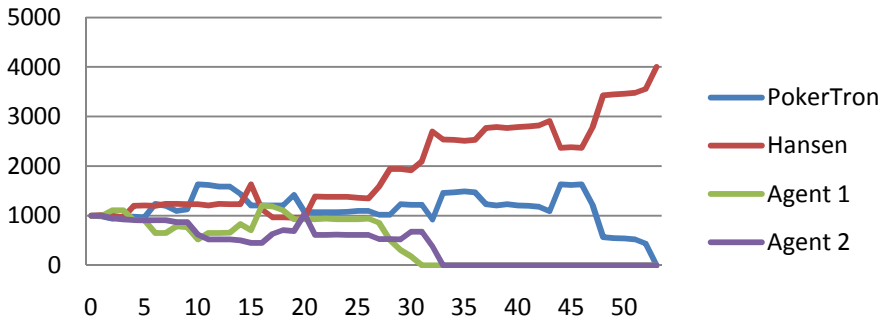


Fig. 3. Stack Evolution from one of the simulated games. Horizontal axis shows the number of hands and the vertical axis displays the agent's stacks.

Both Poker Builder agents gained advantage early in the game, being able to eliminate Agent 1 and Agent 2 in the 31st hand and 33rd hand, respectively. The most important fact to retrieve from these results is that Poker Builder can be used to produce effective agents in a short time and in a very simple way.

These simulations could be made with several thousand games played, but the purpose of these tests was to prove the efficiency of the application and the agent that supports it. The first test showed the effectiveness of the agent reading and running the strategies defined. In the Tournament simulation, the intention was to show how Poker Builder agents would handle different agents from another. Satisfactory results were obtained, despite the fact of running a small number of games.

7 Conclusions

The purpose of this work was to create Poker playing agents more accessible to the common user and, thus, a comprehensible high level language that represents Poker strategies was created. PokerLang filled the gaps of previous approaches like the Poker Programming Language because it allows the definition of much more complex and complete strategies. An intuitive and pleasant graphical application to support the creation of PokerLang files was also created, thus making it easier the creation of Poker playing agents.

Tests and simulations showed that the created agents correctly followed several PokerLang strategies. Moreover, agents made by Poker players were able to beat previously developed agents. However, experiments with PokerLang agents developed using professional Poker players and playing against the best poker playing agents and the best human poker players, are still required to further validate this approach.

In future research, more game concepts can be added to cover up more poker specifications and to make the agents even more effective, such as the customization of abstraction techniques. Another important feature would be the inclusion of an exploration map to allow the agent to assume how to play with information sets that were not defined, instead of just folding. The work will also be concerned with gathering professional poker player models using this language and comparing the models with the real players' behavior in order to fully and further test the expressiveness of the PokerLang language.

References

1. Billings, D., Papp, D., Schaeffer, J., Szafron, D.: Opponent modeling in poker. In: National Conf. on AI, pp. 493–499. John Wiley & Sons (1998)
2. Billings, D., Papp, D., Peña, L., Schaeffer, J., Szafron, D.: Using selective-sampling simulations in poker. In: AAAI Syring Symp. Search Tec. for Problem Solving under Uncertainty and Incomplete Information, pp. 1–6 (1999)
3. Zinkevich, M., Bowling, M., Burch, N.: A new algorithm for generating equilibria in massive zero-sum games. In: AAAI 2007, vol. 1, pp. 788–793 (2007)
4. Johanson, M.: Robust Strategies and Counter-Strategies: Building a Champion Level Computer Poker Player (2007)
5. Johanson, M., Bowling, M.: Data biased robust counter strategies. In: AISTATS 2009, pp. 264–271 (2009)
6. Teofilo, L.F., Reis, L.P.: HoldemML: A framework to generate No Limit Hold'em Poker agents from human player strategies. In: CISTI 2011, pp. 755–760 (2011)
7. Teófilo, L.F., Reis, L.P.: Building a No Limit Texas Hold'em Poker Agent Based on Game Logs Using Supervised Learning. In: Kamel, M., Karray, F., Gueaieb, W., Khamis, A. (eds.) AIS 2011. LNCS, vol. 6752, pp. 73–82. Springer, Heidelberg (2011)
8. Van den Broeck, G., Driessens, K., Ramon, J.: Monte-Carlo Tree Search in Poker Using Expected Reward Distributions. In: Zhou, Z.-H., Washio, T. (eds.) ACML 2009. LNCS, vol. 5828, pp. 367–381. Springer, Heidelberg (2009)

9. Teófilo, L.F., Passos, N., Reis, L.P., Cardoso, H.L.: Adapting Strategies to Opponent Models in Incomplete Information Games: A Reinforcement Learning Approach for Poker. In: Kamel, M., Karray, F., Hagras, H. (eds.) AIS 2012. LNCS, vol. 7326, pp. 220–227. Springer, Heidelberg (2012)
10. Rubin, J., Watson, I.: Case-based strategies in computer poker. AI Communications 25, 19–48 (2012)
11. Rubin, J., Watson, I.: Computer poker: A review. Artificial Intelligence 175, 958–987 (2011)
12. Teofilo, L.F., Reis, L.P., Cardoso, H.L.: Computer Poker Research at LIACC. In: Computer Poker Symposium, AAAI 2012 (2012)
13. Technologies, S.: Poker Programming Language User Guide (2009)
14. Reis, L.P., Lau, N.: COACH UNILANG - A Standard Language for Coaching a (Robo)Soccer Team. In: Birk, A., Coradeschi, S., Tadokoro, S. (eds.) RoboCup 2001. LNCS (LNAI), vol. 2377, pp. 183–192. Springer, Heidelberg (2002)
15. Reis, L.P., Oliveira, E.C.: A Language for Specifying Complete Timetabling Problems. In: Burke, E., Erben, W. (eds.) PATAT 2000. LNCS, vol. 2079, pp. 322–341. Springer, Heidelberg (2001)
16. Pereira, A., Duarte, P., Reis, L.P.: ECOLANG - A Communication Language for Simulations of Complex Ecological Systems. In: Merkuriev, Y., Zobel, R., Kerckhoffs, E. (eds.) ECMS 2005, Riga Latvia, pp. 493–500 (2005)
17. Neves, R., Reis, L.P., Abreu, P., Faria, B.M.: A multi-agent system to help Farmville players on game management tasks. In: CISTI 2012, pp. 409–414 (2012)
18. Reis, L.P., Lau, N.: FC Portugal Team Description: RoboCup 2000 Simulation League Champion. In: Stone, P., Balch, T., Kraetzschmar, G.K. (eds.) RoboCup 2000. LNCS (LNAI), vol. 2019, pp. 29–40. Springer, Heidelberg (2001)
19. Reis, L.P., Lau, N., Oliveira, E.C.: Situation Based Strategic Positioning for Coordinating a Team of Homogeneous Agents. In: Hannebauer, M., Wendler, J., Pagello, E. (eds.) ECAI-WS 2000. LNCS (LNAI), vol. 2103, pp. 175–197. Springer, Heidelberg (2001)
20. Abreu, P.H., Moura, J., Silva, D.C., Reis, L.P., Garganta, J.: Performance analysis in soccer: A Cartesian coordinates based approach using RoboCup data. Soft Computing 16, 47–61 (2012)
21. Mota, L., Reis, L.P., Lau, N.: Multi-robot coordination using Setplays in the middle-size and simulation leagues. Mechatronics 21, 434–444 (2011)
22. Harrington, D., Robertie, B.: Harrington on Hold 'em Expert Strategy for No Limit Tournaments. Strategic Play, vol. 1. Two Plus Two Pub. (2004)
23. Teófilo, L.F.: Estimating the Probability of Winning for Texas Hold'em Poker Agents. In: 6th Doctoral Symposium on Inf. Eng., pp. 129–140 (2011)
24. Sklansky, D.: The Theory of Poker: A Professional Poker Player Teaches You How to Think Like One. Two Plus Two (2007)
25. Felix, D., Reis, L.P.: An Experimental Approach to Online Opponent Modeling in Texas Hold'em Poker. In: Zaverucha, G., da Costa, A.L. (eds.) SBIA 2008. LNCS (LNAI), vol. 5249, pp. 83–92. Springer, Heidelberg (2008)
26. Felix, D., Reis, L.P.: Opponent Modelling in Texas Hold'em Poker as the Key for Success. In: Ghalib, M., Spyropoulos, C.D., Fakotakis, N., Avouris, N. (eds.) ECAI 2008 – 18th European Conference on Artificial Intelligence, Patras, Greece, vol. 178, pp. 893–894. IOS Press (2008)
27. Teófilo, L.F., Rossetti, R., Reis, L.P., Cardoso, H.L.: A Simulation System to Support Computer Poker Research. In: 13th MABS 2012, Valência (2012)
28. Zinkevich, M., Littman, M.L.: The 2006 AAAI Computer Poker Competition. Journal of International Computer Games Association, 166–167 (2006)

New Crime Detection with LBP and GANN

Ahmad Kadri Junoh[1] and Muhammad Naufal Mansor[2]

[1] Institute of Engineering Mathematics (IMK), University Malaysia perlis, Malaysia
[2] Intelligent Signal Processing Group (ISP), University Malaysia perlis, Malaysia
ahmadkadrijunoh@unimap.edu.my, apairia@yahoo.com

Abstract. A current review of media sources indicates crimes of opportunity such as burglaries, purse-snatchings and vehicle theft, are consistently the most topical crime problems. The Malaysian government has taken several steps to increase police effectiveness and reduce crime since 2004. But their effectiveness is limited by low salaries and lack of manpower. Verily to compile a comprehensive afford of crime fighting capabilities. We present an explicit system to detect a crime scene with Local Binary Pattern (LBP) and a fusion of Genetic Algorithm with Neural Network (GANN). This system provided a good justification as a monitoring supplementary tool for the Malaysian police arm forced.

Keywords: Crime rate, Local Binary Pattern, Genetic Algorithm Neural Network and Feed Forward Neural Network.

1 Introduction

Reliable and current crime statistics are very difficult to locate in Malaysia. Some crime index indicators are adjusted year to year making it difficult, if not impossible, to assess consistent trends over time [1]. It's reported that violent crimes comprised only 10% of reported crimes each year and the majority of crimes, 90%, were classified as property crimes by Journal of the Kuala Lumpur Royal Malaysia Police College. Vehicle thefts comprised very nearly one half of all property crimes. A current review of media sources indicates crimes of opportunity such as burglaries, purse-snatchings and vehicle theft, are consistently the most topical crime problems [2]. The Malaysian government has taken several steps to increase police effectiveness and reduce crime since 2004. But their effectiveness is limited by low salaries and lack of manpower [3]. Thus, within these drawbacks, we present a comprehensive study to determine crime behavior with LBP and GANN. This system provided a good justification as a monitoring supplementary tool for the Malaysian police arm forced.

1.1 Face Detection

Loris Nanni et. al, purposes a new method for selecting the most discriminant rotation invariant patterns in local binary patterns and local ternary patterns. The experiment

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 655–660.
DOI: 10.1007/978-3-642-36981-0_60

conducted in several domains using five benchmark databases and obtains the best performance in almost all the datasets [4]. Timo Ahonen presents a novel approach to face recognition which considers both shape and texture information to represent face images. The face is divided into small regions from which Local Binary Pattern (LBP) and tested with FERET database [5]-[6]. Extension in [5], Timo Ahonen employed a local binary pattern histogram fourier features and kernel density estimation [7]-[8]. The results showed a remarkable conclusion. Furthermore, most of the researchers compromised a well known database in their study. In contrast, we refine a new database consists a crime scene. In order to justify the robustness, images were taken in different angle. In addition we compare our experiment with GANN and neural network algorithm.

1.2 System Overview

The proposed face recognition system consists recognition phases as depicted in Fig. 1. It consists of several modules which are, face detection, recognition, feature extraction, GANN and neural network classifier are performed.

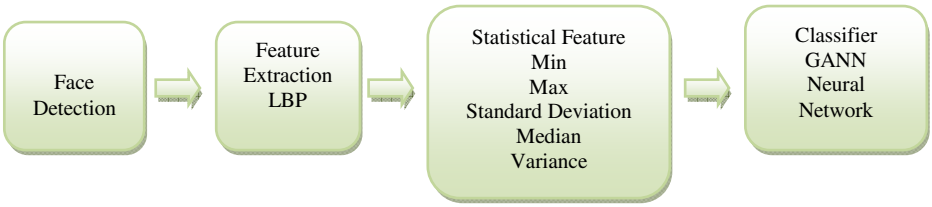


Fig. 1. Block Diagram of Face Recognition System

1.3 Face Database

This database consist 200 hundred images with 10 different persons. The images were captured in different distances, angles and illumination levels. Some of bandit images are shown in Fig. 2

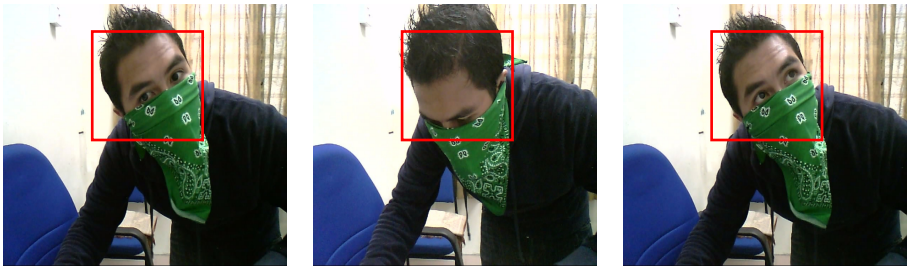


Fig. 2. Bandit Image

1.4 Face Detection

The face recognition module contains of preprocessing, feature extraction, and classification. The input to the face recognition module is the face image, which is derived from two sources from the camera or from the database. During feature extraction, the normalized image is represented as feature vectors. We adopted color detection method suggest by [11]-[13].

1.5 Feature Extraction

The feature extraction algorithms used is Local Binary Pattern (LBAP) by [4]-[8]. LBP was introdued by Ojala et al. [9]. LBP value is computed by comparing it with those of its neighborhoods:

$$LBP_{P,R} = \sum_{i=0}^{p-1} 2^i \times f(g_i - g_c) \tag{1}$$

$$f(x) = \begin{cases} 1x \geq 0 \\ 0x < 0 \end{cases} \tag{2}$$

where g_c is the gray value of the center pixel, gi is the gray value of its neighbors, P is the number of neighbors and R is the radius of the neighborhood. LBP extracted the relevant information in a face image and encoded as efficiently as possible. The image face defined by LBP as shown in Fig 3. Some of statistical feature such as min, max, standard deviation and variance were employed suggest by [14].

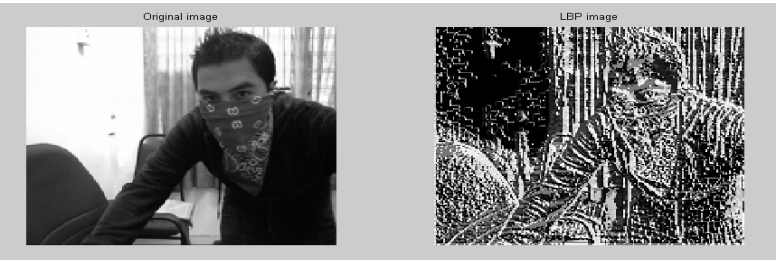


Fig. 3. LBP Face of Bandit images

1.6 GANN Classifier

This algorithm applies a tournament selection (ordinal based) and replacement consists on a steady state update also implemented with a tournament technique. The tournament size is 3. A hundred of generations for the genetic algorithm are carried out in every experiment. All the experiments use a population size of 20. This is a

standard value used in genetic algorithms. Each neural network has 3 hidden layers and is trained over 200 epochs with back propagation. This value is higher than the one usually used to train neural networks, giving enough time to the training to converge, and so taking advantage of the whole potential of each network [10].

2 Results and Analysis

The proposed algorithm was evaluated on a hundred and twenty subjects with different race, gender and age. The average size of each image is 400-500 pixels. The entire subjects were tested for ten trials with GANN and Feed Forward Neural Network with 3 layers, 20 neurons and 5000 epoch. Ten images were taken for each subject. Average accuracy of GANN and neural network algorithm for all subjects was shown in Figure 4.

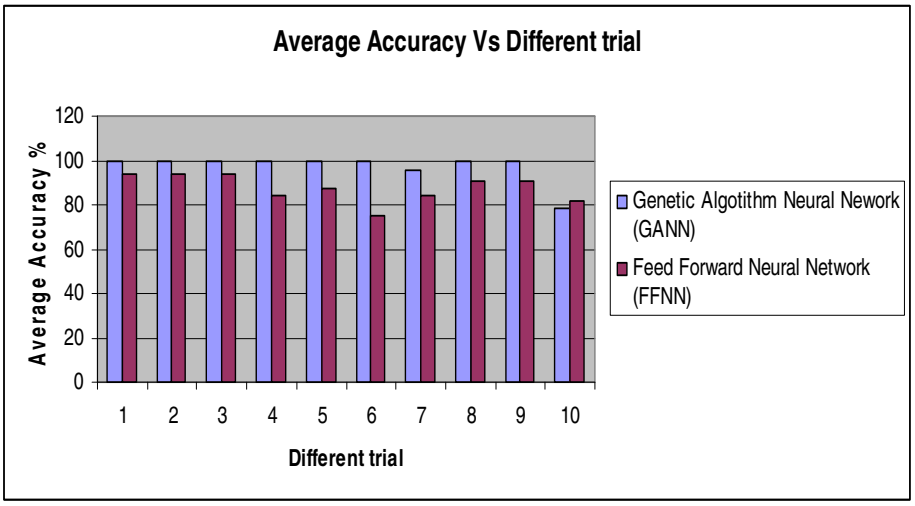


Fig. 4. Average Accuracy with different Trial

Based on the Fig. 4 below, it proved that GANN algorithm concluded a remarkable result. Rigorously, GANN outperformed the FFNN for the entire test route. The average accuracy for the ten trials is 97.12% for GANN and 87.5% for FFNN. Minimum average accuracy is 99.94% for GANN and 93.75% for FFNN respectively. However maximum average accuracy is 75.00% for GANN and 87.5% for FFNN respectively.

3 Conclusion

In this paper, we proudly had proven that LBP and GANN concept is suitable in bandit recognition. Though, framework for a whole system has been developed in this

paper, we concentrated only on database search mechanism. The proposed mechanism can be applied not only in crime monitoring but also in other surveillance system. This successfully provided a good justification as a monitoring supplementary tool for the Malaysian police arm forced.

Acknowledgments. This research was conducted under Fundamental Research Grant Scheme (FRGS) which is contributed by Ministry of Higher Education Malaysia.

References

1. Charles, L.: 'More crimes last year, but fewer violent ones' The Star on Line (January 10, 2008), http://thestar.com.my/news/story.asp?file=/2008/1/10/nation/19964810&sec=nation (accessed January 5, 2010)
2. ACP Amar Singh Sidhu, This Rise of Crime in Malaysia An Academic and Statistical Analysis. Journal of the Kuala Lumpur Royal Malaysia Police College (4), 25 (March 23-24, 2005), http://mpk.rmp.gov.my/jurnal/2005/riseofcrime.pdf (accessed January 6, 2010)
3. Feedom House 2010, Freedom in the World – Malaysia (June 2010), http://www.freedomhouse.org/template.cfm?page=363&year=2010 (accessed September 10, 2010)
4. Nanni, L., Brahnam, S., Lumini, A.: Selecting the Best Performing Rotation Invariant Patterns in Local Binary/Ternary Patterns. In: IPCV 2010, pp. 369–375 (2010)
5. Ahonen, T., Hadid, A., Pietikäinen, M.: Face Recognition with Local Binary Patterns. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
6. Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12), 2037–2041 (2006)
7. Ahonen, T., Matas, J., He, C., Pietikäinen, M.: Rotation Invariant Image Description with Local Binary Pattern Histogram Fourier Features. In: Salberg, A.-B., Hardeberg, J.Y., Jenssen, R. (eds.) SCIA 2009. LNCS, vol. 5575, pp. 61–70. Springer, Heidelberg (2009)
8. Ahonen, T., Pietikäinen, M.: Pixelwise Local Binary Pattern Models of Faces Using Kernel Density Estimation. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 52–61. Springer, Heidelberg (2009)
9. Ojala, T., Pietikainen, M., Harwood, D.: A comparative sudy of texture measures with classification based on feature distributions. Elsevier J. Pattern Recognition 29(1), 51–59 (1996)
10. Fiszelew, A., Britos, P., Ochoa, A., Merlino, H., Fernndez, E., Garca-Martnez, R.: Finding Optimal Neural Network Architecture Using Genetic Algorithms. Research in Computing Science 27, 15–24 (2007)
11. Mansor, M.N., Yaacob, S., Nagarajan, R., Hariharan, M.: Coma Patients Expression Analysis under Different Lighting using k-NN and LDA. International Journal of Signal and Image Processing (IJSIP) 1(4), 249–254 (2010)

12. Mansor, M.N., Idris, A.B., Yaacob, S., Nagarajan, R., Hariharan, M.: Quantifying of Paraplegic Patient Facial Agitation based on Fuzzy k-NN. Journal of Computer Technology and Application (JCTA) 2(1), 24–28 (2011)
13. Mansor, M.N., Yaacob, S., Nagarajan, R., Hariharan, M.: Reckoning of Facial Grimacing Scheme for Patient Agitation in Critical Care. International Journal of Research and Reviews in Artificial Intelligence (IJRRAI) 1(1), 20–24 (2011)
14. Satiyan, M., Nagarajan, R., Hariharan, M.: Recognition of Facial Expression using Haar Wavelet Transform. Transaction on International Journal of Electrical and Electronic Systems Research (IEESR) 3, 91–99 (2010)

Cooperative Scheduling System with Emergent Swarm Based Behavior

Ana Madureira, Ivo Pereira, and Diamantino Falcão

GECAD - Knowledge Engineering and Decision Support Research Center
Institute of Engineering – Polytechnic of Porto (ISEP/IPP)
Porto, Portugal
{amd,iaspe,dfmfa}@isep.ipp.pt

Abstract. This paper presents a Swarm based Cooperation Mechanism for scheduling optimization. We intend to conceptualize real manufacturing systems as interacting autonomous entities in order to support decision making in agile manufacturing environments. Agents coordinate their actions automatically without human supervision considering a common objective – global scheduling solution taking advantages from collective behavior of species through implicit and explicit cooperation. The performance of the cooperation mechanism will be evaluated consider implicit cooperation at first stage through ACS, PSO and ABC algorithms and explicit through cooperation mechanism application.

Keywords: Cooperation, Swarm Intelligence, Scheduling Systems, Multi-Agent Systems.

1 Introduction

Cooperation can be empirically stated as the common effort of a group for their mutual benefit. Cooperation in economic and social contexts may be viewed as a relationship based on collaboration between individuals or organizations as a means of achieving their objectives [1].

Software systems developing involving autonomic interacting software agents present new challenges in Computer Science and Software Engineering. Agent based technologies provide a way to conceptualize complex and dynamic systems as comprising interacting social and autonomous entities, acting, learning and evolving separately in response to interactions and stimuli in their local environment [2]. Additionally, these techniques provide theoretical understanding and practical implementation of Multi-Agent based systems.

Social issues in the organization of Multi-Agent Systems (MAS) have been increasingly relevant over the last decades as a platform to define the structure of agent interactions in open and dynamic environments [2]. Nature provides several and diverse examples of social systems, such as: insect colonies like ants, termites, and bees and its foraging behavior for food; bacteria which appear able to act in a

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 661–671.
DOI: 10.1007/978-3-642-36981-0_61

finalized way; the human brain considering that intelligence and mind arises from the interaction and coordination of neurons; the molecule and cell formation considering homeostasis and the capability of adapting and reproducing arise from protein interactions and antibody detection. Several efforts and contributions have been made to take social nature as an inspiration and basis for algorithms developing based on analogy with social and self-organized behavior. These approaches have been generally referred as Swarm Intelligence (SI) [3].

The proposed cooperation mechanism is embedded on AutoDynAgents scheduling system architecture [4]. The architecture consists in a system in which communities of agents model a real manufacturing system subject to perturbations. Agents must be able to manage their internal behavior and their relationships with other agents, through a cooperative mechanism in accordance with business policies defined by the user manager. In the proposed Swarm based Cooperation Mechanism for scheduling in dynamic environments, agents coordinate their actions automatically without human supervision considering a common objective – global scheduling solution taking advantages from collective behavior of ants (Ant Colony Optimization- ACO), bees (Artificial Bee Colony - ABC) and particles (Particle Swarm Optimization-PSO) through implicit cooperation.

The remaining sections are organized as follows. In section 2 theoretical foundations of SI paradigm are described. Section 3 describes cooperation aspects on MAS. In section 4 the addressed cooperation mechanism is described. Section 5 presents the computational study and discusses the obtained results. Finally, the paper presents some conclusions and puts forward some ideas for future work.

2 Swarm Intelligence

Optimization algorithms inspired from the collective behavior of biological populations that can observed in nature such as ants, bees, fish, and birds are associated to the paradigm Swarm Intelligence and have been referred as a creative approach to optimization problem solving. ACS, PSO and ABC algorithms are among the most promising SI optimization techniques for scheduling resolution [3].

2.1 Ant Colony System

The Ant Colony Optimization (ACO) algorithm takes inspiration from the foraging behavior of some ant species. These ants deposit pheromone on the ground in order to mark some favorable path that should be followed by other members of the colony. ACO exploits a similar mechanism for solving optimization problems. The ACO algorithm is a probabilistic technique for solving computational problems, which can be reduced to finding good paths through graphs. This algorithm, initially proposed by Marco Dorigo in his PhD thesis [5], is a member of ant colony algorithms family.

The first ACO algorithm is known as Ant System that was aiming to search for an optimal path in a graph. It was based on the foraging behavior of ants seeking a path between their colony and a source of food. The original idea has since diversified to

solve a wider class of numerical problems, and as a result, several problems have emerged, drawing on several aspects of the behavior of ants [6].

Several ACO algorithms have been proposed in the literature, which differ in some decisions characterizing the construction process of solutions and update pheromone procedures [3][4][6].

2.2 Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a population based stochastic optimization technique proposed by Eberhart and Kennedy [7] inspired by social behavior of bird flocking or fish schooling. PSO shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA).

In PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles. The particle swarm concept originated as a simulation of simplified social system. The original intent was to graphically simulate the choreography of a bird block or fish school. However, it was found that particle swarm model can be used as an optimizer [7]. Each particle (individual) adjusts its flying according based on its own and group experiences.

2.3 Bees Based Algorithms

The Bees Algorithm is a new population-based search algorithm, first developed in 2005 by Pham et al. [8] and Karaboga et al. [9] independently. The algorithm mimics the food foraging behaviour of swarms of honey bees. In its basic version, the algorithm performs a kind of neighbourhood search combined with random search and can be used for optimization problems. In 2005, Pham proposed a *Bees Algorithm* in a technical report [8] inspired in the foraging behaviour of honey bees to find food sources. At the same time Karaboga [9] proposes an Artificial Bee Colony (ABC) algorithm that proposes a similar inspiration in the foraging behaviour of the bees.

Real bees are social insects living in organized group called hive. In a beehive, the individuals have some specific tasks performed by specialized individuals. The goal of this organization is to maximize the amount of nectar in the colony getting the utmost of the food sources.

3 Cooperation in Multi-Agent Systems

Cooperation in economic and social contexts may be viewed as a relationship based on collaboration between individuals or organizations as a means of achieving common objectives. The cooperation contrasts with the concept of competition where the participants are selfish.

Agent-based Computing arises as an important area of research and development that have emerged in information technology to support Self-* Systems and Autonomic Computing paradigms [2][10][11]. Techniques to design and implement agent based systems could be categorized into three classes [2]: organization level

(concerning organizational structure related to agent societies as a whole, trust, norms and obligations, …), interaction level (concerning agent communication, interaction and decision making) and Agent Level (concerning individual agents, like reasoning and learning). According to Wooldridge [11], intelligent agents are defined as agents, capable of flexible autonomous action to meet their design objectives. They must include:

- **Reactivity** to perceive and respond to changes occurring in their environment;
- **Pro–activeness** as the ability to exhibit goal-directed behaviour by taking the initiative, responding to perturbations;
- **Sociability** as the capability of interacting with other agents.

In context of intelligent agents additional properties are also mentioned [11]: Self-analysis as the ability to analyse and explain its behaviour and detect error and success; and Learning, adapting and improving through interaction with the environment.

Different agents interactions have associated different models, architectures, agents typologies and classifications. Nwana [10] proposes a typology of agents that identifies some dimensions of classification according to: Mobility, as static or mobile; Presence of a symbolic reasoning model, as deliberative or reactive; Roles, as information or Internet; Exhibition of ideal and primary attributes, such as autonomy, cooperation and learning; and Hybrid philosophies, which combine two or more approaches in a single agent.

Based on the exhibition of ideal and primary attributes, Nwana [10] proposes a classification where agents may be classified considering several ideals and primary attributes which agents should exhibit such as:

- **Autonomy** refers to the principle that agents can operate freely without the need for human guidance, even though this would sometimes be invaluable. A key element of their autonomy is their proactiveness [11];
- **Cooperation** – Cooperation with other agents is relevant since it is the reason for having multiple agents in the first place in contrast to having just one. In order to cooperate, agents need to possess a social ability [11];
- **Learning** – considering that for an Agent-based system to be truly smart, it would have to learn as they react and/or interact with their external environment considering that a key attribute of any intelligent being is its ability to learn [11].

Different MAS approaches are described on literature to implement cooperation between agents that could be categorized in two main classes [4][12][13][14]: first, each agent is able to communicate with the others requesting their needs to the group. It requires a higher degree of agents' intelligence, since they should be able to analyze the task and communicate with each other to obtain the solution. In the second class of approaches, a coordinator agent analyzes the problem and, based on their characteristics; send the tasks to each agent individually.

4 Coordination Mechanism

The proposed Cooperation Mechanism is considered and embedded on a MAS architecture where a community of agents models a real manufacturing system subject to perturbations [4]. The scheduling approach followed in this proposal considers a specific kind of social interaction - Cooperative Problem Solving - where a group of agents work together to achieve a good solution for the scheduling problem.

A common feature of real world systems is that an organized behavior could emerge from agents' interactions. Ants organize to build colonies, neurons organize to produce adaptive human behavior, and businesses organize to create economies. To address this problem we intend to integrate and explore: Multi-Agent Systems, Cooperation, Autonomic Computing and the evolution of Artificial Intelligence paradigm, like Swarm Intelligence.

The purpose of this paper is concerned with the definition of cooperation mechanism to be embedded on a self-managed scheduling system able to exhibit intelligent and self-configuring behaviour.

The architecture of the prototype is based in three main modules [4]: the User Interface, the Self-organization Module and the Scheduling Module. Our focus in this paper is related with the scheduling module results in order the increase system effectiveness through cooperation. Self-organization Module is out of streamline objectives for this paper.

The Scheduling Module is performed by Job Agents, Resource Agents, and UI Coordinator Agent that implement coordination mechanisms to enable solutions refinement. The coordination mechanism procedure is performed in two steps: the system waits for the solutions obtained by Resource Agents and then apply a repair mechanism [15] to accomplish local solutions considering jobs precedence constraints and resource occupation times till a feasible solution is obtained; In the second step, a cooperation mechanism is established between related resource agents in order to interact with each other to improve schedule quality. This coordination mechanism is prepared to accept agents subjected to dynamism [4].

4.1 Model

The scheduling problem is decomposed into a series of Single Machine Scheduling Problems (SMSP). The Resource Agents (with a SI method associated) obtain local solutions and later cooperate in order to overcome inter-agent constraints and achieve a global schedule [15].

The system architecture is illustrated on Fig. 1. In order to support communication and interaction with the user, a User Interface Agent (GUI) is implemented. This agent, apart from being responsible for the user interface, dynamically generates the necessary Job Agents according to the number of jobs that comprise the scheduling problem and assign each job to the respective Job Agent. It is also responsible for the verification of feasible schedules and identification of constraint conflicts on each job and the decision on which Resource Agent is responsible for solving a specific conflict. The User Interface Agent assumes the coordinator role on scheduling process.

The Job Agents are responsible for the generation of the earliest and latest processing times. It is also responsible for assigning the jobs' operations to the respective Resource Agent. The Resource Agents are responsible for the scheduling of the operations that require processing in the machine supervised by the agent. These agents implement SI algorithms (ACO, PSO or ABC) in order to find best possible operation schedules and are also responsible to deal with dynamism (new jobs arriving, cancelled jobs, changing jobs attributes, etc), to change/adapt the parameters of the basic algorithm according to the current situation and to cooperate with other agents.

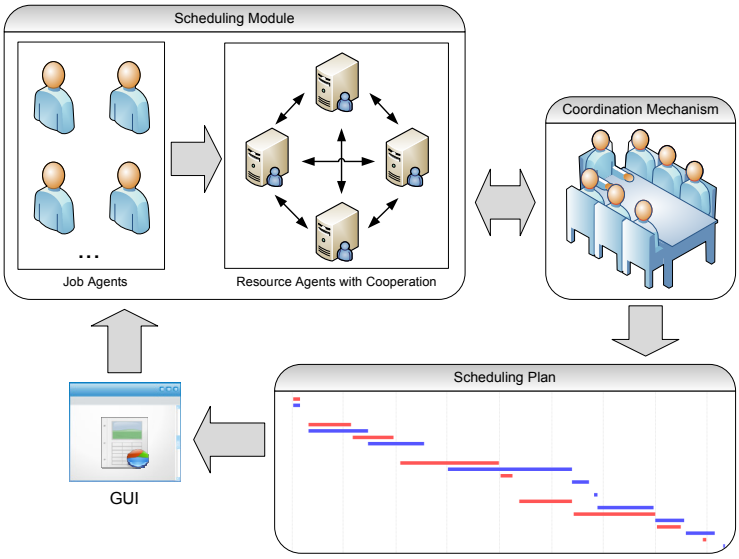


Fig. 1. System architecture

After solution construction, Resource Agents communicate the obtained solutions to the Coordination Mechanism for a feasibility check. Agents must be able to learn and manage their internal behavior and their relationships with other agents, using a cooperative mechanism in accordance with business policies defined by the user manager.

4.2 Cooperation Mechanism Algorithm

The main objective of the cooperation mechanism is to provide the system with cooperative intelligence in order to analyze the scheduling plan generated by the Resource Agents and improve it by reducing idle times and completion times, making small changes to improve the final solution (Table 1).

The cooperation mechanism minimizes the idle times in the machines through swapping operations, keeping all constraints imposed by the problem. To minimize the idle times it is created a backup copy of the best scheduling solution reached that

is updated every time the cooperation mechanism reaches a new best solution. This behavior guaranties that even when the cooperation mechanism does not improve the final solution the system have always a possible solution and the system never degrades the current solution. When the cooperation mechanism reaches to a better solution it updates the backup copy; when a worse solution is reached then it is discarded and the old solution in backup is used.

Table 1. Cooperation Mechanism Algorithm

Algorithm 4: Cooperation Mechanism Algorithm	
Step 1: Data loading	Stores the best solution
	Analyze resources searching for idle times between operations
	Stores the *makespan* value and the ID of the delayed operation
	Sorts the operations by idle times descending order
Step 2: Computational tests	Verify if idle times to improve still exists
	Tests if it is possible to exchange the delayed operation with the following operation or if it is possible to exchange the precedence of the delayed operation with its precedence operation
Step 3: New plan construction	Creates a new scheduling plan with the changes obtained from step 2.

The verification of idle times is made after the first application of the repair mechanism and every time that exits a better solution from the cooperation module. As a result we will obtain, if possible, a continuous better scheduling plan in the resource that will reduce the conclusion time of all jobs.

Without integration of the cooperation mechanism, the system is entirely dependent of the initial agent's solutions and unable to evolve the scheduling plans. With the proposed cooperation mechanism we pretend to give the system the ability of optimize the global solution by cooperation, allowing it to evolve and produce better scheduling plans.

5 Computational Study

A software tool was developed to perform the computational study aiming to analyse and evaluate the performance and advantages of the proposed cooperation mechanism, on minimizing the *makespan* (C_{max}). The computational tests were carried out on a PC with Intel Xeon W3565 at 3.20 GHz. The performance was tested on 30 benchmark instances of Job-Shop Scheduling Problem (JSSP) from different sizes, available at OR-Library [16]. The instances were selected based on their dimension (number of jobs).

We consider that academic benchmark problems are an effective evaluation framework since multiple authors and diverse application areas have used them over the years. Additionally, they permit an insight of global behavior and performance on a class of scheduling problems, which are our main objective.

5.1 Discussion of Results

In order to evaluate the performance of the proposed cooperation mechanism the system was carried out for three Swarm Intelligent techniques, ACS, PSO and ABC. Each algorithm computed n=5 simulations for each instance under analysis, leading to 150 simulations in total. The cooperation mechanism (CM) applied to ACS, PSO and ABC improved most of the simulations (Table 2).

Table 2. Improved solutions through Cooperation Mechanism

ACS		PSO		ABC	
Total	**%**	**Total**	**%**	**Total**	**%**
106	70,67%	117	78,67%	96	64%

From the obtained results it is possible to conclude on the advantages of CM associated to PSO solutions when analyzed the number of improved solutions, summarized on Table 2, and compared with the solutions obtained before applying CM. The percentage of improvements was 78,67%, leading to a total of 117 improvements from the 150.

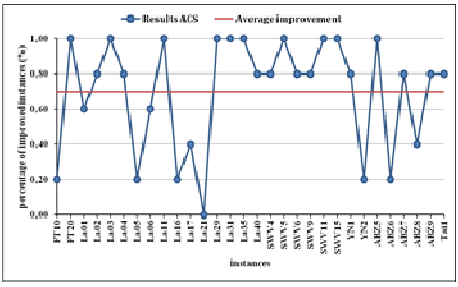


Fig. 2. ACS with the cooperation mechanism

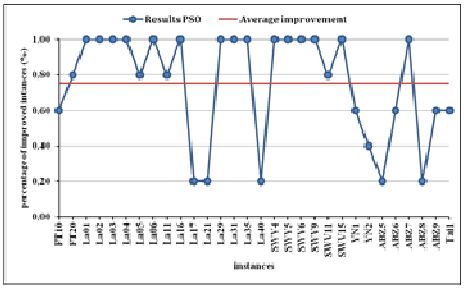


Fig. 3. PSO with the cooperation mechanism

Figures 2, 3 and 4 illustrates the obtained results by ACS, PSO and ABC with the cooperation mechanism. It is clear by the results that ACS achieved an improvement of 100% (5 out of 5) in 10 instances, 80% (4 out of 5) in other 10 instances, and in the remaining ones an improvement less than or equal to 60% (3 out of 5). Moreover, through the results it is possible to see that instance La21 represents an extreme case, i.e., none of the 5 simulations was improved by the CM. As for PSO the results show that most of the solutions have improved. In fact, most of the problem instances present an improvement greater or equal to 60% (3 out of 5), 24 to be more precise, where 15 of these 24 present a 100% improvement (5 out of 5).

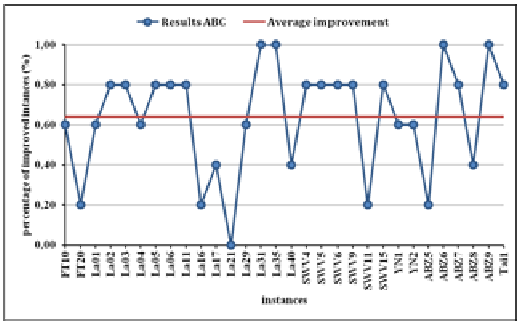


Fig. 4. ABC with the cooperation mechanism

In ABC case the results show that the improvement in most instances is greater or equal to 60% (3 out of 5). However, when comparing with the results obtained by the ACS and PSO, ABC presents the worst overall results. For instance, ACS obtained 100% (5 out of 5) improvement in 10 of the 30 instances under analysis, PSO managed to improve 15 instances, and ABC only 4 instances. But, this can be mainly explained by the fact that ABC manages to get good results before applying the CM, and therefore the mechanism offered no improvement to the solution.

In general, most of the instances were solved in relatively short computational time. For example, instance SWV11 with 50 jobs and 10 machines took 14 seconds with ACS, 11 seconds with PSO and 15 seconds with ABC. In average the instances were solved in 3 seconds with ACS, 12 seconds with PSO, and 9 seconds with ABC.

5.2 Significance Analysis

The evaluation of performance of SI techniques was based on use Friedman's test and Post-hoc Pairwise Comparisons following a systematic methodology described by Villegas [17], using *makespan* average values obtained by the different Swarm Intelligence based techniques (ACO, PSO and ABC) on the resolution of 30 test problems of JSSP for *makespan.* From the analysis, we can conclude that ACO and PSO have the same performance and ABC is the most effective, outperforming ACO and PSO. Thus, is it possible to conclude about statistic evidence that permits to say, with a confidence level of *99%* and *60 degrees of freedom* that ABC is the most effective when the objective is the minimization of *makespan*.

6 Conclusions

In this paper we study the effectiveness and efficiency of using a cooperation mechanism to solve realistic scheduling problems under dynamic environments.

A cooperative dynamic scheduling system was implemented and evaluated using nature inspired techniques, ACS and PSO. Hence, a cooperation mechanism was implemented in order to enable a level of interaction between the agents. This kind of social interaction, allow agents to work together in achieving a good global solution to the problem.

The obtained results show that proposed cooperative mechanism performs remarkably well and revealed to be a significant improvement to the system. In fact, with the mechanism almost every instance improved.

Acknowledgments. This work is supported by FEDER Funds through the "Programa Operacional Factores de Competitividade - COMPETE" program and by National Funds through FCT "Fundação para a Ciência e a Tecnologia" under the project: FCOMP-01-0124-FEDER-PEst-OE/EEI/UI0760/2011. The second author would like also to thanks FCT for the Ph.D. scholarship SFRH/BD/63404/2009.

References

1. Diekmann, A., Lindenberg, S.: Sociological aspects of cooperation. In: International Encyclopaedia of the Social Sciences and Behavioral Sciences. Elsevier (2001)
2. Luck, M., McBurney, P., Shehory, O., Willmott, S.: Agent Technology: Computing as Interaction, A Roadmap for Agent-Based Computing, AgentLink III (2005)
3. Dorigo, M.: Swarm Intelligence. Springer, New York (2007)
4. Madureira, A., Pereira, I.: Intelligent Bio-Inspired System for Manufacturing Scheduling under Uncertainties. Int. Jour. of Comp. Inform. Sys. and Ind. Manag. Appli. 3 (2011)
5. Dorigo, M.: Optimization, Learning and Natural Algorithms. PhD Thesis, Politecnico di Milano, Italy (1992) (in Italian)
6. Dorigo, M., Gambardella, L.M.: Ant Colony System: A cooperative learning approach to the traveling salesman problem. IEEE Trans. on Evol. Comp., 53–66 (1997)
7. Kennedy, J., Eberhart, R.C.: Particle Swarm Optimization. In: Proceedings of the IEEE International Conference Neural Networks, pp. 1942–1948 (1995)
8. Pham, D.T., Ghanbarzadeh, A., Koc, E., Otri, S., Rahim, S., Zaidi, M.: The Bees Algorithm, Manufacturing Engineering Centre. Cardiff University, United Kingdom (2005)
9. Karaboga, D., Akay, B.: A modified Artificial Bee Colony (ABC) algorithm for constrained optimization problems. Applied Soft Computing 11, 3021–3031 (2011)
10. Nwana, H.S.: Software Agents: An Overview. Knowledge Engineering Review 11(3), 205–244 (1996)
11. Wooldridge, M., Jennings, N.R.: Intelligent Agents: Theory and Practice. Knowledge Engineering Review 10(2) (1995)
12. Allen-Williams, M.: Coordination in multi-agent systems. Ph.D. Thesis (2005)
13. Huang, Z., Gao, P., He, Y., He, Q.: Multi-Agent Cooperation Based on Interest Group. In: International Conference on Computational Intelligence and Software Engineering (2009)

14. Ouelhadj, D., Petrovic, S., Cowling, P.I., Meisels, A.: Inter-agent cooperation and communication for agent-based robust dynamic scheduling in steel production. Adv. Eng. Inform. 18 (2004)
15. Madureira, A., Ramos, C., Silva, S.C.: A Coordination Mechanism for Real World Scheduling Problems Using Genetic Algorithms. In: IEEE World Congress on Computational Intelligence (CEC 2002), Honolulu - Hawai (EUA) (2002)
16. OR-Library, http://people.brunel.ac.uk/~mastjjb/jeb/info.html
17. Villegas, J.G.: Using nonparametric test to compare the performance of Metaheuristics (2012), http://www-labsticc.univ-ubs.fr/or/sites/default/files/Friedman%20test%20-24062011.pdf

Finding the Suitable Number of Resources to Maximize System Throughput

M. Carmen Ruiz, Diego Pérez, Juan José Pardo, and Diego Cazorla

Escuela Superior de Ingeniería Informática
Universidad de Castilla-La Mancha
Campus Universitario s/n. 02071 Albacete, Spain
{mcarmen.ruiz,diego.perez,juanjose.pardo,
diego.cazorla}@uclm.es

Abstract. This paper presents a performance analysis tool (BAL) for concurrent systems that involves resources and synchronization. It is based on timed process algebra (BTC) for specifying process systems and their resource requirements. The output of the tool is the minimum time and resources needed to reach a given state. BAL makes the syntactic analysis of the system specification, draws up its relevant transition graph by applying the rules of the operational semantics and solves a performance optimization problem relevant to the minimization of the maximum completion time. A case study is used to assess how BAL works, the degree of usefulness of the results and the time required to obtain them.

Keywords: Tool, Formal Methods, Performance Evaluation, Available Resources, Timed Process Algebra.

1 Motivation

Twenty years ago, formal methods were perceived mainly as techniques for system verification and were suitable only for small academic problems. Today, formal methods have evolved considerably in the analysis they can perform and in the characteristics and sizes of the systems to study. Many formal methods have appeared to reflect different system properties that initially were excluded (such as probability or time) there also being a wide variety of analysis on system specifications (e.g., correctness and model checking) but many of them limit their study to toy systems mainly due to the exponential growth in components when systems enlarge. Nowadays, it is necessary to use formal methods for the study of real-world systems of significant scale. For this aim, there is a clear need to develop tools based on these well-established formal methods that are able to perform the different analyses in an automatic manner. There are already tools of this type but there is still a long way to go.

In this paper, we present a tool based on formal methods capable of carrying out temporal performance analysis in real-world systems. This new tool is called BAL (not an acronym). BAL takes as its input system specifications made using the timed process algebra BTC (Bounded True Concurrency) and performs its syntactic analysis.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 673–684.
DOI: 10.1007/978-3-642-36981-0_62

BTC is based on CSP [4] and extends CSP syntax in order to consider the duration of actions by means of a timed prefix operator. Likewise, the operational semantics has also been extended to consider the context (resources) in which processes are executed. BTC is able to take into account the fact that system resources must be shared by all the processes. So, if there are more processes than resources then not all of them can be simultaneously executed. A process has to wait until it allocates the resources needed to continue its execution. This means that we can find two kinds of delays in the execution of a process: delays related to the synchronization of processes, and delays related to the allocation of resources. The former is usual in a (theoretical) concurrent context, but the latter is only taken into account if we consider a limited amount of available resources.

Once the formal resource-aware model has been defined and we are able to execute clear and precise specifications, we focus our attention on performance evaluation. The performance of a system may be different depending on the amount of available resources. If we denote by $[P]_N$ a process which is executed in the scope of N resources, we will have $[a|b]_1 \equiv [a.b + b.a]_1$ but $[a|b]_n \not\simeq [a.b + b.a]_n \ \forall n > 1$. In general, given a process P, $\forall n, m \in \mathbb{N}$ with $n \neq m$, we find that the performance of $[P]_n$ may - or may not - be different from the performance of $[P]_m$. Actually, one of the main applications of this process algebra consists in finding a natural number n such that $\forall i \in \mathbb{N}, i \geq n$, $[P]_i$ is equivalent (from a performance point of view) to $[P]_n$, i.e., n represents the maximum degree of parallelism we can exploit in the system, and, as a result, we obtain the optimum number of resources needed in order to speed up system performance.

Given that this evaluation has to be done automatically, the BAL tool has been developed. This tool is able to make the syntactic analysis of systems specified in BTC, build its relevant transition graph by applying the rules of the operational semantics and calculate the minimum time needed to reach the final state from the initial one.

This evaluation made by BAL can be used in two different ways. On the one hand, if we have a fixed number of resources we can ascertain the time needed to evolve from the initial state to the final one (or between states), so we can check different configurations for a system before using it, thus saving an immense amount of time and money. On the other hand, if we start with some specification, we can find the appropriate number of resources of any type that we need in order to fulfil some time requirements. We can even calculate the minimum number of resources needed to obtain the best performance from a system.

Apart from the usefulness of the analysis that BAL performs, the main advantage of this tool is the fact that it is capable of dealing with real-world systems of a considerable size. This has been achieved through painstaking work linking different implementation strategies that will be discussed later.

In this paper, first the language BTC will be outlined. Secondly, BAL tool will be presented. Following this presentation, BAL is used in a case study which will highlight how BAL works and how the results obtained can be useful. Finally, conclusions and outlines for current and future work will be presented.

2 Related Work

We have found suggestive tools based on formal methods in the literature but all of them, for one reason or another, stray away from the characteristics and goals of the BAL tool.

The vast majority are based on Petri nets or timed automata. Among the former the closest to our work are timed Petri nets like PEP [6] or TINA [7] but both of these are software environment to edit, simulate and verify Petri nets and do not deal with performance evaluation.

The main tools based on timed automata are Kronos [2] and UPPAAL [1, 5] and it is a branch of UPPAAL (UPPAAL Cora [3]) which performs a similar evaluation to that carried out by BAL although the underlying formal methods used in each one are different. UPPAAL Cora uses a priced (or weighted) timed automata as the underlying modelling formalism. Priced timed automata extend classical timed automata with cost information on location (and edges) giving the price per time unit for staying in that location. In this way every run of a priced automaton has a global cost. BAL uses a process algebra (BTC). BTC specifications are straightforward and intuitive, as they are very similar to the manner in which any user can conceive a system. This is thanks to a feature of BTC which marks a great difference from other algebras, namely that resources do not need to be modelled as processes. This technique is widely used in formal modelling, but it is just an approximation to reality. The possibility then arises that, once the system has been specified by BTC, it may be worthwhile translating it to automata and using a tool (like UPPAAL CORA) for performance evaluation. Clearly, it is easier to make a translation than a performance evaluation tool. Unfortunately, this is not feasible for several reasons, of which the following two are the most important. First, an action may eventually need more than one resource for execution, which can be modelled by BTC where, moreover, a clear distinction in the use of preemptable and non-preemptable resources can be made. We have not been able to model this situation in UPPAAL Cora, which does not mean it is not possible but simply not at all intuitive. But the second –and main– problem has arisen with the fact that BTC can split the execution time of actions to achieve both real parallelism and interleaving (noted in next section) while using UPPAAL Cora this is not possible.

3 The BTC Language

In this section, we will not make an attempt to explain in depth the timed process algebra BTC. As the newest approach can be found in [8], we shall limit our explanation to a brief overview. We begin by listing its main features:

- Timed algebra
- It takes into account that the available resources in a system must be shared by all the processes. This evolves two types of delays:
    - related to the synchronization processes.
    - related to the allocation of resources.

- *True Concurrency* —> $(a|b) \not\equiv (a.b + b.a)$
- Homogeneous / Heterogeneous Resources
- Preemptable / Non-preemptable Resources

The specifications obtained by this algebra are quite easy to perform because it is fairly straightforward to represent concurrency, parallelism, resource sharing, conflicts, mutual exclusion and non-determinism.

We will deal with three kinds of actions: *timed actions* (Act_T) which use time (and resources if needed); *untimed actions* (Act_U) which use neither time nor resources (actions for synchronization); and *special actions* (Act_S), which use resources but no time. Let $c \in Act_S$ be a *special action*, which is used to deal with non-preemptable resources which must be requested and released. So, for each non-preemptable resource we have an action c to request the resource and the corresponding *conjugate* action $\hat{c} \in Act_S$ which is executed when the resource is released.

The syntax of BTC is defined by the following BNF expression:

$$P::= stop \mid a.P \mid \langle b, \alpha \rangle.P \mid P \oplus P \mid P + P \mid P \parallel_A P \mid recX.P$$

where $A \subseteq Act_U$, $a \in (Act_U \cup Act_S)$, $b \in Act_T$, and $\alpha \in \mathbb{N}$, $\mathbb{N}$ represents the set of natural numbers. Furthermore, we assume a set of process variables Id ranged over by X, X', a set of processes P ranged over by P, Q, R, and a set of actions $Act = Act_U \cup Act_T \cup Act_S$.

Furthermore, this syntax has been extended with a view to representing both the actions that use the shared resources and the amount of resources the system has at its disposal:

$$[[P]]_{Z,N}$$

where $Z = \{Z_1, Z_2, \ldots, Z_m\}$. This means Z is a set consisting of a set for each different type of shared resource (Preemptable/Non-preemptable). Moreover, $m \in \mathbb{N}$ is the number of different types of shared resources available on the system and $x_i \in \mathbb{N}$ is the number of actions which need at least one resource of type i for their execution. Each of these sets is defined as follows:

$$Z_i = \{b_1, b_2, \ldots, b_{x_i}, c_1, c_2, \ldots, c_{x_i}\}$$

and represents the set of actions which require for their execution at least one of the shared resources of the type i. Note that in Z_i, *conjugate* actions are not included just for the sake of clarity since if action $c \in Z_i$ then $\hat{c}$ should follow. We consider

$$N = \{n_1, n_2, \ldots, n_m\}$$

where $n_i \in \mathbb{N}$ represents the amount of shared resources of type i available in the system.

By incorporating these changes, our timed process algebra BTC is able to model processes which need to use different types of resources in their execution (heterogeneous resources), dealing with preemptable and non-preemptable resources and taking into account the number of all available resources in the system at any given

time. This is an important step since we are able to model any kind of delay which can appear in a system, i.e., we are able to deal with delays related to the synchronization of processes and delays related to the allocation of resources.

By means of operational semantics, we provide the language operators with a meaning and an accurate interpretation by describing how one process is able to convert into another.

4 BAL Tool

With this formal resource-aware model the timed characteristics of the system are captured. The next step is concerned with carrying out performance evaluation: we want to be able to estimate the minimum time needed to reach a given state. By applying the rules of the operational semantics, we build a transition graph where we can abstract the information about actions and consider only the information about time (duration of actions). This graph is a weighted directed graph, where weights are always positive numbers, and the problem to solve is finding the shortest path from the initial node. Usually, the number of states in the transition graph for a real system is huge so the ability of a tool to perform this task automatically becomes relevant. With this idea in mind the BAL tool has been developed, which, moreover, has been improved with some other useful features: a syntactic analyzer and a graphic interface with a user assistant.

Thus, the BAL tool can be divided into three stages. First, a syntactic analyzer checks the specification of the system to be studied. If the system specification is syntactically correct, the second stage initiates. Here the tool builds the relevant transition graph by applying the rules of the operational semantics and lastly the minimum time needed to reach the final state from the initial one is calculated and the path obtained shown. Evidently, the most delicate part in the tool development has been relative to graph analysis where it was necessary to join different algorithms of pruning, dynamic load balancing and parallelization. This is so, since one of the main aims is for the BAL tool to be able to deal with real systems and not be limited to toy examples. The techniques used will be commented in the appendix section where some remarks about the implementation is shown. A general view of the BAL tool can be found in Figure 1.

4.1 Graphic Interface and Assistant

Besides being effective and efficient, the BAL tool should also be user-friendly and to that end, BAL includes a graphic interface and assistant. The BAL tool can thus be executed interactively in the command line or by making use of the graphic interface. Working in text mode it is necessary to supply a file with the specification of the system to be analyzed as an argument to the command *BAL*. The graphic interface requires little or no explanation because our standardization of the tools makes it easily recognisable for the user, and the new functions are also quite intuitive. A quick view of this interface can be found in Figure 2.

More attention needs to be given to the assistant, which has been developed to make specification tasks easier and avoid syntax errors whenever possible. This assistant consists of two main parts:

- Resource Assistant Z_i: This helps to set the number of shared resources of any type in the system and the sets of actions that need these resources during their execution. Depending on the type of resource, the type of the action will be different. For preemptable resources the actions will be timed actions while for non-preemptable ones the actions belonging to set Z_i will be special actions to request and release the resource.
- Process Assistant: When the actions that make up the process must be supplied, it is easy to make a syntax error, so the assistant deals with the syntax details and the user just needs to give the action name and its type.

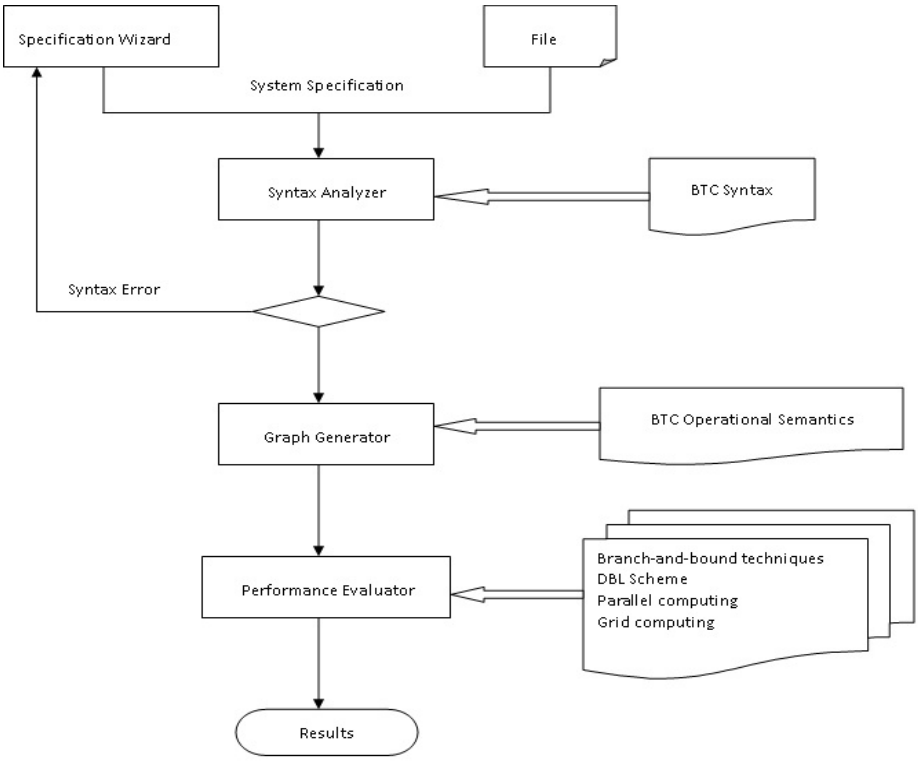


Fig. 1. Structure of the BAL tool

Once the BAL tool has the system specification, the syntactic analysis following BTC syntax is performed by the BAL tool. Later, the graph is built and the performance evaluation is carried out. The results are shown in the *notification area* and they reveal the following information:

- The weight of the shortest path to evolve from the initial state, i.e., the minimum units of time needed to execute all the processes in the system.
- The state of the nodes and the transitions in the path chosen as the shortest. That is, what actions are prepared to execute in any moment and the ones selected.
- The time that the tool needs to calculate this result.

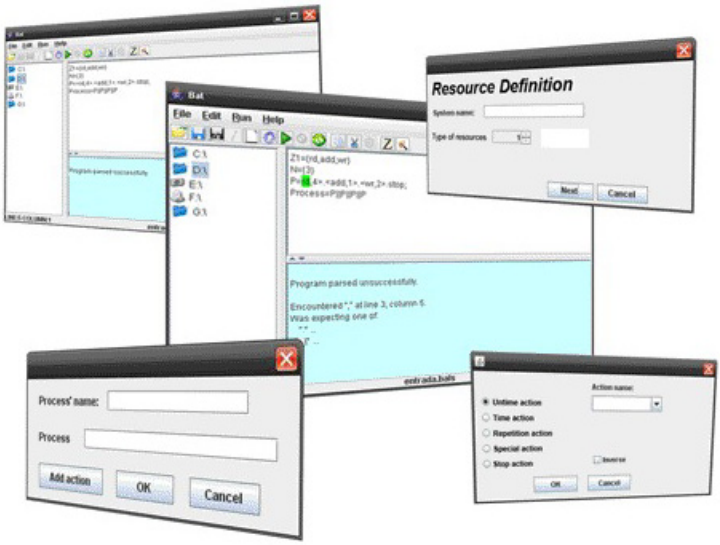


Fig. 2. Graphic Interface

5 Academic Case Study – Addition of Matrices

The following example serves to assess the developed tool. We check how BAL works, the degree of usefulness of the results and the time required to obtain them. Obviously, the process algebra BTC is able to specify more complex systems than this one but it has been chosen because its size is suitable to make a comparison between the different versions of the tool. The case study should not be unnecessarily large, since otherwise the state explosion problem would be hindered or prevented from drawing results in sequential version without pruning.

In the matrices addition system, only one type of resource is taken into account - the processor-, which will be named RecProc in order to avoid misunderstanding in the subsequent evaluation, when the number of processors in the platforms where BAL has been tested is indicated. Also, for the sake of clarity, it is assumed that there is the same number of processes as rows in the matrices to add. This assumption does not hinder the assessment and would be equally valid if we were to consider processes instead of threads or columns instead of rows. Hence, the system specification for the matrices addition in BTC is as follows:

$$Z_1 = \{add\_i\} \quad N = \{x\}$$

$$[[\ Sys\_add\ ]]_{Z,N} \equiv [[P_1 \parallel P_2 \parallel \cdots \parallel P_i \parallel \cdots \parallel P_n]]_{Z,N}$$

$$P_i \equiv < rd_i, 4 >. < add_i, 1 >. < wr_i, 2 >. stop$$

where rd_i is an action meaning "read a row", add_i is "add a row" and wr_i is "write a row". Note that according to the BTC syntax, the number that accompanies the action means the time required to complete its execution.

Then, from the system specification and by means of the BAL tool, the transition graph is obtained by applying the rules of the operational semantics, and the minimum time needed to finish the operation, depending on the size of the matrices and the number of processor resources (RecProc), is computed. Considering two matrices with three rows each and two RecProc ($N = \{2\}$) in the system, the smaller case, we find that the number of nodes is not substantial (79). Since, however, a previous simplification has been considered, the largest number of available actions which can be executed simultaneously has always been chosen. Without this simplification, even this case acquires unwieldy proportions, exceeding 500 nodes. There are several paths with minimum running time.

The information provided by BAL is that using these parameters the minimum time required to carry out the addition is 11 units of time. Now, using BAL the same evaluation is made for a 10x10 matrix and a different number of RecProc. The results obtained are displayed in Table 1 and in Figure 5 the results for various matrix sizes and a different number of RecProc are shown.

These results help us to show one of the ways in which BAL can be used. Given a system specification, BAL carries out an analysis of different possible configurations with the aim of achieving the appropriate number of resources that fulfil timed system requirements. From the results obtained, it is easy to deduce that when 10x10 matrices are added, and having 3 RecProc instead of 2, are at its disposal, a considerable improvement is obtained. Nevertheless, when there are 7 Recproc, it is useless to add 1 more (or even 2 more) because the time necessary to complete the task does not improve and will remain at 11 units. Hence, the other RecProc become idle.

So far, this case study has been used to show how the tool works and how the results obtained can be used. Furthermore, this example is one that we can use to check if the results obtained by BAL are correct. The verification was manually performed and experimentally obtained the same results.

Another analysis on the BAL tool worthy of mention is that of the improvements achieved by pruning and parallel processing. Table 2 shows a comparison made between the sequential version with no pruning and with pruning. 10x10 matrices have been added and the time needed by BAL to carry out the assessment based on the number of RecProc into the system has been obtained. The gain observed is quite substantial even in this simple example and it is worth noting that many cases, which were impossible to study previously due to memory constraints or an unacceptable execution time, can now be studied in a reasonably short period of time.

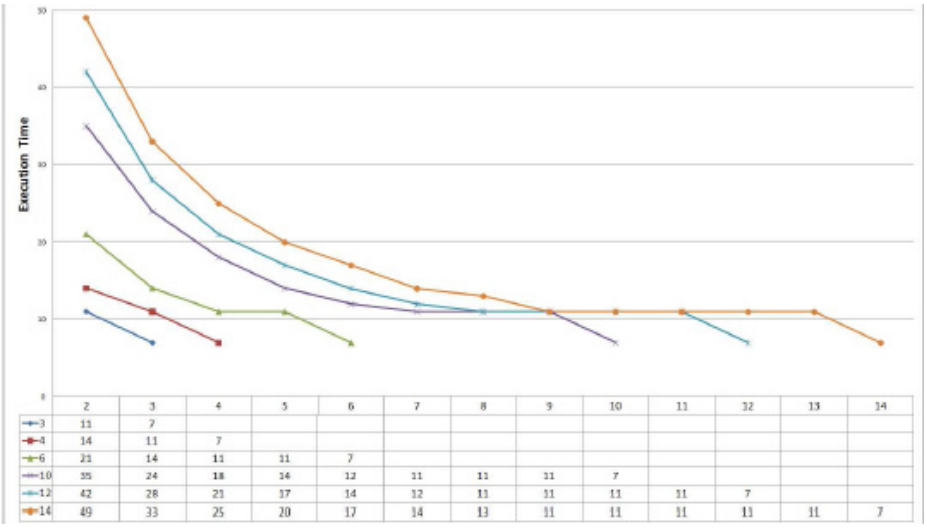


Fig. 3. Sum of 2 matrices with various matrix sizes and different number of processors

Table 1. BAL Results

10 x 10 Matrices	
RecProc	**Min Time** for the addition
2	35
3	24
4	18
5	14
6	12
7	11
8	11
9	11
10	7

Table 2. Execution times

Sequence Version		
RecProc	10 x 10 Matrices	
	Without prunning	**With** prunning
2	----	2’ 41’’
3	170h 21’ 3’’	8’ 42’’
4	51h 12’ 4’’	6’ 17’’
5	22h 45’ 21’’	1’ 42’’
6	5h 12’ 34’’	9 ‘’
7	1h 23’ 5’’	2 ‘’
8	22’ 24’’	<1’’
9	22’’	<1’’
10	<1’’	<1’’

Moreover, as discussed above, in order to reduce as much as possible the memory requirements and time needed to perform the evaluation, a parallel version of the tool has been developed. The time taken by BAL with this new version depends on the platform used but, as a reference, it can be said that in a cluster with 4 machines consisting in 8 cores each and 32 GB RAM, BAL only needs 9'13" to analyze a system with 100x100 matrices and 2 RecProc and 1000x1000 matrices in less than 30' (depending on the number of RecProc).

6 Conclusions and Future Work

In this work we have presented the BAL tool which has been developed to carry out performance evaluation. Besides this main goal, BAL tool has been improved using a syntactic analyzer and a graphic interface including a help assistant when the system specification is written. Therefore, the BAL tool consists of the following parts:

- Graphic interface
- Assistant for system specification
- System analyzer
- Graph generator
- Performance Evaluator

A case study has been made to show how the results obtained by BAL can be used; with a fixed number of resources we can ascertain the time needed to evolve from the initial state to the final one, so we can check different configurations for a system before using it, hence saving a huge amount of time and money. We can even calculate the minimum number of resources needed to obtain the best performance from a system.

Currently, our work focuses on two different lines. The first is concerned with improving the tool and the second one deals with its application.

As stated above, the tool uses parallel and grid computers with distributed memory, however we have decided to take advantage of improvements in hardware architecture and multicore machines and have used threads which will communicate by means of shared memory.

The main field where BTC is proving its worth is in the performance evaluation of Flexible Manufacturing Systems (FMS). A competitive manufacturing system is expected to be flexible enough to respond to small batches of customer demand. Due to the fact that the construction of any new production line is a large investment, current production lines must be able to be reconfigured to keep up with increased frequency of new product designs. Therefore, in the development of a FMS, particular attention must be paid to the design phase, where the performance of the system has to be accurately evaluated so that the most appropriate number and type of resources are selected from the initial design stages.

In this context, BTC and BAL are of great utility given that they allow us to check different configurations for a flexible manufacturing system before being established or for future improvements.

References

1. Behrmann, G., David, A., Larsen, K.G.: A Tutorial on UPPAAL. In: Bernardo, M., Corradini, F. (eds.) SFM-RT 2004. LNCS, vol. 3185, pp. 200–236. Springer, Heidelberg (2004)
2. Bozga, M., Daws, C., Maler, O., Olivero, A., Tripakis, A., Yovine, A.: Kronos: A Model-Checking Tool for Real-Time Systems. In: Vardi, M.Y. (ed.) CAV 1998. LNCS, vol. 1427, pp. 298–302. Springer, Heidelberg (1998)
3. Behrmann, G., Brinksma, E., Hendriks, M., Mader, A.: Production Scheduling by Reachability Analysis: A Case Study. In: 19th International Parallel and Distributed Processing Symposium (IPDPS 2005), p. 19 (2005)

4. Hoare, C.: Communicating Sequential Processes. Prentice Hall (1985)
5. Larsen, K.G., Pettersson, P., Yi, W.: Uppaal in a Nutshell. Int. Journal on Software Tools for Technology Transfer, 134–152 (1997)
6. Pep - Programming Environment based on Petri nets, http://parsys.informatik.uni-oldenburg.de/pep/
7. Tina - TIme petri Net Analyzer, http://www.laas.fr/tina/
8. Ruiz, M.C., Cazorla, D., Cuartero, F., Pardo, J.J.: Analysis of the SET e-commerce protocol using a true concurrency process algebra. In: 21st ACM Symposium on Applied Computing (SAC 2006), pp. 879–886 (2006)

Appendix: Implementation

As discussed above, to solve the problem of minimizing the time needed to reach the final state (from initial) a recursive algorithm has been implemented. To find such a solution an in-depth traverse of the graph is executed. For each of the explored nodes, all its successors are generated before another node is explored. After each expansion the new child is again selected for expansion. At this moment the first pruning may be introduced. If the actual cost is greater than one obtained from a previously analyzed branch, then there is no need to return to the nearest point of unexplored decision alternatives because it can be guaranteed not to improve on the solutions found so far.

Despite this pruning, it can be observed that as the size of the system increases, the number of nodes in the corresponding graph increases considerably, so the time needed to obtain the final result and memory requirements increases significantly. In a detailed study on the graph of transitions generated during the analysis of these systems, the possibility of further pruning methods, which improve the performance of the tool, was discovered. These new techniques are based on branch and bounds algorithms, which can be used to simplify the search in the space of states.

The first action taken was not to study those states that had already been analyzed previously in some branch. Obviously, this increases the complexity of the search strategy, requiring as it does, storing in memory all the generated states. Therefore, the new modification of the algorithm is similar to the standard state-space traversal algorithm, which uses two data structures for storing nodes; nodes to be explored (BROWSE) and nodes that have already been explored (VIEW). Therefore, when a node is going to be analyzed, before calling upon the recursive function to continue the search in depth, it is necessary to check if the node is stored in the list VIEW.

BAL has also been further improved by parallel and grid computing to try to tackle the state explosion problem. In fact, by using parallel and distributed computing, the execution time has been dramatically reduced and, in addition, memory constraints have become less restrictive. That is to say, the initial sequential algorithm has been parallelized so that multiple processors are able to study different parts of the graph. The strategy used for parallelizing the search algorithm involves partitioning the graph into subdomains (subgraphs). The subgraph can then be distributed over the processors responsible for conducting the calculation in parallel. In this task a parallel branch and bound algorithm is used. Ideally, the search time would be reduced in proportion to the number of processors that cooperate in the search but the overload introduced by the

cooperation of the processors must be taken into account, which reduces the speedup. It is therefore necessary to examine the quality of the partitioning based on its effect on the application code, which requires an in-depth study of the load balance.

The computational work of each processor should be balanced, but unfortunately the state space generated by BAL is not balanced. The graph has an unknown distribution, so that one of the problems of this algorithm is the distribution of the search space in each processor. The difficulties in partitioning the search space in a balanced way, leave us with two alternatives to distribute the workload among processors: static load balancing (SLB) and dynamic load balancing (DLB). The first is useful when the work load is known from the start, as the work is distributed equally across processors and no extra cost for balancing the load is involved. But BAL generates the graph dynamically so that dynamic load balancing is needed. The main idea in this type of balancing is that when a process finishes its work, it gets more from a process which still has unstudied branches.

To make the initial assignment of subgraphs to different processors, a breadth-first traverse is carried out at the first levels of the graph until the number of nodes is greater than or equal to the number of available processors. Then, each of these nodes is assigned to a processor. From this moment on, each processor analyzes the assigned subgraph by means of a depth-first algorithm by applying branch-and-bound techniques, which generates new nodes to study. The technique used to make the work balance between the processors is Random Polling (RP), a simple yet effective randomized DBL scheme. Every processor works on at most one subgraph at a time. In addition, each processor has a stack, which stores the expanded nodes. A processor whose subgraph is completely studied polls randomly determined processors until it finds one which is busy. The busy processor selects one of its expanded nodes and transmits it to the idle processor. In the original RP scheme, the work space is initially assigned to a processor and other processors have no work. But, in our experience, by breadth distribution at the beginning, the best results are achieved.

Step towards Paper Free Hospital through Electronic Health Record

Maria Salazar[1], Júlio Duarte[2], Rui Pereira[2], Filipe Portela[3], Manuel Filipe Santos[3], António Abelha[2], and José Machado[2]

[1] Centro Hospitalar do Porto, Oporto, Portugal
msalazar@chporto.min-saude.pt
[2] Universidade do Minho, Departamento de Informática, CCTC, Braga, Portugal
a47060@mail.uminho.pt,
{jduarte,abelha,jmac}@di.uminho.pt
[3] Universidade do Minho, Departamento de Sistemas de Informação, ALGORIMI Center, Guimarães, Portugal
{cfp,mfs}@dsi.uminho.pt

Abstract. Information technology has great potential for transforming the health care system, improving quality of care. With the increasing expansion of health information systems, the Electronic Health Record (EHR) has become one of the finest sources for clinical information aggregators in the context of digital health. The EHR is a core part of a hospital information system, as well as a service on duty of the patient to improve the treatment of patients. It can be considered as a longitudinal electronic record of patient heath information, for example vital signs, medical history or laboratory data, generated by one or more encounters in any care delivery setting. As the EHR offers many potential opportunities for healthcare systems, it is important to take steps to improve the system. With this in mind, a study of the features present on a Portuguese EHR was made. The basis of this study was an adoption model that evaluates the EHR system accordingly to its current features. After this study, the EHR will be ranked into one of the existing eight stages.

Keywords: EHR, EMRAM, HIMSS.

1 Introduction

The sustained demand by healthcare organizations to improve the quality of patient care and patient safety boosted the adoption of information and communication technologies (ICT). Along with the computerization of hospitals, a new stage began and with it the emergence of new concepts, paradigms and methodologies to solve problems in this domain of knowledge. Healthcare, among other industries, has already profited extensively by the growth of ICT, with the objective of serving not only physicians but also other professionals in their daily work with patients. The most common application of those types of systems is the electronic health records (EHR) system.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 685–694.
DOI: 10.1007/978-3-642-36981-0_63

More than just a computerized version of a paper medical chart, over time, the EHR can allow a patient's providers to share important information, across different health care settings, while maintaining EHR patient confidentiality in accordance with federal and state privacy and security requirements. The patient health information in an EHR includes allergies, radiology images, lab and test results, medical history, diagnoses, medications, and immunization dates. Besides the inclusion of all the functions provided by the process clinic in paper, other features are present, such as decision support systems, personified warning or automatic alerts and messages. The EHR has proved to be an excellent tool for healthcare organizations. But the continuously seek for improving the system is necessary. The task of maintaining a well-functioning health service requires continual adjustments and sometimes also the introduction of new functions and it is also possible to turn the system more usable to the healthcare providers. Those modifications will increase the efficiency; reduce errors, which leads to improve the patient safety, along with other benefits.

In this context, this article highlights an EHR hospital system that was evaluated in accordance with an adoption model. This study was made to understand the level of the current EHR, and ascertain what can be improved in order to turn the hospital **into** a paperless environment.

This article contains six others sections. The subsequent section explains the notion of electronic health record. In the third section, the HIMSS organization is addressed. The fourth section describes the methods used to perform the categorization of the EHR's level. The remaining sections feature the results, as well as its discussion and lastly the final conclusion is presented.

2 The Electronic Health Record

The Electronic Health Record (EHR) is a core application, which covers horizontally the health care unit and makes possible a transverse analysis of medical records along the several services, units or treated pathologies [1]. Despite the fact that there is not one exact definition for EHR, it can be defined as the computerized records of patients' clinical data. This information, which can be clinical, administrative or financial, is inserted in an electronic system that enables the capture, maintenance, transmission and storage of clinical information which is essential, not only for the monitoring of the health status of each patient but also for proposes such as cost management [2] [3] Thus, an EHR is an assembly of standardized documents, ordered and concise, directed to the register of actions and medical procedures; a set of information that can be compiled either by physicians or other health professional; a register of compiled facts, containing all the information regarding patient health data; and a follow up of the risks values and clinical profile [1].

All the data are collected accordingly to the Problem Oriented Medical Record (POMR) method. This is a format for clinical recording consisting of a problem list, a database including the patient history with physical examination and clinical findings, diagnostic, therapeutic and educational plans and a daily SOAP (Subjective, Objective, Assessment and Plan) progress note. The problem list serves as an index

for the reader, each problem being followed through until resolution. This system widely influences note keeping by recognizing the four different phases of the decision making process: data collection; the formulation of problems; the devising of a management plan; and the reviewing of the situation and the revising of the plan if necessary [4] [5].

With the adoption of the EHR it was possible to acquire the versatility of a device capable of storing a vast sum of data. This was the great advantage obtain with the transition of Paper Clinic Process (PCP) to EHR. The data legibility, the continuous data processing, the ability to detect errors or releasing alarms concerning eventual pathological anomalies can also be considered as positives features acquired with the implementation of EHR. With the inclusion of this advantages, the patient assistance has become more effective, faster and with better quality.

Despite the fact that the EHR system is fully implemented at the hospital, there are always particular modifications that can be made in order to improve the system, which is crucial to achieve success. The referred adjustments can be involved in the integration of new features or by improving the work conditions (reduce errors, improve patient safety and care quality, increase efficiency). Furthermore, although the EHR system is a replacement of the PCP, the hospital still uses paper charts to deliver and manage patient care and has a mixture of discrete data, document images and medical images within its EHR environment.

With those considerations in mind, it was extremely important to evaluate the level of the EHR. Furthermore, it is vital understand with new features can be included to achieve the main goal, which is to turn the hospital into a paperless environment. One organization produced one adoption model that can be a useful tool in this task. Lets know a bit more about that organization, HIMSS, and its adoption model, the EMRAM.

3 The HIMSS

The Healthcare Information and Management Systems Society (HIMSS), founded over 50 years in Georgia Institute of Technology, is a cause-based, not-for-profit organization exclusively motivated on providing global leadership for the optimal use of information technology (IT) and management systems for the furtherance of healthcare [6]

HIMSS, along with its related organizations, are headquartered in Chicago with additional offices in the United States, Europe and Asia, representing more than 44,000 individual members, of which more than two thirds work in healthcare providers, governmental and non-profit organizations [6].

HIMSS frames and leads healthcare practices and public policy through its content expertise, professional development, research initiatives and media vehicles designed to promote information and management systems' contributions to improve the quality, safety, access and cost-effectiveness of patient care [6].

More recently, two affiliates were created: HIMSS Foundation and HIMSS Analytics. The mission of the last one is to provide the highest quality data and

analytical expertise to support improved decision-making for healthcare providers, healthcare IT companies and consulting firms. This subsidiary is also responsibly for the development of the Electronic Medical Record Adoption Model (EMRAM), which is an eight steps process to understanding the level of Electronic Medical Record (EMR) capabilities in hospitals [7]. The model identifies the levels of EMR capabilities ranging from ancillary department systems through a paperless EMR environment. It is important to highlight that the term EMR is similar to EHR. The only difference between those two terms is that an EMR only includes the patient information of one healthcare organization and the EHR contains information of different healthcare organizations [2].

3.1 Electronic Medical Record Adoption Model (EMRAM)}

The EMRAM was created in 2005 to reflect how individual hospitals and integrated delivery systems (IDS) in the United States and Canada adopt information technology. Along with this, the EMRAM can be used as a guideline to achieve higher levels of IT adoption in way to improve patient safety, care quality, health information exchange and a paperless environment [8]:

To submit a hospital to this model, a comprehensive questionnaire that lists up to 250 potential applications and the hardware supporting the application has to be completed. In addition to this, information about hospital's budget and staffing, the information system (IS) department budget and staffing can be collect in order to a benchmarking can be built. After gathering this up, HIMSS Analytics software analyses the gathered data and generates a score from zero to 7000, which will indicate the hospital's stage. This way, all aspects of adoption, even portions of stages not yet completed, are awarded \cite{[WEB-Hoyt]}. However, each step has criteria that must be met before the stage level can be awarded. For instance, a hospital cannot be a stage 5 if they have not completed all the criteria present on stage 4, even though they may meet all conditions for stage 5. The logic is all aspects of a stage must be met to complete a stage. Brief descriptions of all criteria present in each stage are presented following [8]:

Stage 0: At this stage are present all the organizations that do not have installed any of the key ancillary department systems, such as laboratory, pharmacy or radiology. In addition to this, it is required that the organization is not processing laboratory, pharmacy and radiology data output online from external service providers.

Stage 1: To achieve this stage, the major ancillary clinical systems have to be installed. Another possibility to accomplish it is if the laboratory, pharmacy and radiology information system data output is delivered to the hospital for online access and processing in case of those ancillary services are not provided in-house, but by external service providers.

Stage 2: It is required at this stage that the major ancillary clinical systems feed data to a system that provides physician access for retrieving and reviewing patient-centered results. The referenced system may be an Electronic Patient Record (EPR) system or a clinical data repository (CDR) which is fed by and feeding back into sub-systems. It might contain a controlled medical vocabulary (CMV) tool such as SNOMED to transfer results into a format that can be incorporated into the EMR as structured data. It also may contain the clinical decision support/rules engine for rudimentary conflict checking. Information from document imaging systems may be linked to the system at this stage. The hospital should be health information exchange (HIE) capable at this stage and can share information in the EPR/CDR with other patient care stakeholders.

Stage 3: Nursing/clinical documentation (e.g. vital signs, flow sheets, nursing notes, care plan charting) and/or the electronic medication administration record (eMAR) system and Order Entry/Communications are required, and are implemented and integrated with the EPR/CDR for at least one service in the hospital. The first level of clinical decision support may be implemented to conduct error checking with order entry (i.e., drug/drug, drug/food, drug/lab conflict checking normally found in the pharmacy). Some level of medical image access from picture archive and communication systems (PACS) may be available for access by physicians outside the Radiology department, e.g. via the organization's intranet.

Stage 4: Computerized Practitioner Order Entry (CPOE) for services (e.g. radiology, laboratory, operating room, etc.) and/or medication (i.e. ePrescribing) is added to the nursing/clinical documentation and EPR/CDR environment. For instance, if one patient service has implemented CPOE for use by any clinician and with physicians entering orders and completed the previous stages, then this stage has been achieved. Second level of clinical decision support capabilities related to evidence based medicine may be available.

Stage 5: This stage has only one requirement: the existence of a full complement of PACS systems provides medical images to physicians via an intranet and displaces all film-based images. Therefore, if a hospital contains it and has completed all the previous stages, this stage five has been achieved.

Stage 6: It is obligatory at this stage to have implemented, for at least one patient care service area, a full physician documentation/charting. A clinical decision support system (CDSS) provides guidance for all clinician activities related to protocols and outcomes in the form of variance and compliance alerts (i.e. third level of clinical decision support). Some form of structured templates is required to capture discrete data for physician documentation interaction with CDSS. The closed loop medication

administration environment is fully implemented. The electronic medication administration record (eMAR) is implemented and integrated with CPOE/ePrescribing and/or pharmacy to maximize point of care patient safety processes for medication administration. Bar coding or other auto identification technology, such as radio frequency (RFID), automated dispensing machines (ADM) or double e-signature by administering nurses ensure the protection of the 5 Rights'.

Stage 7: At this stage, the hospital no longer uses paper charts to deliver and manage patient care and has a mixture of discrete data, document images and medical images within its EMR environment. Also, Clinical Data Warehouses are being used to analyze patterns of clinical data to improve quality of care and patient safety and to feed outcomes reports, Quality Assurance and Business Intelligence. The hospital demonstrates summary data continuity for all hospital services (e.g. inpatient, outpatient, ED and with any owned or managed ambulatory clinics). Clinical information can be readily shared via standardized electronic transactions (e.g. Continuum of Care Document) with all entities that are authorized to treat the patient or a health information exchange (i.e. other non-associated hospitals, ambulatory clinics, sub-acute environments, employers, payers and patients in a data sharing environment).

3.2 The Prestigious Stages

Of the eight stages present in the EMRAM, the stage six and the stage seven are the most prestigious being the stage seven the pinnacle of an environment where paper charts are no longer used to deliver patient care. Hospitals that achieve these two stages are recognized to represent best practices in implementation of the EHR, mastering the key challenges such as closed loop medication administration, use of CCD transactions to share data with other organizations or the inexistent of no more film in radiology for example.

To achieve those stages, other requirements aside the collecting of data have to be completed. The validation of stage six is only completed with a phone interview. The final evaluation of the stage seven is conducted by an on-site visit \cite{[WEB-Europe]}. This shows the importance of assign those stages to a hospital.

In the United States, there are a lot of hospitals awarded with the stage seven, more precisely 88 hospital [7]. California is the state that leads with 37 stage seven hospitals. However, in Asia there are twelve stage six hospitals and only one hospital has been recognized with the stage seven award, the Seoul National University Bundang Hospital [7]. In Europe, the University Medical Center Hamburg-Eppendorf (UKE) in Germany is the only one that achieved stage seven of the EMRAM [8]. There is not a single Portuguese hospital awarded with a stage seven or six.

4 Methods

The major objective in this phase was to identify all the features that are available at the current EHR system in the hospital. Although it is whispered that a one-on-one interview or a brainstorm are the greatest methods to gather this type of information, this was not done. Instead, an inside study was performed. The decision for this method in opposition to those referred before is simple. The healthcare providers do not have availability to spend time to join interviews. The study solution is a better choice in this area because it not affects the work of the health professionals.

In order to conduct this analysis properly, a guideline was created. As the guideline was elaborated accordingly to the EMRAM, it is based on the different requirements listed along the eight stages of the model. Thus, the proposal of this guideline is to identify which characteristics are present in the system. Along with the identification of each feature, this study turn also possible to find out which features are present but are not used by the users. Or even if the users know the existence of those features. This way, those features can be recognized and further investigation can be performed in order to establish the reason of not being used. With this knowledge, adjustments can be implemented in order to overpower that drawback. Therefore, this study was conducted at the totality of the hospital's units.

Below is the list of some points presented in this guideline:

- Reading reports in electronic format;
- Management of the list of inpatients;
- Requests for analyzes or examinations;
- Prescription of drugs;
- Registration / reading of nursing notes;
- Usage of paper documentation in the management and delivery of patient care;
- Existence of electronic medication administration record.

With the outcome of this study, it is expect to rank the EHR into one of the eight levels of the EMRAM. It would be of great pleasure to the hospital if the EHR achieve one of the prestigious stages (stage six or stage seven). However, if the EHR cannot achieve such stages, there are no reasons for disappointment. Instead, it can be essential to understand what have to change in order to turn it into a better system. The objective of the hospital is to become a paperless environment, which means be recognized with a stage seven.

5 Results

This was an extremely meticulous work, analyzing everything present in the hospital, to further comparison with the EMRAM and see in what level the EHR is nowadays. This exercise has never been done before to this specific EHR system, so comparisons with previous work do not fit in here.

Through the results obtained from this studied in the hospital, it was possible to identify all the features present in the system. All the major ancillary clinical systems, such as laboratory, pharmacy and radiology, were installed in the hospital. Furthermore, all the information derived from those systems is feed to the EHR, which allows the physicians to have access for retrieving and reviewing the patient-centered results.

In order to help those clinical systems, along with others, the hospital implemented a medical imaging technology, in order to provide economical storage and convenient access to images from different modalities. The referred technology is a picture archiving and communication system (PACS), which allows the digital transmission of images and reports, eliminating the need of manually files or films.

The hospital is Health Information Exchange (HIE) as exchange information with other hospitals and health centers. This means that it has the capability to electronically move clinical information among disparate health care information systems while maintaining the meaning of the information being exchanged. Being HIE the access to clinical data is facilitate, as well as its retrieval, in order to provide safer and more timely, efficient, effective, and equitable patient-centered care. The hospital owns equally Clinical Data Warehouses, which are used, for example, to analyze patterns of clinical data to improve quality of care and patient safety and to feed outcomes reports.

The EHR contains a controlled medical vocabulary, the Systematized Nomenclature of Medicine (SNOMED), and it also has installed other useful tools like the ninth revision of the International Classification of Diseases (ICD-9). Those tools, not only facilitates the work of the entire workers in the hospital, but also improve other tasks such as indexing. All the information stored at the EHR can be seen by all the patient care stakeholders. This system also contains nursing/clinical documentation where the healthcare providers can record the vital signs of patients and nursing notes for example. All the order entries are made electronically and integrated with the EHR at the totality of the hospital's services. Along with this, it has mechanisms to prevent errors, such as error checking with order entry.

Physicians have the possibility of, electronically, entry instructions for the treatment of patients under his or her care. This is possible once the EHR has integrated the Computer Practitioner Order Entry (CPOE). However, the CPOE has other benefits, like patient-centered decision support, real-time patient identification, drug dose recommendations or check for treatment conflicts. Integrated with the CPOE is the electronic medication administration record (eMAR), which is responsible to enlist bar coding technology in order to submit and fill prescriptions with hand-held scanners that read bar codes and transmit them to the terminal/workstation using Bluetooth wireless technology. The eMAR relays drug dosage data to the pharmacy, which is then connected to the nursing stations.

Finally, it was possible to observe that the intensive care, one of the different patient care service areas present in the hospital, is a paperless environment. All the physician documentation/charting is made electronically, without paper. However, the majority of the patient care services areas still feel the necessity to use paper charts to deliver and manage patient care. That is not inevitably for everything accomplished in that service, but in the fulfillment of some actions, they have to use paper.

6 Discussion

Comparing the results obtained with the specification of the EMRAM, it is possible to identify the current stage of the EHR. Once the major ancillary clinical systems are installed in the hospital, the stage one is achieved. The data from those systems is included in the EHR, which has a controlled medical vocabulary. This, and the fact that the hospital is capable of exchange health information make the EHR ranking the second stage.

The third and fourth stages are also taken, because the EHR fulfills all the requirements present in those stages. The nursing and clinical documentation, electronic medication administration record system and order entry/communication are implemented successfully in the hospitals' EHR, as well as the CPOE. The system PACS is also present in the hospital, therefore the stage five is achieved.

It was possible to state that the intensive care was the only one patient care service area, which was a paperless environment. This means that the stage seven cannot be achieved, as the entire hospital has to be free of paper. But what is the conclusion about the stage six? Can the EHR be recognized as level six or 'only' level five?

Full physician documentation present in, at least, one of the services of the hospital is fulfill, as seen before. The third level of clinical decision support is also present in the EHR. The last requirement to achieve stage six is verified as well. The electronic medication administration record is fully implemented, and the medication administration is automatized by bar coding. Therefore, the EHR present in the hospital can be considered as a stage six.

Being recognized as a stage six hospital is very prestigious. However, more improvements have to be made to aim to the top, turn the hospital a paperless environment and become a reference regarding the EHR system. With this study, it was possible to state the level of the EHR. In addition to this, it turned possible to identify the remaining steps that need to be taken in order to improve the system and become a stage seven EHR.

A stage seven hospital demonstrates superior implementation and utilization of health IT systems, resulting of true sharing, information exchange and immediate delivery of patient data to improve process performance, quality of care and safety. That is the aim of the hospital, reach the higher level of the EMRAM, providing the best service not only for the patients, but also to all their workers.

7 Conclusions

In this paper it is presented an evaluation of the current EHR system, with the help of the EMRAM provided by the HIMSS. The awareness of the level of the EHR system is vital to know what has to be done in order to improve the system to achieve the main goal of turn the hospital into a paperless environment. Having the hospitals' EHR to be recognized as a stage six is a source of pride. However, it is pretended to achieve the stage seven and be in the top. With this in mind, this evaluation turn possible to identify what have to be done to aim that objective, which means

achieving the seventh level. After these modifications have been made, it is vital to done this evaluation again to see if the level seven in really achieved. The expectation is high as is the ambition.

Acknowledgments. This work is financed with the support of the Portuguese Foundation for Science and Technology (FCT), with the grant SFRH/BD/70549/2010 and within project PEst-OE/EEI/UI0752/2011. It is also supported by FEDER through Operational Program for Competitiveness Factors – COMPETE and by national funds though FCT in the scope of the project FCOMP-01-0124-FEDER-022674. The work of Julio Duarte was supported by the grant SFRH/BD/70549/2010 from FCT. The work of Filipe Portela was supported by the grant SFRH/BD/70156/2010 from FCT.

References

1. Machado, J., Alves, V., Abelha, A., Neves, J.: Ambient intelligence via multiagent systems in the medical arena. Engineering Intelligent Systems for Electrical Engineering and Communications 15(3), 151–157 (2007)
2. Machado, J., Abelha, A., Novais, P., Neves, J., Neves, J.C.: Quality of service in healthcare units. Int. J. Computer Aided Engineering and Technology 2(4), 436–449 (2010)
3. Duarte, J., Portela, C.F., Abelha, A., Machado, J., Santos, M.F.: Electronic Health Record in Dermatology Service. In: Cruz-Cunha, M.M., Varajão, J., Powell, P., Martinho, R. (eds.) CENTERIS 2011, Part III. CCIS, vol. 221, pp. 156–164. Springer, Heidelberg (2011)
4. Peixoto, H., Santos, M., Abelha, A., Machado, J.: Intelligence in Interoperability with AIDA. In: Chen, L., Felfernig, A., Liu, J., Raś, Z.W. (eds.) ISMIS 2012. LNCS, vol. 7661, pp. 264–273. Springer, Heidelberg (2012)
5. Miranda, M., Abelha, A., Santos, M., Machado, J., Neves, J.: A Group Decision Support System for Staging of Cancer. In: Weerasinghe, D. (ed.) eHealth 2008. LNICST, vol. 1, pp. 114–121. Springer, Heidelberg (2009)
6. Healthcare Information and Management Systems Society, `http://www.himss.org/ASP/index.asp`
7. HIMSS Analytics, `http://www.himssanalytics.org/home/index.aspx`
8. HIMSS Europe, `http://www.himss.eu/`

Remote Scientific Computing over the Internet: An Example in Geometry

Miguel Ferreira[1] and Miguel Casquilho[2]

[1] Department of Computer Science and Engineering, Instituto Superior Técnico, Technical University of Lisbon, Ave. Rovisco Pais, IST, 1049-001 Lisboa, Portugal
[2] Centre for Chemical Processes, Department of Chemical Engineering, Instituto Superior Técnico, Technical University of Lisbon, Ave. Rovisco Pais, IST, 1049-001 Lisboa, Portugal
mcasquilho@ist.utl.pt

Abstract. Scientific computing over the Internet can suit many activities that have not, in the authors' opinion, been explored enough in general. Resources such as executables, languages, packages, can be used from a remote computing system. In this study, largely based on academic practice, a simple illustrative example in Geometry is implemented on a distributed system that outsources the computing-intensive tasks to remote servers that may be located in other universities or companies, linked to grids and clusters and so on. The software stack and software developed to support the communication is explained in detail. The architecture developed stresses the interoperability of the software, and a high degree of decoupling between components hosted in various locations. The results of this study motivate further work and serve a practical purpose that may be useful to everyone doing scientific computing.

Keywords: Scientific computing, Internet, remote executables, university-industry links.

1 Introduction

Scientific computing over the Internet can nowadays suit many areas of activity, but, in these authors' opinion, has not been, in general, fully explored. The second author has, since more than a decade, intensively used this mode of computing in research and teaching at IST[1], in domains related to Engineering Mathematics, as Operational Research or Statistics. The computing has been done in a server of IST's information technology centre, intended mainly to lodge webpages. The present study, largely based on previous academic practice, focuses on the establishment of a link between two universities, one wishing to execute software made available by the other. This would also apply to any two entities, such as two companies or in a university-company linkage (a particular application of [5]). In the Internet context, resources adequate to the particular technical purpose, such as executables, languages, or packages, can be used, if accessible at this level, from a remote computing system.

[1] IST: *Instituto Superior Técnico*, engineering school of the Technical University of Lisbon, Lisbon (Portugal).

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 695–702.
DOI: 10.1007/978-3-642-36981-0_64

With the ease of communication of the Internet, a step can be taken to reap benefits from using remote resources. There are, of course, many resources for computing on the Web, dealing with small tasks, ranging from conversions of units to more complex mathematical problems. An extensive example of scientific computing over the Web is the original work by Ponce ([7]), containing various (Fortran) programs to solve problems dealing with hydraulics and related areas. However, most of these applications are deployed wholly on single nodes, which also host the web interface and logic. The present work intends to take this topology into the next stage, allowing further decoupling of components, by introducing an intermediate communication layer between distributed nodes, which together form the web computing system.

The adoption of Internet-based computing has been deemed by the 2nd author indispensable to his activities as a tool in the academic practice, and a gateway to the university-industry linkage, in an era of cheap information technology gear.

This study is based on a simple, yet surprising, illustrative example in Geometry, as might be used in a lecture, chosen both to be clear to a wide readership and to avoid beclouding the underlying software structure. From a webpage of one entity, the computing is done, without the user's perception, at another one (this suggesting the extension to more), allowing a certain software to be accessed.[2]

In the following sections: the illustrative example is briefly described in its mathematical aspects, and its resolution through the Internet is presented; the software architecture is shown; and some conclusions are drawn about the technique.

2 Illustrative Example

A problem in Geometry was chosen as an illustrative example. Let the minimum distance be sought between points A and B, source and destination, as seen in Fig. 1, both on the X-axis, passing by point P, to be determined, on the half line s making an angle γ with the axis. The problem is treated in [1] and solved by differential calculus. The analytical solution for $P = (X, Y)$ is given in Eq. (1).

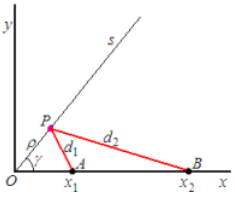


Fig. 1. Route from A to B, passing by P on s, for minimum distance

$$\frac{1}{X} = \frac{1}{2}\left(\frac{1}{x_1} + \frac{1}{x_2}\right)\sec^2\gamma \qquad (1)$$

[2] The first entity is the Faculty of Sciences, University of Lisbon, Lisbon (Portugal), and the second is IST, but their roles might be reversed.

It is $Y = X \tan\gamma$, thus Eq. (1) leads to Eq. (2):

$$X = 2\frac{x_1 x_2}{x_1 + x_2}\cos^2\gamma \qquad Y = \frac{x_1 x_2}{x_1 + x_2}\sin(2\gamma) \tag{2}$$

More concisely, in polar coordinates, (ρ, θ), with $\theta \equiv \gamma$, the radial coordinate is

$$\rho = 2\left(1/x_1 + 1/x_2\right)^{-1}\cos\gamma = 2\frac{x_1 x_2}{x_1 + x_2}\cos\gamma \tag{3}$$

The interest of this problem —the reason it was chosen— lies in the unexpected result as γ decreases towards 0. In Fig. 2, the optimum routes are shown, to which correspond the optimum positions of P, for various descending values of γ, always with $x_1 = 1$ and $x_2 = 3$. The results come from the authors' website ([2]). Now, intuition would possibly lead ρ to the *arithmetic mean* of x_1 and x_2.

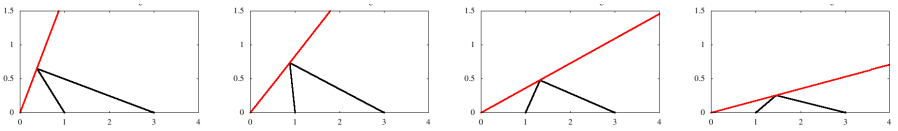


Fig. 2. Optimum routes for γ= 60, 40, 20, 10° (from left to right)

Observation of the sequence in Fig. 2, however, disputes intuition, and confirms Eq. (3): ρ tends to the *harmonic* mean of x_1 and x_2. Images for small angles, 5 and 2°, in Fig. 3, show the limiting ρ to be not 2, the arithmetic mean of $x_1 = 1$ and $x_2 = 3$, but 1.5, their harmonic mean. Another interesting property of the optimum routes is that, for varying γ (with fixed x_1, x_2), the locus of the optimum points P is a circle with radius $R = x_1 x_2 / \left(x_1 + x_2\right)$ (same physical units of the x's, of course) centred at (0, R), here R = 3/4. These facts, out of the scope of this study, corroborate the adequacy of the Internet also to reveal noteworthy features.

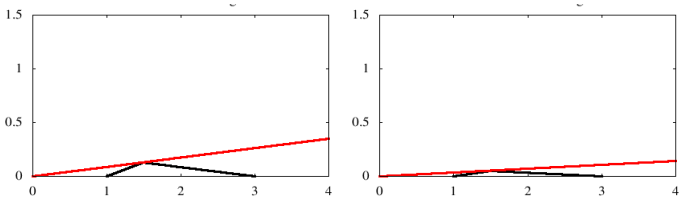


Fig. 3. Optimum routes for γ= 5, 2° (from left to right).

3 Software Architecture

The present study inherits and expands former applications utilized to solve many problems using the Internet, in previous and current academic practice. In this work, we developed a decentralized computing architecture, distributed on a network, using

the HTTP protocol to communicate between the servers, in what is usually known as a web[3] service. The architecture is composed by two types of servers:

a) front end servers, which provide the computing services to the clients, with a simple, practical web interface that can be easily accessed by any browser; and
b) back end servers, which receive the computing tasks from the front end and have the required software to execute them.

The remote call may take a while, depending only on the complexity of the problem and computation power of the server machine.

The back end addresses must be known to the front end servers, so that they can be located on the network. Likewise, the front end must be publicly accessible to the users/clients, and have a well-known address. In the architectural layout described, both the front and back end servers are highly decoupled between them and from the other servers, having no structural dependencies on any single network point [no SPOF[4] (*e.g.,* [4])]. Therefore, they can be easily brought up and down, and change location, without disturbing the overall functioning of the system. The only requirement is just one front end and one back end servers for the system to work.

The decoupling is highly beneficial for two reasons: i) load balancing of requests between the front ends, and of computing tasks between the back ends; and ii) fault tolerance against possible node crashes.

The front and back end support parallel task/requests that require a separation and isolation of execution contexts. This is guaranteed by the HTTP server and script engine used, PHP, with additional safeguards required in the code to carefully avoid any conflict in the resources used (filenames, etc.).

The system is illustrated with the geometric example above, implemented on an Internet link between two semi-closed local networks, the Sigma cluster of IST, and the web servers of FCUL, following the steps described in the next two subsections.

The IST server is deployed on a cluster of amd64 Opteron processors running Debian Linux, Apache 2.2.16, and PHP 5.3.3-7. The FCUL server runs on a cluster of i386 Intel Xeon processors with Red Hat Linux, Apache 2.2.3 and PHP 5.1.6.

3.1 Local Execution

In the starting point, i.e., the previous work done, the system runs in a single local server, which combines the front end and back end functionalities. This is a simple case scenario that served to develop and test the basic computing service. The system uses the following five files in turn:

a) Webpage, such as [2], in a well-known address of a front end server — It is a PHP file containing an HTML 'form' to receive the user's data, which is then sent via an HTTP POST method to a processing PHP script (next);
b) PHP script 'interface.php', which

[3] In *web* (as attributive) or *Web*, the Chicago Manual of Style Online ([4]) was roughly followed.
[4] "single point of failure".

1. Extracts the user's arguments from the HTTP request;
2. Launches the required program in a new process (via PHP 'proc_open') with redirected streams to new process pipes, open to the calling PHP process;
3. Feeds it with the given arguments through the child process read pipe;
4. Waits to read the output of the called program from the other, write pipe; and
5. Closes the pipes and terminates the child process.

c) Binary program ('angDist.exe', compiled from a Fortran 90 source), which also writes to a file the data required for a graphic to be created afterwards.

Now, the 'interface.php' script [in *b)*] constructs a dynamic webpage from:

d) 'interfacetop.php' (constant), the *top* of the webpage;
e) *body* (main) section, in HTML 'pre' format, with the results of the program call, and (typically) a graphic with plotted results, closing HTML *bottom*.

The screenshots are shown in Fig. 4 for the user data and Fig. 5 for the results.

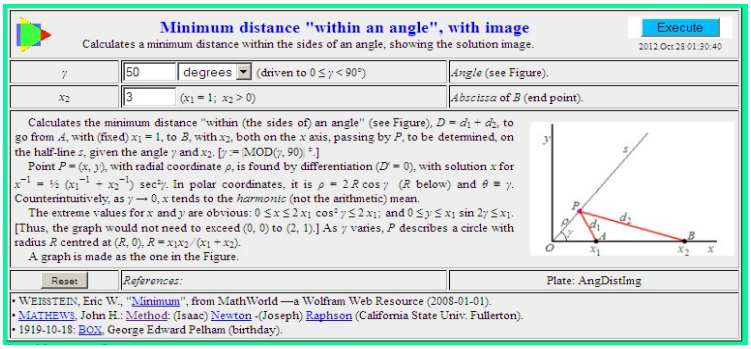


Fig. 4. Webpage for the user data

3.2 Remote Execution

This mode is the focus of the present work, and its website is [3]. Starting from the local execution system described in the previous section, the interface between the computing program and the web front end was greatly modified to support the distribution of both parts, mainly the computing intensive tasks. To this end, the following changes were made:

a) Split the PHP 'interface' between a local, front end part, 'interface-local.php', and a back end web service interface for the remote program, 'interface-remote.php';
b) The front end interface loads the list of known back end servers' addresses, and polls them to find one available;
c) The front end makes an HTTP request to the available server, by invoking the PHP script on the back end. The front end forwards the input data using the HTTP POST method, specifying in the request which service is required (i.e., 'angDist' in the example);

d) The back end interface calls the binary program in a manner similar to the local execution mode;
e) The back end sends the results back to the front end, i.e., both the main results and the parameters of the to-be-created graphic, packaged in the same HTTP response body;
f) The front end process receives the output of the task, and unpacks the two blocks of data (results and graphic's parameters); and
g) The front end retains the responsibility of generating the graphic with the parameters received from the remote request, using the GNU tool *gnuplot*,

The choice was made not to send the graphic itself over the Web, for it could lead to problems of text data encoding (one of the tenets of web services being the use of textual ASCII data), and it would considerably increase the messages' payload size.

The results are, of course, the same as previously. A different HTML background was chosen to differentiate between a service running in local execution mode (the front end at IST) and another in remote mode (the one at FCUL [3]). The remote execution network is schematically shown in Fig. 6.

Results

```
2012-10-28 UTC+0000  1:36:55.952
*** Distance within an angle ***                                  (Oct-2012, MC)

gamma (deg., rad),  50.0      0.8727    |
x_1, x_2,             1.       3.000    |
------------------------------------------+-------------------------------------
COS, TAN(gamma), 0.64279      1.1918    | COS^2,                      0.41318
                                        |
If x_2 = x_1:                           | (absolute minimum, d0)
  x_P, y_P,      0.41318      0.49240    | d0 = x_1 cos^2 = x_P =      0.41318
If x2 = infinity:                       |
  x_P, y_P,      0.82635      0.98481    | x_P = 2 d0
                                        |
MINIMUM DISTANCE for given x_2          |
x_P, y_P,        0.61976     0.73861    | Cartesian coordinates
rho_P,                       0.96418    | (Polar) radial coordinate
d_1, d_2,         0.8307       2.492    | A--P, P--B
Min. distance,                3.3229    | A--P--B

2012-10-28 UTC+0000  1:36:55.952
2012-10-28 UTC+0000  1:36:56.386        CPU:       0.4 sec.          End
```

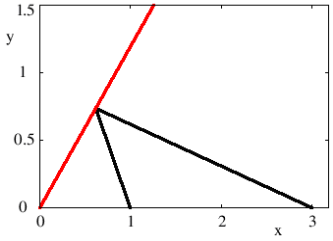


Fig. 5. Webpage for the results

The system performed as expected, namely, the communication latency introduced by the network was negligible when compared to the typical computing time for scientific problems, and it is a constant delay depending only on the size of input and output data, and the underlying network infrastructure.

4 Conclusions

The present study inherits former extensive work in scientific computing over the Internet by one of the authors, close to the work by [7]. Our work has been done in one server of IST, where the webpages and their respective executables are located. The study extrapolates that approach to a two server solution permitting a webpage on a new server, at FCUL, to access an executable placed on the other server, at IST, without the user's perception. The access is governed by two PHP scripts, each placed in one of the servers.

This shows the ease of use of an executable in a remote locus possessing required resources (executables, languages, packages), thus avoiding the breach of the source webpages' style. With the current ease of communication, this points to the use of remote software among collaborating entities, such as companies or universities or in the university-industry linkages. Thus, some software components topologically isolated from a web gateway or from unsecure locations outside its LAN may be accessed by a trusted web server and provided to the worldwide web users.

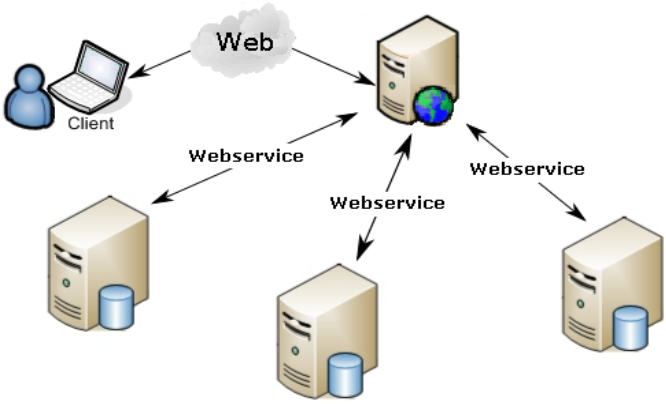


Fig. 6. Remote execution network

Acknowledgements. The research was done at (1.st author) the Department of Computer Science and Engineering and (2.nd author) *Centro de Processos Químicos* (Centre for Chemical Processes), Department of Chemical Engineering, both of the Technical University of Lisbon, Lisbon, Portugal. Thanks are due to the CI *(Centro de Informática*, Informatics Centre) of IST and, for their special effort, CI of FCUL.

References

1. Casquilho, M., Buescu, J.: A minimum distance: arithmetic and harmonic means in a geometric dispute. International J. of Mathematical Education in Science and Technology 147, 399–405 (2011)
2. Casquilho, M., `http://web.ist.utl.pt/~mcasquilho/compute/com/Fx-angdistImg.php`

3. Casquilho, M., http://webpages.fc.ul.pt/~macasquilho/compute/Fx-angdistRemote.php
4. Chicago Manual of Style Online (The), http://www.chicagomanualofstyle.org/
5. Dikaiakos, M.D., Katsaros, D., Mehra, P., Pallis, G., Vakali, A.: Cloud computing: distributed internet computing for IT and scientific research. IEEE Internet Computing 13(5), 10–13 (2009)
6. Dooley, K.: Designing large-scale LAN's. Reilly Media, Inc., Sebastopol (2002)
7. Ponce, V.M.: Vlab, San Diego State University, http://onlinecalc.sdsu.edu/

Achieving Multiple Dispatch in Hybrid Statically and Dynamically Typed Languages

Francisco Ortin, Miguel Garcia, Jose M. Redondo, and Jose Quiroga

University of Oviedo, Computer Science Editorial,
C/ Calvo Sotelo s/n, 33007, Oviedo, Spain
{ortin,garciamiguel,redondojose}@uniovi.es

Abstract. Multiple dispatch is a feature provided by some programming languages, in which the actual method to be executed is dynamically determined by the runtime types of its arguments. Although some programming languages such as CLOS, Xtend and Dylan provide this feature, many widespread languages such as Java, C# and C++ only support *single* dispatch by means of dynamic binding. Therefore, different implementation techniques are commonly used to obtain multiple dispatch from within these single-dispatch languages. We propose a new approach to obtain multiple dispatch in single-dispatch hybrid static and dynamic typing languages. Using C# as an example, the proposed approach is compared with the existing ones, considering factors such as software maintainability and readability, code size, parameter generalization, compile-time type checking, and runtime performance. The example implementation of the hybrid statically and dynamically typed approach in .NET has shown a runtime performance 1.26 times lower than the statically typed approach, and 6.27 times higher than the dynamically type one.

Keywords: Multiple dispatch, multi-method, hybrid static and dynamic typing, dynamic binding, reflection, method overload.

1 Introduction

Most existing object-oriented programming languages provide dynamic binding as a mechanism to implement adaptable and maintainable code. Dynamic binding is a dispatching technique that postpones until runtime the process of associating a message to a specific method. Although dynamic binding is a powerful tool, widespread languages such as Java, C# and C++ only support *single* dispatch. In these languages, multiple-dispatch is obtained by using specific design patterns, inspecting the dynamic type of objects, or using reflection.

With multiple-dispatch, a message can be dynamically associated to a specific method based on the runtime type of all its arguments. These multiple-dispatch methods are also called multi-methods [1]. For example, if we want to evaluate binary expressions of different types with different operators, multi-methods allow modularizing each operand-operator-operand combination in a single method. In the

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 703–713.
DOI: 10.1007/978-3-642-36981-0_65

example C# code in Fig. 1, each `Visit` method implements a different kind of operation for two concrete types, returning the appropriate value type. In this example, the values and operators implement the `Value` and `Operator` interface, respectively. Taking two `Value` operands and an `Operator`, a multi-method would be able to receive these three parameters and dynamically select the appropriate `Visit` method to be called. It works like dynamic binding, but with multiple types. In our example, a triple dispatch mechanism is required.

```
public class EvaluateExpression {
  /******************** Addition ***************************/
  IntegerValue Visit(IntegerValue op1, AddOperator op, IntegerValue op2) {
    return new IntegerValue(op1.Value + op2.Value); }
  DoubleValue Visit(DoubleValue op1, AddOperator op, IntegerValue op2) {
    return new DoubleValue(op1.Value + op2.Value); }
  DoubleValue Visit(IntegerValue op1, AddOperator op, DoubleValue op2) {
    return new DoubleValue(op1.Value + op2.Value); }
  DoubleValue Visit(DoubleValue op1, AddOperator op, DoubleValue op2) {
    return new DoubleValue(op1.Value + op2.Value); }
  StringValue Visit(StringValue op1, AddOperator op, StringValue op2) {
    return new StringValue(op1.Value + op2.Value); }
  StringValue Visit(StringValue op1, AddOperator op, Value op2) {
    return new StringValue(op1.Value + op2.ToString()); }
  StringValue Visit(Value op1, AddOperator op, StringValue op2) {
    return new StringValue(op1.ToString() + op2.Value); }
  /******************** EqualsTo ***************************/
  BoolValue Visit(IntegerValue op1, EqualToOperator op, IntegerValue op2) {
    return new BoolValue(op1.Value == op2.Value); }
  BoolValue Visit(DoubleValue op1, EqualToOperator op, IntegerValue op2) {
    return new BoolValue((int)(op1.Value) == op2.Value); }
  BoolValue Visit(IntegerValue op1, EqualToOperator op, DoubleValue op2) {
    return new BoolValue(op1.Value == ((int)op2.Value)); }
  BoolValue Visit(DoubleValue op1, EqualToOperator op, DoubleValue op2) {
    return new BoolValue(op1.Value == op2.Value); }
  BoolValue Visit(BoolValue op1, EqualToOperator op, BoolValue op2) {
    return new BoolValue(op1.Value == op2.Value); }
  BoolValue Visit(StringValue op1, EqualToOperator op, StringValue op2) {
    return new BoolValue(op1.Value.Equals(op2.Value)); }
  /******************** And ***************************/
  BoolValue Visit(BoolValue op1, AndOperator op, BoolValue op2) {
    return new BoolValue(op1.Value && op2.Value); }
 /******************** The rest of combinations *************/
  Expression Visit(Value op1, Operator op, Value op2) { return null; }
  /* Multi-method implementation (to be done with different approaches) */
}
```

Fig. 1. Modularizing each operand and operator type combination

In this paper, we analyze the common approaches programmers use to simulate multiple dispatching in those object-oriented languages that only provide single dispatch (e.g., Java, C# and C++). To compare the different alternatives, we consider factors such as software maintainability and readability, code size, parameter generalization, compile-time type checking, and runtime performance. The main contribution of this paper is a new approach to obtain multiple dispatch in languages

that provide hybrid dynamic and static typing, such as C#, Objective-C, Boo and Cobra. The proposed solution provides high maintainability and readability, requires reduced code size, allows parameter generalization, and shows a significantly better runtime performance than the reflective approach.

2 Existing Alternatives

2.1 Statically Typed Approach

The *Visitor* design patter is a very common approach to obtain multiple-dispatch in object-oriented languages than do not implement multi-methods [2]. By using method overload, each combination of non-abstract types is implemented in a specific `Visit` method (Fig. 1). Static type checking is used to modularize each operation in a different method. The compiler solves method overload selecting the appropriate implementation depending on the static types of the parameters. Suppose an *n*-dispatch scenario: a method with *n* polymorphic parameters, where each parameter should be dynamically dispatched considering its dynamic type (i.e., multiple dynamic binding). In this *n*-dispatch scenario, the *n* parameters belong to the H_1, H_2… H_n hierarchies, respectively. Under these circumstances, there are potentially $\prod_{i=1}^{n} CC_i$ `Visit` methods, CC_i being the number of concrete (non-abstract) classes in the H_i hierarchy.

Using polymorphism, parameters can be generalized in groups of shared behavior (base classes or interfaces). An example of this generalization is the two last addition methods in Fig. 1. They generalize the way strings are concatenated with any other `Value`. This feature that allows grouping implementations by means of polymorphism is the parameter generalization criterion mentioned in the previous section.

As shown in Fig. 2, the *Visitor* pattern places the `Visit` methods in another class (or hierarchy) to avoid mixing the tree structures to be visited (`Value` and `Operator`) with the traversal algorithms (`Visitor`) [3]. The (single) dispatching mechanism used to select the correct `Visit` method is dynamic binding [2]. A polymorphic (virtual) method must be declared in the tree hierarchy, because that is the hierarchy the specific parameter types of the `Visit` methods belong to. In Fig. 2, the `Accept` method in `Value` provides the multiple dispatch. When overriding this method in a concrete `Value` class, the type of `this` will be non-abstract, and hence the specific type of the first parameter of `Visit` will be known. Therefore, by using dynamic binding, the type of the first parameter is discovered. This process has to be repeated for every parameter of the `Visit` method. In our example (Fig. 2), the type of the second operand is discovered with the `Accept2` method in `Operator`, and `Accept3` in `Value` discovers the type of the third parameter before calling the appropriate `Visit` method.

In this approach, the number of `Accept` methods grows geometrically relative to the dispatch dimensions (i.e., the *n* in n-dispatch, or the number of the `Visit` parameters). Namely, for H_1, H_2… H_n hierarchies of the corresponding *n* parameters in `Visit`, the number of `Accept` methods are $1+\sum_{i=1}^{n-1} \prod_{j=1}^{i} CC_j$. Therefore, the

code size grows geometrically with the number of parameters in the multi-method. Additionally, declaring the signature of each single `Visit` method is error-prone and reduces its readability.

Adding a new concrete class to the tree hierarchy requires adding more `Accept` methods to the implementation (see the formula in the previous paragraph). This feature reduces the maintainability of this approach, causing the so-called *expression problem* [4]. This problem is produced when adding a new type to a type hierarchy involves changes in other classes.

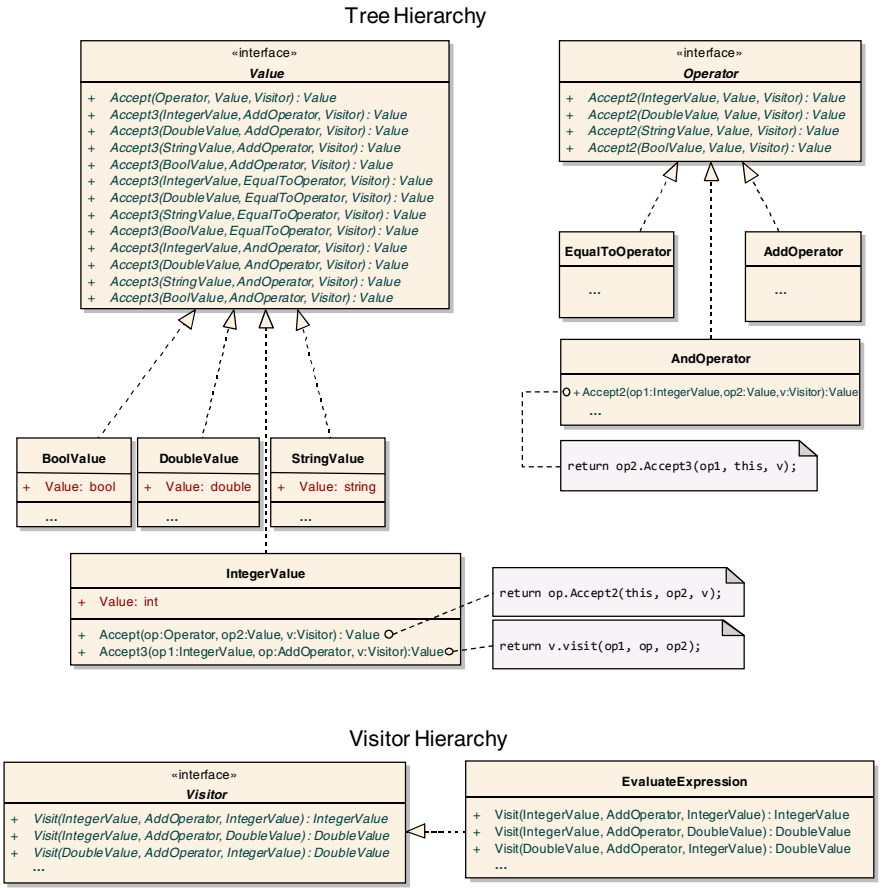


Fig. 2. Multiple dispatch implementation with the statically typed approach (ellipsis obviates repeated members)

This alternative provides different advantages. First, the static type error detection provided by the compiler. Second, this approach provided the best runtime performance (see Section 4). Finally, parameter generalization, as mentioned, is also provided.

2.2 Runtime Type Inspection

In the previous approach, the dispatcher is implemented by reducing multiple-dispatch to multiple cases of single dispatch. Its high dependence on the number of concrete classes makes it error-prone and reduces its maintainability. The following approach implements a dispatcher by consulting the dynamic type of each parameter in order to solve the specific `Visit` method to be called. This type inspection could be performed by either using an *is type of* operator (e.g., `is` in C#) or asking the type of an object at runtime (e.g., `GetType`). Fig. 3 shows its implementation in C#[1]. Notice that this single `Accept` method is part of the `EvaluateExpression` class in Fig. 1 (it is not added to the tree hierarchy).

```
public class EvaluateExpression {
  … // * Visit methods in Fig. 1
  public Value Accept(Value op1, Operator op, Value op2) {
    if (op is AndOperator) {
      if (op1 is BoolValue) {
        if (op2 is BoolValue)
          return Visit((BoolValue)op1, (AndOperator)op, (BoolValue)op2);
        else if (op2 is StringValue)
          return Visit((BoolValue)op1, (AndOperator)op, (StringValue)op2);
        else if (op2 is DoubleValue)
          return Visit((BoolValue)op1, (AndOperator)op, (DoubleValue)op2);
        else if (op2 is IntegerValue)
          return Visit((BoolValue)op1, (AndOperator)op, (IntegerValue)op2);
      }
      else if (op1 is StringValue)  {  …  }
      else if (op1 is DoubleValue)  {  …  }
      else if (op1 is IntegerValue) {  …  }
    else if (op is EqualToOperator) {
      …
    }
    else if (op is AddOperator) {
      …
    }
    Debug.Assert(false, String.Format(
            "No implementation provided for op1={0}, op={1} and op2={2}",
             op1, op, op2) );
    return null;
  }
}
```

Fig. 3. Multiple dispatch implementation using runtime type inspection (for the sake of brevity, ellipsis is used to obviate repeating code)

Fig. 3 shows the low readability of this approach for our triple dispatch example with seven concrete classes. The maintainability of the code is also low, because the dispatcher implementation is highly coupled with the number of both the parameters of the `Visit` method and the concrete classes in the tree hierarchy. At the same time,

[1] Only the *is type of* operator approach is shown. The one that uses the names of the dynamic types is available for download; the URL is specified in Section 5.

the code size of the dispatcher grows with the number of parameters and concrete classes.

This approach makes extensive use of type casts. Since cast statements perform type checks at runtime, this approximation looses the robustness of full compile-time type checking. Runtime performance is not as good as the one in the previous approach (Section 4), and parameter generalization is provided due to polymorphism.

2.3 Reflection

The objective of the reflection approach is to implement a dispatcher that does not depend on the number of concrete classes in the tree hierarchy. For this purpose, not only the types of the parameters but also the methods to be invoked are discovered at runtime. The mechanism used to obtain this objective is reflection, one of the main techniques used in meta-programming [5] (programming applications that manipulate programs, including themselves). Reflection is the capability of a computational system to reason about and act upon itself, adjusting itself to changing conditions [6]. Using reflection, the self-representation of programs can be dynamically consulted and, sometimes, modified. As shown in Fig. 4, the dynamic type of an object can be obtained using reflection (`GetType`). It is also possible to invoke the specific `Visit` method implemented by its dynamic type (`GetMethod`), passing the dynamic types of the parameters. It also provides the runtime invocation of dynamically discovered methods (`Invoke`).

The code size of this approach does not grow with the number of concrete classes. Moreover, the addition of another parameter does involve important changes in the code. Consequently, this approach is more maintainable than the previous ones. Although the reflective `Accept` method in Fig. 4 may be somewhat atypical at first; we think its readability is certainly higher than the one in Fig. 3.

The first drawback of this approach is that no static type checking is performed. If `Accept` invokes a nonexistent `Visit` method, an exception is thrown at runtime, but no compilation error is produced. Another limitation is that parameter

```
public class EvaluateExpression {
  … // * Visit methods in Fig. 1
  public Value Accept(Value op1, Operator op, Value op2) {
    MethodInfo method = this.GetType().GetMethod("Visit",
              BindingFlags.NonPublic | BindingFlags.Instance, null,
              new Type[] { op1.GetType(), op.GetType(), op2.GetType() }, null);
    if (method == null) {
      Debug.Assert(false, String.Format(
                  "No implementation provided for op1={0}, op={1} and op2={2}",
                   op1, op, op2) );
      return null;
    }
  return (Value)method.Invoke(this, new object[] { op1, op, op2 });
  }
}
```

Fig. 4. Multiple dispatch implementation using reflection

generalization is not provided because reflection only looks for one specific `Visit` method. If an implementation with the exact signature specified does not exist, no other polymorphic implementation is searched (e.g., the last `Visit` method in Fig. 1 is never called). Finally, this approach has showed the worst runtime performance in our evaluation (Section 4).

3 The Hybrid Statically and Dynamically Typed Approach

Hybrid static and dynamic typing (henceforth referred to simply as *hybrid typing*) languages provide both typing approaches in the very same programming language. Programmers may use one alternative or the other depending on their interests, following the *static typing where possible, dynamic typing when needed* principle [7]. In the case of multiple dispatch, we have used static typing to modularize the implementation of each operand and operator type combination (`Visit` methods in Fig. 1). We propose the use of dynamic typing to implement multiple dispatchers that dynamically discover the suitable `Visit` method to be invoked.

In a hybrid typing language, its static typing rules are also applied at runtime when dynamic typing is selected. This means that, for instance, method overload is postponed until runtime, but the resolution algorithm stays the same [8]. We have used this feature to implement a multiple dispatcher that discovers the correct `Visit` method to be invoked at runtime, using the overload resolution mechanism provided by the language. At the same time, parameter generalization by means of polymorphism is also achieved.

Fig. 5 shows an example multiple dispatch implementation (`Accept` method) in C#. With `dynamic` the programmer indicates that dynamic typing is preferred, postponing overload resolution. The first maintainability benefit is that the dispatcher does not depend on the number of concrete classes in the tree hierarchy (the *expression problem* [4]). Another dispatching level can be provided by declaring one more parameter, passing it as a new argument to `Visit`. The dispatcher consists in a single invocation to the overloaded `Visit` method, indicating which parameters require dynamic binding (multiple dispatching) with a cast to `dynamic`. If the programmer wants to avoid dynamic binding for a specific parameter, this cast to `dynamic` will not be used. This simplicity makes the code highly readable and reduces its size considerably. At the same time, since the overload resolution mechanism is preserved, parameter generalization by means of polymorphism is also provided (i.e., polymorphic methods like the two last addition implementations for strings in Fig. 1).

In C#, static type checking is disabled when the `dynamic` type is used, lacking the compile-time detection of type errors. However, there are research works on hybrid typing languages, such as the *StaDyn* programming language [9], that provide static type checking when the `dynamic` type is used. When this feature is not supported, the best approach is to use static types to declare the `Accept` parameters using polymorphism (restricting their types to `Value` and `Operator`, as shown in Fig. 5). At the same time, exception handling is another mechanism that can be used to make

the code more robust –notice that parameter generalization reduces the number of possible exceptions to be thrown, compared to the reflection approach.

```
public class EvaluateExpression {
  … // * Visit methods in Fig. 1
  public Value Accept(Value op1, Operator op, Value op2) {
    try {
      return this.Visit((dynamic)op1, (dynamic)op, (dynamic)op2);
    } catch (RuntimeBinderException) {
      Debug.Assert(false, String.Format(
                  "No implementation provided for op1={0}, op={1} and op2={2}",
                  op1, op, op2));
    }
    return null;
  }
}
```

Fig. 5. Multiple dispatch implementation with the hybrid typing approach

Finally, the proposed approach has showed a runtime performance between the statically typed implementation and the reflective one (see Section 4). Hybrid typing languages, including C#, commonly implement a dynamic cache to improve runtime performance of dynamically typed code [10]. This technique provides a significant runtime performance improvement compared to reflection [11].

4 Runtime Performance Evaluation

In order to compare the runtime performance of the proposed approaches, we have developed in C# an interpreter of the simple expression language defined in Fig. 6. Expressions are integer, double or string literals, plus the logical-and, addition and equals binary operators. Two multi-methods were used to implement the interpreter: one single dispatch to traverse the abstract syntax tree (AST); and another triple-dispatch that evaluates every node in the AST (the code shown in Figs. 1, 2, 3, 4 and 5). The program evaluated was `(1+2) == (2+1) && (1.1+2.2) == (2.2+1.1) && ("hello"+"world") == ("world"+"CIST")` in a loop of 10,000 iterations.

Expression	→	*Expression* (**+** | **==** | **&&**) *Expression*
	|	*IntegerLiteral*
	|	*DoubleLiteral*
	|	*StringLiteral*

Fig. 6. Syntax of the expression language used in the evaluation

Regarding the data analysis, we followed the methodology proposed in [12] to evaluate the runtime performance of JIT-compiled programming languages. The elapsed time of each test is measured 30 times, taking the mean value for a 95% confidence level (using the *Student's t*-distribution) to eliminate measurement errors [13]. All the tests were carried out on a lightly loaded 2.67 GHz Intel Core I7 920

system with 8 GB of RAM running an updated 64-bit version of Windows 7 Professional SP1.

Fig. 7 shows the mean execution time plus the confidence interval for a 95% confidence level. As mentioned, the statically typed approach is the fastest one because, apart from dynamic binding, no other dynamic type operation is performed at runtime. The type inspection approach is the second one but, unlike the other approaches, its runtime performance increases with the number of concrete classes [14]. As shown in [14], this dispatch implementation may even show the worst runtime performance for a specific number of concrete classes. Finally, hybrid typing is 1.26 times slower than the statically typed approach, but 6.27 times faster than reflection. This difference with reflection is caused by the call site cache implemented by the .NET Framework [15].

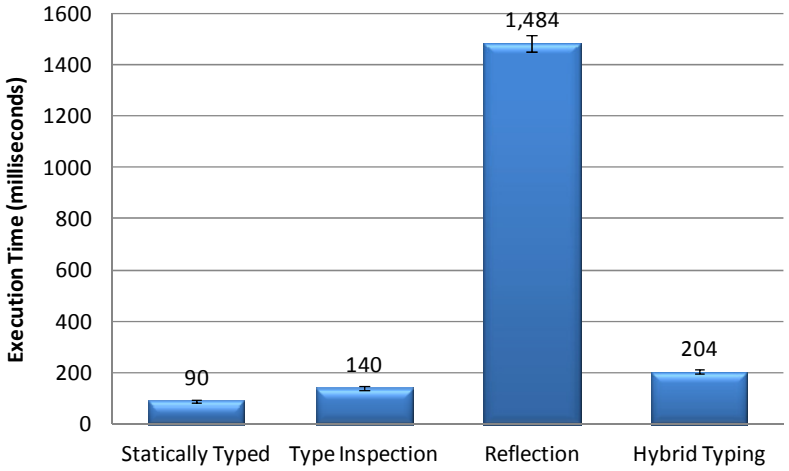


Fig. 7. Execution time (expressed in milliseconds) of the different approaches

5 Conclusions and Future Work

The combination of hybrid dynamic and static typing supported by some programming languages can be used to obtain several benefits in the implementation of multiple dispatch, with a lower runtime performance cost than reflection. The proposed approach is highly readable and maintainable, being loosely coupled with the number of concrete classes in the tree hierarchy and the dispatching dimensions. It also provides parameter generalization, obtained with the use of polymorphism for each parameter type. Compile-time type error detection is not provided, but the number of runtime errors is lower than the reflective approach. Finally, a runtime performance evaluation using C# has revealed that the proposed solution in the .NET platform is 1.26 times slower than the *Visitor* design pattern and 6.27 times faster than reflection.

Future work will be focused on improving compile-time type error detection and runtime performance of the proposed approach. We have developed an extension of C# that performs type inference over dynamic references [9]. By using union and intersection types [16], many type errors of dynamically typed code are detected by the compiler and runtime performance is significantly improved [17]. We are currently implementing the overload resolution algorithm when dynamic types are passed as parameters, a key feature used in this paper.

All the implementations presented in this paper, including the expression language used in the evaluation, are freely available at http://www.reflection.uniovi.es/stadyn/download/2012/worldCIST.zip

Acknowledgments. This work has been partially funded by Microsoft Research to develop the project entitled *Extending Dynamic Features of the SSCLI, awarded in the Phoenix and SSCLI, Compilation and Managed Execution Request for Proposals*. This work has also been funded by the Department of Science and Innovation (Spain) under the National Program for Research, Development and Innovation: project TIN2011-25978, entitled *Obtaining Adaptable, Robust and Efficient Software by Including Structural Reflection in Statically Typed Programming Languages*.

References

1. Chambers, C.: Object-Oriented Multi-Methods in Cecil. In: ECOOP 1992. LNCS, vol. 615, pp. 33–56. Springer, Heidelberg (1992)
2. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley Longman, Boston (1995)
3. Ortin, F., Zapico, D., Cueva, J.M.: Design Patterns for Teaching Type Checking in a Compiler Construction Course. IEEE Transactions on Education 50(3), 273–283 (2007)
4. Torgersen, M.: The Expression Problem Revisited. In: Odersky, M. (ed.) ECOOP 2004. LNCS, vol. 3086, pp. 123–146. Springer, Heidelberg (2004)
5. Ortin, F., Lopez, B., Perez-Schofield, J.B.G.: Separating Adaptable Persistence Attributes through Computational Reflection. IEEE Software 21(6), 41–49 (2004)
6. Maes, P.: Computational Reflection. Ph.D. Thesis, Laboratory for Artificial Intelligence, Vrije Universiteit, Amsterdam, the Netherlands (1987)
7. Meijer, E., Drayton, P.: Static Typing Where Possible, Dynamic Typing When Needed. In: OOPSLA Workshop on Revival of Dynamic Languages, pp. 1–6 (2004)
8. Bierman, G., Meijer, E., Torgersen, M.: Adding Dynamic Types to C#. In: D'Hondt, T. (ed.) ECOOP 2010. LNCS, vol. 6183, pp. 76–100. Springer, Heidelberg (2010)
9. Ortin, F., Zapico, D., Perez-Schofield, J.B.G., Garcia, M.: Including both Static and Dynamic Typing in the same Programming Language. IET Software 4(4), 268–282 (2010)
10. Ortin, F., Redondo, J.M., Perez-Schofield, J.B.G.: Efficient virtual machine support of runtime structural reflection. Science of Computer Programming 74(10), 836–860 (2009)
11. Redondo, J.M., Ortin, F., Cueva, J.M.: Optimizing Reflective Primitives of Dynamic Languages. International Journal of Software Engineering and Knowledge Engineering 18(6), 759–783 (2008)
12. Georges, A., Buytaert, D., Eeckhout, L.: Statistically rigorous Java performance evaluation. In: Object-Oriented Programming Systems and Applications, pp. 57–76. ACM, New York (2007)

13. Redondo, J.M., Ortin, F.: Efficient Support of Dynamic Inheritance for Class and Prototype-based Languages. Journal of Software and Systems, 1–24 (2012) (to be published)
14. Cunei, A., Vitek, J.: PolyD: a flexible dispatching framework. In: Object-Oriented Programming Systems and Applications, pp. 487–503. ACM, New York (2005)
15. Chiles, B., Turner, A.: Dynamic Language Runtime, `http://www.codeplex.com/Download?ProjectName=dlr&DownloadId=127512`
16. Ortin, F., Garcia, M.: Union and intersection types to support both dynamic and static typing. Information Processing Letters 111(6), 278–286 (2011)
17. Ortin, F.: Type Inference to Optimize a Hybrid Statically and Dynamically Typed Language. The Computer Journal 54(11), 1901–1924 (2011)

Open Source Technologies Involved in Constructing a Web-Based Football Information System

Pedro Rodrigues[1], António Belguinha[1], Carlos Gomes[1], Pedro Cardoso[1], Tiago Vilas[1], Renato Mestre[2], and J.M.F. Rodrigues[1,3]

[1] Institute of Engineering, University of the Algarve, 8005-139 Faro, Portugal
[2] Inesting, S.A., 8005-146 Faro, Portugal
[3] Vision Laboratory, LARSyS, University of the Algarve, 8005-139 Faro, Portugal
{pcardoso,jrodrig}@ualg.pt

Abstract. The current information systems and match analysis software associated to professional football output a huge amount of statistics. Many football professionals are particularly interested in real time information about the tactical plan occurring during the match, and the relations between that information and what was prepared in the training sessions. It is fundamental to have on the bench, and on-the-fly, the most relevant information each time they have to take a decision. In this paper, we present a set of open source technologies involved in building a multi-platform web based integrated football information system, supported in three main modules: user interfaces, databases, and the tactical plan detection and classification. We show that the selected technologies are suitable for those modules, allowing field occurrences to trigger meaningful information.

Keywords: Football, open source, web technologies, information system.

1 Introduction

All historically renowned football teams have displayed a recognizable footprint in their gameplay, which has always been thought of as something observed by football experts rather than described by game statistics and scientific principles. Within the coaching process, great emphasis is placed on the coach's ability to observe and recall all the critical discrete incidents from a sport and tactical performance [1]. However, it has been shown that coaches cannot accurately observe and recall all of the detailed information that is required for a complete understanding or interpretation of the team performance [2].

Footdata is a project to build a new multi-platform product for football, based on technologies of information and communication systems, which integrates two fundamental components of the football world: i) A social network, which will have all the typical features and ii) the professional component, which features an acquisition and information system to meet all the football management needs. The latter component includes an automated platform to gather information from the teams, not only on the competition part but also on the preparation model (training

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 715–723.
DOI: 10.1007/978-3-642-36981-0_66

sessions). This platform will be based on a prototype that will process images acquired live on-site (in matches or training sessions), or images from recorded or broadcasted matches. The processing of this information goes beyond the traditional statistical data compilations extracted from automated observation of the teams actions, adding features which will allow the analysis of the football match structure, namely in its two fundamental ways: width and depth of the team, allowing to rationalize and optimize the team's actions regarding occupation of spaces. However, the main goal is to automatically collect information which concerns the tactical plan, and on-the-fly alert the technical team for specific events. All the above should be presented in a web (browser) environment, accessible from personal computer or a mobile device (smartphone or tablet).

There are in the market commercial systems that integrate partially some of the components of this project, e.g. Kizanaro® [3], Tracab® [4], Prozone® [5], Amisco® [6], and SportVU® [7]. We can also find in the literature examples that explore some aspects of the project, like spatial and spatiotemporal analysis [8], local positioning systems in the game [9], and analysis of football strategies [10], [11] and [12].

In this paper we present a preview of what the system will be, and focus on the open source technologies to develop the professional component of Footdata(-PRO). Due to the limited size of the paper, we will not enter into full details about the reasons that led us select each specific technology. The main contribution is a compilation and discussion of the open source technologies necessary to develop a web-based football information system (IS).

The rest of the paper is organized as follows. Section 2 shows the main Footdata modules, Section 3 and 4 presents the user interface and the database technologies and in Section 5 the tactical plan detection and classification. Finally, we conclude the paper with the discussion and future work in Section 6.

2 Footdata-PRO Modules

Has mention in the Introduction, Footdata is a project to build a social network and the professional component (-PRO) in a web environment. Footdata-PRO features an acquisition and information system divided in five main modules: (a) Field, where the ball and players location in the pitch is done by computer vision (during the match or training sessions). This information is sent to a server through an internet connection, where the match tactical plans are verified, based on the information available from prepared tactical schemes, trainings, individual players, and collective information from each team. Based in the above procedure, in the coach's (user) profile and in what is occurring in real-time in the field (or in practice), selected specific information is sent to the coach's mobile device. A request can be sent to the server every time the coach requires extra information. (b) Center, where all the information and the main computation is congregated (including the processing and management of previously recorded videos). (c) Coach, where the coach's team inputs and accesses to detailed information of every topic related to the football match, scouting, medical department, etc. (d) Player, used to send specific information to the player, and the (e) Presentation, where semi-automatic presentations are prepared for the players in preparation moments.

An integrated web interface, where usability is a key factor, will combine all the above features in a single place, including video manipulation tools, tactical planning/drawing tools, semi-automatic presentation tools, etc. Figure 1 illustrates the Footdata-PRO main structure.

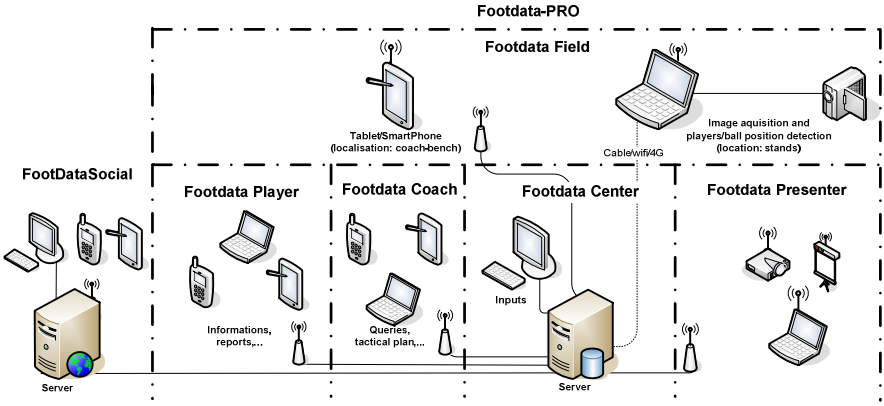


Fig. 1. Footdata-PRO main modules: Field, where the ball and players location in the pitch is done by computer vision; Center, where all the information and the main computation is congregated; Coach, where the coach inputs and accesses to detailed information; Player, used to send specific information to the players; and Presentation, where semi-automatic presentations are prepared for the players

3 User Interface Technologies

A prerequisite for the user interface was that it should be delivered as a web (browser) application. Furthermore, it should be a multi-platform system running in multiple operating systems (OS) with special emphasis to Windows, Linux, Mac, iOS, and Android. This requirement can be seen as an asset, as it is deliverable to a larger number of users, requiring only a web browser on a personal computer or on some other mobile device (phone or tablet). After the analysis of the possibilities, the decision of using HTML5 [13], [14] to its largest extension was made, maintaining the necessary fallbacks to web browser plugins or specific applications for the mobile devices. This decision was made based on two main assumptions: the majority of the HTML5 functionalities are, or soon will be, supported by the most used browsers (OS independent) and the will/need to use open source software to develop the applications (the use of open source software was a previous agreement done by all the project consortium members). Nevertheless, several difficulties are easy to find, as the embedding of video (see below).

The second phase, required the selection of a programming platform, in particular a web development framework. The choice was Django [15]. As their providers define it, Django is a high-level Python [16] web framework that encourages rapid development and clean, pragmatic design [15], which provides among others:

an object-relational mapper, elegant URL design, template system, cache system, and internationalization [17]. The option of using Django was also supported by the fact that the programming language behind it (Python) provides a very rich set functionalities on its standard library, including internet data handling, interprocess communication and networking, and concurrent execution. Furthermore, Python has also a large set of third-party extensions, in the form of modules, which extends the Python's standard library, as for example: connectors to the majority of the database systems (e.g., MySQL, PostgreSQL, Oracle, Microsoft SQL Server, MongoDB, CouchDB, etc. [18]), data analysis (e.g., Pandas, NumPy, SciPy [18]), and image processing and manipulation (e.g., OpenCV [19]).

One of the specifications made for the interface is to allow video manipulation, including the selection and displacement of a player (see Section 2). The solution planned to the video section, without using any third party plugins, will use HTML5 video support. However the current incompabilities between browsers/platforms and the video formats stand as an obstacle quite costly to solve. We have to keep in mind that legacy browsers/platforms may not support the HTML5 video element [20], this problem can addressed with the inclusion of fallbacks into plugins. By analysing the browsers supported video formats, we found out that a good format combination to work with would be MP4/WebM or MP4/Ogg. With one of those solutions, displaying compatibility is attained by the largest majority of the web browsers. Despite the format of the video being correct for a certain browser, we have to be careful with the video and audio codecs. In this case we would have the H.264 video with AAC audio codecs in a MP4 container (.mp4), VP8 video codec with Vorbis audio codec in a WebM container (.webm), and finally, Theora video codec with Vorbis audio codec in a Ogg container (.ogv) [21]. To discover if a particular browser supports a particular video format, libraries like the Modernizr Javascript Library [22] can be used. HTML5media, MediaElement.js and Video.js are also solutions to take into account in this matter.

To provide an efficient video manipulation, the native video controls that come in each browser seem to be insufficient. Fortunately, using JavaScript we can implement our own controls and design (with the help of CSS - Cascading Style Sheets) [20]. Different approaches can be found in [23], [24] where the HTML5 Video and Canvas are combined together. jQuery User Interface - jQueryUI [25] is also worth mentioning because besides being a tool for the controls design it can also solve some incompatibilities (e.g., at the time of writing, the HTML5 input range for the video seekbar, is not correctly rendered by Internet Explorer nor Firefox).

Figure 2 shows two examples of the interfaces: on the left, the team details are presented, and on the right, the video tool (some information is omitted due to the confidentiality agreement).

4 Database Technologies

The database behind this project has several levels of information to store. From the classic user profile, to the tracking data (players/ball) from matches and practice

sessions. This leads to very different types of data. For example, to prepare a training session each coach has its own methodology, which goes from a detailed level, where every exercise is thought, to the ones prepared in a macro structure. Also different strategies are used to prepare a match.

The conjunction of several facts provided by the football consultants, gave us a set of uncertainties which made impossible to deliver a complete database structure at an earlier stage. As matter of fact, the process has been mainly iterative, where new ideas are constantly being proposed, which somehow led us to the adoption of an agile development strategy, attaining the principles that underlie the Agile Manifesto.

Those uncertainties made us decided to use relational databases management systems (RDBMS) [26] and NoSQL [27] databases. Comparing both we can state that NoSQL have simple access patterns, compromise on consistency for performance and Ad-hoc data format. On the other hand RDBMS allow complex queries with joins, ACID (Atomicity, Consistency, Isolation, and Durability) transactions and have well defined schemas. In particular, the project uses MySQL [28] and MongoDB [29] databases, allowing a very flexible structure where "well-structured/known" data (e.g., users data) is left on the MySQL server and data which can vary or be altered (e.g., definition of the game model or presentations from the Footdata-PRO module) is left on MongoDB.

Fig. 2. Two examples of web interfaces done in HTLM5. On the left some team information, and on the right the video tool (some information is omitted due to the confidentiality agreement).

5 Tactical Plan Detection and Classification

We can describe a football match, as the confrontation between two teams, guided by their own behaviours and rules. These rules and behaviours are called offensive and defensive principles, methods of attack and defence, and processes that help the players actions in the field, forming the team. All these concepts are part of the tactical organization, that can be grouped together forming what is known as the tactical plan ("game model"). Several systems have been developed for the above

purpose and are being used by some of the well-known football clubs in Europe, e.g. [30]. Since a football match is composed with 23 moving objects (22 players and a ball), it is an application area of spatial information science, where spatial and spatiotemporal approaches can be applied to tactical analysis. However, few researches of spatial and spatiotemporal analysis have been done for sports tactical analysis [8]. Nevertheless, an approach to the analysis of time-based event records and real-time behaviour records on sport performance known as T-pattern detection is presented in [13]. Latter, a framework for analysis of soccer tactics with trajectory data is presented in [8], and in [10] the authors used tools from network theory to describe the strategy of football teams. For other approaches see [11] and [12].

For the player detection, and localization in the field we use an open source computer vision library, the OpenCV [19], [31], it has C++, Python and Java interfaces running on multiple OS. In our case C++ and the CodeBlocks IDE [32] were used. To speed up the code, the OpenMP API [33] was also used, which supports multi-platform shared-memory parallel programming.

In terms of computer vision, we prepare the system to process videos and realtime acquiring. To detect and localize the player, the algorithm can be summarized as: (a) Definition of the game zone (pitch delimitation), using the HSV colour space [34] combined with the lateral lines detection with the Canny edge detector and Hough transform [35]. We must refer that this step is only necessary on broadcasted videos. When the images are acquired by a static camera that captures the entire pitch, the key points for the contours of the field are mark by hand at the beginning of a match, using the implemented interface.

The next step (b) is to detect and assign the players to each team, for which we use a combination of median and morphological filters [35] over the binary field image, to obtain players blobs. Then, to assign them to each team, the equipment range of the Hue was used. The final step (c) consists in tracking each player by calculating the shortest Euclidean distance between each player in the present and in the previous frame. Occlusions, overlapping and players "collisions" are also solved.

In order to be able to analyse the tactical plan, it is required to obtain the correct coordinates of the players with respect to the football field. Which is attained with a perspective transformation (see e.g. [19]), from the original image (video frame) to the "model" field. In the Footdata, despite computing many of the traditional statistics, we are more interested in detecting in real-time (or in video) where the different pre-defined (by each coach; saved in the IS) tactical plan situations is accomplished or fails, labelling the different situations, and sending the information to the coach. We must refer that most game models are built from several relations: players position in the field, their position relatively to the teammates and adversaries, and the position of the ball. This is still work in progress.

Figure 3 shows a frame from a broadcasted video sequence (left), the players detection and classification, and the projection to the model field (right), where the distances are in meters, and the origin of coordinates is the interception point of the top lateral line with the midfield line.

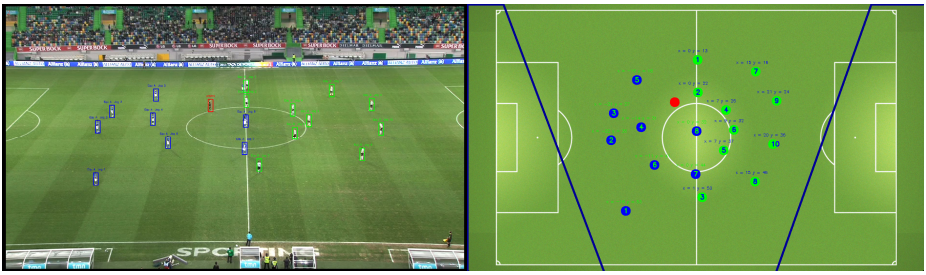


Fig. 3. The players classification (left), and the projection to the model field (right)

6 Conclusions

In this paper we present and discuss some of the main open source technologies which were involved in the building of a web integrated information system for football. The selected technologies present the most promising functionalities and compatibilities to allow us the implementation of the pretended web/cloud IS. Despite not profoundly discussed due to the lack of space, other technologies have been tested.

Other subject also not discussed in this paper is the usability of the interface. This is a fundamental key feature for the success of the product, since the final users (technical staff and players) have a very particular language (professional way to see the football), which is very different from the usual fans. Due to this reason, the interface has been discussed step-by-step with the designers, several football coaches and elements of the technical staffs. Different outputs have been (and still are) in discussion, depending on the video source to be analysed, i.e., from broadcasted videos or from a static (system) camera in the stands.

In the context of the used open source technologies, it is also worth mentioning the management and documentation of the project is done with ChiliProject [36], a web based project management system; the Git [37] is used for the distributed revision control and source code management system; and Doxygen which is used for the code documentation [38].

As future work, we plan to continue the development of the Footdata-PRO user interface. In the tactical plan detection and classification, we have already implemented the most trivial statistics, as the distance travelled by a player, the detection of passes, etc. We are now focused in the automatic detection of the offensive and defensive principles, methods of attack and defence. Also, we are still improving the players and ball tracking methods, particularly in the case of occlusions, focusing always on the fact that the system has to work in real time and has to be portable (easy to mount in any pitch or practice field by an assistant coach), even if for those reasons we have to lose a little bit of reliability.

Acknowledgments. This work was partly supported by the Portuguese Foundation for Science and Technology (FCT), project PEst-OE/EEI/LA0009/2011 and project FootData QREN I&DT, n.º 23119. We also thanks to project leader Inesting, S.A. [www.inesting.com], and the consultant football coach Domingos Paciência.

References

1. Armatas, V., Yiannakos, A., Sileloglou, P.: Relationship between time and goal scoring in soccer games: Analysis of three World Cups. Int. J. of Performance Analysis in Sport 7(2), 48–58 (2007)
2. Franks, I., Miller, G.: Eyewitness Testimony in Sport. J. of Sport Behaviour 9, 39–45 (1986)
3. Kizanaro, http://www.kizanaro.com (accessed November 2012)
4. Tracab, http://www.tracab.com (accessed November 2012)
5. Prozone, http://www.prozonesports.com (accessed November 2012)
6. Amisco, http://www.sport-universal.com (accessed November 2012)
7. Sportvu, http://www.sportvu.com (accessed November 2012)
8. Kim, H., Kwon, O., Li, K.: Spatial and Spatiotemporal Analysis of Soccer. In: Proc. 19th ACM SIGSPATIAL GIS 2011, Chicago, USA, November 1-4, pp. 385–388 (2011)
9. Leser, R., Baca, A., Ogris, G.: Local positioning systems in (game) sports. Sensors 11(10), 9778–9797 (2011)
10. Peña, J.L., Touchette, H.: A Network Theory Analysis of Football Strategies, pp. 1–6 (June 28, 2012), arXiv:1206.6904v1[math.CO]
11. Duarte, R., Araújo, D., Folgado, H., Esteves, P., Marques, P., Davids, K.: Capturing complex, non-linear team behaviours during competitive football performance. J. Syst. Sci. Complex, 1–11 (2012)
12. Moura, F., Martins, L., Anido, R., Barros, R., Cunha, S.: Quantitative analysis of Brazilian football players' organisation on the pitch. Sports Biomech. 11(1), 85–96 (2012)
13. Pilgrim, M.: HTML5 - Up and Running. O'Reilly Media, Inc. (2010)
14. HTML5, http://www.w3.org/TR/2011/WD-html5-20110525/ (accessed November, 2012)
15. Django, http://www.djangoproject.com (accessed November 2012)
16. Python, http://www.python.org/ (accessed November 2012)
17. Kaplan-Moss, J., Holovaty, A.: The Definitive Guide to Django: Web Development done Right (Expert's Voice in Web Development), 2nd edn. Apress (2009)
18. Oliphant, T.E.: Python for Scientific Computing. Computing in Science & Engineering 9(3), 10–20 (2007)
19. Bradski, G., Kaehler, A.: Learning OpenCV: Computer Vision in C++ with the OpenCV Library. O'Reilly (2012)
20. Powers, S.: HTML5 Media. O'Reilly (2011)
21. Pfeiffer, S.: The Definitive Guide to HTML5 Video. Apress (2010)
22. Casario, M., Elst, P., Brown, C., Wormser, N., Hanquez, C.: HTML5 Solutions - Essential Techniques for HTML5 Developers. Appress (2011)
23. Meyer, J.: HTML5 and JavaScript Projects. Apress (2011)
24. Schmitt, C., Simpson, K.: HTML5 Cookbook. O'Reilly (2012)
25. jQuery:jqueryui.com (accessed November 2012)
26. Balasubramaniam, R.: Data Security in Relational Database Management System. Int. J. of Computer Science and Security 6(4), 203–210 (2012)
27. Cattell, R.: Scalable SQL and NoSQL Data Stores. ACM SIGMOD Rec. 39(4), 12–27 (2011)
28. MySQL, http://www.mysql.com (accessed November 2012)
29. mongoDB, http://www.mongodb.org (accessed November 2012)

30. Borrie, A., Jonsson, G.K., Magnusson, M.S.: Temporal Pattern Analysis and its Applicability in Sport: An Explanation and Exemplar Data. J. of Sports Science 20(10), 845–852 (2002)
31. OpenCV, http://opencv.org (accessed November 2012)
32. Code: Blocks, http://www.codeblocks.org/ (accessed November 2012)
33. OpenMP, http://openmp.org/wp/ (accessed November, 2012)
34. Dearden, A., Demiris, Y., Grau, O.: Tracking Football Player Movement from a Single Moving Camera Using Particle Filters. In: Proc. 3rd Euro. Conf. on Visual Media Production CVMP 2006, London, UK, pp. 29–37 (2006)
35. Rao, D., Patil, S., Babu, N., Muthukumar, V.: Implementation and Evaluation of Image Processing Algorithms on Reconfigurable Architecture using C-based Hardware Descriptive Languages. Int. J. of Theoretical and Applied Computer Sciences 1(1), 9–34 (2006)
36. ChiliProject, http://www.chiliproject.org (accessed November 2012)
37. Git, http://git-scm.com (accessed November 2012)
38. Doxygen, http://www.stack.nl/~dimitri/doxygen/ (accessed November 2012)

A Collaborative Tourist System Using Serious Games

Rui Pedro Araújo Fernandes[1], João Emilio Almeida[1,2], and Rosaldo J. F. Rosseti[1,2]

[1] FEUP – Faculdade de Engenharia da Universidade do Porto
[2] LIACC – Laboratório de Inteligência Artificial e Ciência de Computadores
Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal
{ei06123,joao.emilio.almeida,rossetti}@fe.up.pt

Abstract. Smartphones are trending and most tourists already own one. Many applications exist for smartphones, in order to help tourists in their journeys. Applications available share the same problem, though: content may not be complete or updated and, they lack user feedback. The best information about a city is usually held by locals, or by recent visitors. An application fed and maintained by its users, seems a good solution to improve the information contained in mobile applications for tourists. Motivating travelers to share their knowledge, in a long-term basis, poses a problem. We propose to tackle that problem through the use of Serious Games, to achieve user's collaboration in the building and maintenance of a tourism database. A prototype for Android platform was developed and deployed. A group of 20 individuals tested the application in the terrain. Data was collected from their experience. A survey to the users showed their feelings towards this new experience. Results showed that Serious Games are a viable approach to achieve several objectives. Users from the test group showed to be interested in the gaming environment, enabling them to share their knowledge, possibly in a long term basis.

Keywords: Serious Games, Tourism, Android.

1 Introduction

Tourists have many tasks to be performed prior to the travelling itself. They need to research what they would like to see, whether using books, online resources or friends. After this research period, they are still required to create a route that allows them to visit the places they are interested in, taking into account the time that they have available. All these tasks are very time consuming.

Many of today's tourists have a smartphone and the tendency is that most of them will have one in a near future. With this in mind, there are several applications in the market to support tourism activities. A whole range of these applications exist, but they typically share the same problems, considering that they are created and maintained by third-party entities. Those problems are the following: i) Contents are incomplete or outdated; ii) An application focuses in a specific area; iii) Contributions from the users themselves are reduced to classifications and comments.

A solution to these problems could be having the users be the sources and the maintainers of the information available in an application of this kind. But users

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 725–734.
DOI: 10.1007/978-3-642-36981-0_67

would need to be motivated in contributing with their knowledge and, to reach a good quality standard in the information, in spending their time confirming the data of other users. This motivation would have to be kept, as well, for a large period of time, so that users could feel part of a community and reach long term collaboration.

The approach followed in this project, was the merging of serious games with a tourism application, thus creating a serious game for tourists. It's intended that, through games, tourists will be eager to contribute to the application in order to progress further in the game, while they should feel joy in using the application.

2 Background and Related Work

2.1 Serious Games

Although that, at a glance, Serious Games (SG) appear to be a recent phenomenon, the term as we know today was first presented long before the computer's and electronic devices' introduction to entertainment. In 1970, Clarck C. Abt [1] introduced SG with a very close definition to the one that what we use today.

To understand why SG should be used, it is important to first realize how games engage players. Prensky [2], proposes that six structural elements characterize games and makes them motivating: i) Rules; ii) Goals and objectives; iii) Outcome and feedback; iv) Conflict, competition, challenge and opposition, leading to player's excitement; v) interaction, the social aspect in the game; vi) representation or story exaggerating interesting aspects of reality. If players are having fun, they will stay motivated and will strive to achieve harder and longer objectives [3].

2.2 Gathering PoI

A Point of Interest, or PoI is is a specific point location that someone may find useful or interesting. The term is widely used in GPS navigation software. PoI are needed to establish the route that ultimately the tourist will use in their visits.

In order to populate a database of valid PoI, mainly, three sources of information can be used: public databases, social networks and the users of the application.

Public databases are available to the average user and provide some way to freely download the information contained in them. Besides the easiness to obtain information, these databases are also valuable because they are usually built by the users. Since the users posted the contents, it's expected that these contents may have interest to other users looking for data in the same context. The most common examples are: OpenStreetMap [Ope11] and Google places.

Social Networks are being used by millions of users all over the world [4]. Facebook has an enormous amount of data with possible interest to many parties. Useful Facebook data can be extracted, from their new application: Facebook Places. This application allows users to create places that will be stored as Facebook profiles. To access these profiles, it is necessary to use the Facebook Graph API to access every object in Facebook's social graph by a series of HTTP request, which return results in JSON. From a Place profile, it is possible to retrieve, if available, the

following information: likes; category; description; city; country; GPS coordinates; check-ins; website; location (street, number); photos. This information is, assuredly, useful for a database of PoIs. The main problem resides in finding the profiles which are Places, since Facebook doesn't provide any way of searching only for them. A way to overcome this obstacle is crawling all the profiles available, identifying those which are Places. But, the 64-bit range ID adopted by Facebook poses a hard to overcome temporal barrier, since it would take years for a single computer to scroll through all the IDs. Therefore, to capture useful data, would be necessary either, too much time, or a large number of machines. Neither of those was available during the development of this application, fact that leads to let the Facebook option aside.

2.3 User's Contribution to Gathering PoI

The PoI databases described previously are mainly user fed. Users of those applications take the initiative of adding their own PoI, making these databases increasingly complete, at the same time that they comment and rate others additions. So, it's acceptable to think that the users of the application described in this paper will want to participate by adding their own PoI in addition to those already existent, - collected elsewhere - and they should also be interested in rating and commenting the PoI already available. Therefore, users might be classified as another source of information, as viable as the information collected automatically by the application's agents.

2.4 Routes Generation

Route generation consists in generating a route, from a given set of PoI. Considering that we are generating routes for tourists, each PoI should have a score to the tourist to whom the route is being generated for. The goal is to find the optimum route, i.e. the route which satisfies the tourist the most (the set of PoI in which it passes sums up the highest score possible). Although several algorithms exist [5-8], in this work the Branch&Bound algorithm was used, due to its simplicity.

3 Project Description

3.1 Serious Games in Smartphones Using Android

To evaluate if Serious Games are a good solution, a test-bed was created, which consisted in an Android application, to achieve long-term collaboration from tourists in the creation and maintenance of a tourism database.

The android application rewards users for exploring a new place, as well as for contributing with their knowledge about their new discoveries or about their hometown. The application communicates with a remote server, through an internet connection, in order to obtain data. All the communications are done through Web Services, so that the server's data can be used by a multitude of platforms.

The purpose is to create the largest database of PoIs possible, being the users/players the providers of the information. With this in mind, it is important to identify what challenges the players will look for in a game of this genre. Players should be pushed to share their knowledge about the cities that they know and, at the same time, get rewards for sharing that knowledge. In order to create the desired competitive gaming environment, a point system was created, so that users would be rewarded for their actions in the game. This happens so that users are encouraged to contribute with their knowledge.

3.2 Rewards and the Point System

Players are rewarded for actions on the game. These rewards build the player's score. Since a game should engage its players both in short and in long-term, these two kinds of rewards were implemented. Short-term rewards aim to keep the player motivated. They consist on a badge system, to classify users, depending on the completion of games, and a certain number of actions. User starts with a "Beginner" badge, then moves forward to "Good", "Pro" and "TOP tourist". There will be a list with the "Contributor of the day" and "Contributor of the month".

Long-term rewards are related with the amount of points earned and a players ranking. So, users can check where they stand in the overall picture, and strive to get better, by comparing themselves with others.

People are known to be extra motivated when some action is rewarded with money or equivalent - a prize - which has value for them. It was not implemented, but if this application gets launched in the market, some physical rewards, could be used to incentive use. For example, brands could use the application's georeferencing properties to advertise their products or shops, and users could get bonuses for going to the place at a determined time or for being the best contributor in a city in the day where a determined product is being advertised.

3.3 System's Architecture and Implementation

The system's architecture is client-server based. A high-level representation of the architecture is presented on Figure 1. In this representation is visible that the system is composed by three main parts. The "Server" is composed by three elements: the "Crawler", the "System Database" and the "Web Services". The "Crawler" is responsible for making use of the open databases' APIs in order to request, parse and subsequently add PoIs to the "System Database". The "System Database" contains all the information regarding the project, from users to PoIs. "Web Services" were built in order to allow the communication between the "System Database" and the "Clients". "Open Databases" are all the databases from where the system collects PoI for, in a first phase, the beginning of the project, so that users have something to start with and, in a second phase, for a synchronization in which the "Crawler" should check which new PoI from these databases can be added to the system. "Clients" are all the smartphones using the application, being the system's interface with its users.

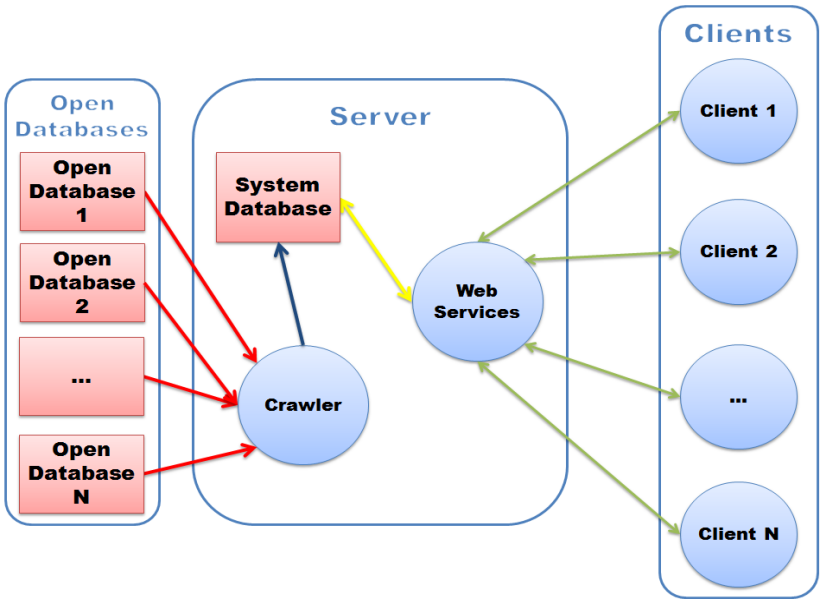


Fig. 1. High level system architecture diagram

The client's side was developed for the Android platform, used by the majority of smartphones. All the Web Services and API calls done in this project are implemented in JSON. RESTful Web Services were implemented, instead of SOAP, due to its lightweight feature. For the database, the relational open-source MySQL was chosen.

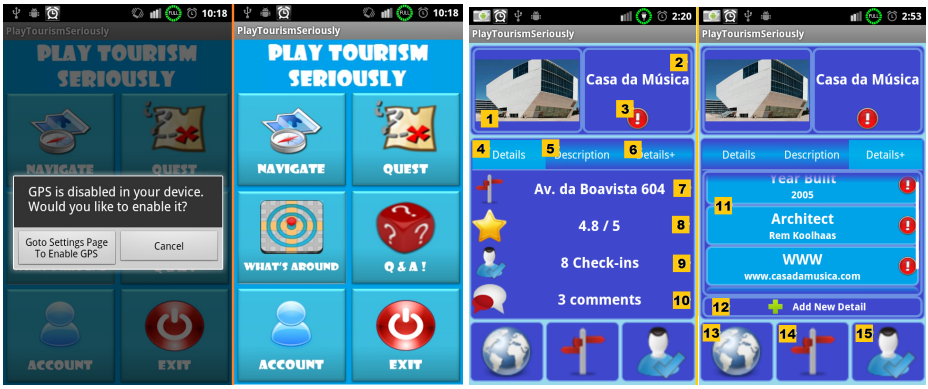
4 Experimental Setup

After developing the application, a test group was selected and the application was tested, by the users belonging to the test group, in order to obtain real results from its use. All the data gathered from the user's interaction was stored. When the test-period was over, the testers were submitted to a survey, in order to evaluate the application's effects on the users, as well as the player's opinion.

4.1 Description of the Application

Figure 2 a) shows the main menu when the GPS signal is not available or is turned off. Figure 2 b) shows the main menu when GPS is active. When showing a PoI, application screen looks like figure 2 c) and d). The areas and their components, are explained below.

Face Area: 1 –This area contains the PoI picture. With a tap, users can check all the pictures available about the place, or add their own. 2 - Title - PoI name. 3 - Report PoI - By tapping in this button, one can report the content available in the database.



(a) (b) (c) (d)

Fig. 2. A few screenshots: a) main menu when the GPS is turned off; b) when GPS is active; c) a PoI when the "Details" tab is selected; d) a PoI when the "Details+" tab is selected

Details Area: 4 – Details. 5 – Description. 6 - Dynamic Details. 7 – Address. 8 – Classification. Shows the average classification of the PoI. 9 - Check-ins. Shows how many people have checked in into this PoI; 10 - Comments - Shows how many comments exist for the PoI. By tapping in this area one can read the comments available; 11 - Dynamic details. 12 - Add a dynamic detail.

Actions Area: 13 - Taping this button shows the localization of the PoI in a map. 14 - Taping this button a user can obtain the directions to the PoI. 15 - By tapping this button, if is near the PoI, the user checks-in and is able to comment and rate it.

The data inserted by users and, specially, when a reward is given for it, should be tracked and controlled. It's intended that the users turn into the controllers of the content added by others. These controlling actions can be done by tapping in the exclamation mark icons that can be seen in figure 2 c) and d).

Users are awarded points when information that they reported is considered to be faulty by the system. All the users that reported the same get points, and the users that suggested the same alteration, in the case of the dynamic details, get the bonus for adding a dynamic detail as well.

4.2 The Game Concept

The application can be used in various modes: "Navigate", "Circuit", "Quest", "What's Around" and "Questions / Answers".

"Navigate" is a functionality that allows the users to make a trip around a city, taking into account their current location and available time. This trip is calculated according to the user's defined interests, and it is possible to customize the trip.

In the game "Circuit" users are challenged to complete the circuit that "Navigate" generates. The user gets points for each PoI that he visits (when a check-in is made) and bonus points if he finishes the full circuit. The user is able to set the trip's

maximum time and a finishing point. The routes try to maximize the players' pleasure, by including the most possible PoIs that fall in his categories of interest.

All the distances, and corresponding times, between those points are calculated using The Google Distance Matrix API [9] and the values are stored in the server's Database. With all the time necessary to travel between each point, a Branch&Bound algorithm is applied, being the nodes the score of the PoI, and the branches the time necessary to travel from the parent PoI to the son PoI. Taking into account that the server's computational power is higher than a mobile device's one, and in order to minimize data transfers, the route generation is done server-side. The device asks, via Web Services, for a route, sending the initial variables (current position, time available, interests and end point), and receives the directions of the calculated route.

"Quest" is a game that intends to boost the inner explorer inside of any tourist. In this game, players are challenged to explore a certain area, in order to find a PoI that can be anywhere inside the region with the radius defined by the user. The area "1" (fig.2c) shows the name of the PoI which the user is looking for. When tapped, information about the place, like pictures and other details, is shown. Buttons "3" and "5" let the user zoom in and out the map. Button "4" enables the user to start a new quest. Button "6" lets the user play the "Questions/Answers" game, talked further, while he is playing the "Quest" game. Button "2" enables/disables the help for this game. This helping system works like the "hot and cold" game [10], providing immediate feedback when the user changes position. The closer the user gets to the point he is trying to find, the closer the bar moves to the fire symbol, and the opposite is also verifiable. Upon successful conclusion of the "Quest game", points are awarded. In the case of forfeit, compensation points are awarded as well.

"What's around" allows users to search for nearby PoIs, which meet the criteria defined by them. Users can search for name and/or category and the places are presented by order of distance to the current location. After selecting the search parameters, a request is sent to the server, returning all the PoIs that meet the criteria, in a 15km radius. When the results are presented, users can either view the information about a place, or ask a route to it.

"Questions/Answers" is a game that intends to test the players' general culture. It poses questions to the players about the PoIs of the city in which they encounter themselves. Correct answers will give user reward points.

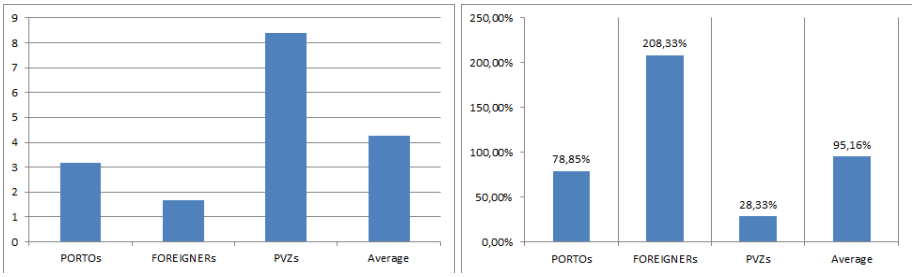
4.3 Evaluation Methodology

In order to obtain data related to the use of the application, a group of 20 individuals were invited to test it in a real scenario. All these testers were previous android users and, therefore, no adaptation period to the hardware was required. Users installed the application in their Android mobile phones, and used it for the period of a week, without any supervision. Utilization data was collected by the server, in order to obtain quantitative data about the application's usage and, when the testing period was over, a survey was conducted on the users, in order to obtain qualitative data. This survey focused in their opinion about the application, but it also covered the effects from it that might have yielded on the users.

5 Results and Discussion

The test group used the application in the city where they live in, for a period of 5 days (a working week). Users knew that a ranking would be created. The servers of the application were loaded with information's about two cities: Porto and Póvoa de Varzim. Users were subdivided in three groups: PORTOs are the users from that city, inhabitants. PVZs the inhabitants of Póvoa do Varzim and FOREIGNERS, a group of foreigners that tested the application in Porto (fig.3).

Interesting to refer that, on top of the 353 PoI that were contained in the database in the beginning of the test week, 85 more were added by users. This represents an increase of almost 25%, showing that the competition created by the point system motivates users to contribute with their own knowledge. Not surprisingly, in average, FOREIGNERs contributed with less PoIs than the other groups, since they were visiting a new city. Still, they added some that were not on the database, as they were passing by them. The fact that PVZs added more PoIs than PORTOs is probably due to the fact that most of the PoIs contained in the beginning of the testing period were in the city of Porto. With less PoIs it was easier to add information (fig.3 a).



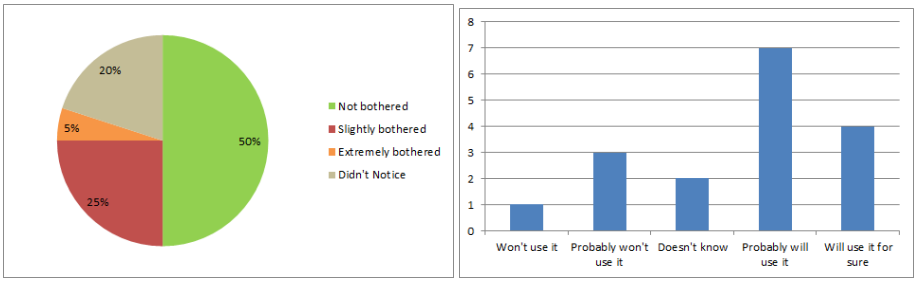
(a) (b)

Fig. 3. a) average PoIs added by group; b) "Questions/Answers" game results by group

To refer, as well, that users produced a total of 181 check-ins in 88 different PoIs. These are used to validate the existence of a point, besides allowing users interactions with the PoI. A group of 20 users, in just a period of a week of non-continuous usage, confirmed the existence of almost a quarter of the total PoIs available.

A survey was conducted through e-mail, sent after the testing period was over, having all the answers been received in the three days following the dispatch of it.

Subtle advertisement was used in the application, in order to assess whether users would recognize it and to study if they would feel bothered by it. Only 20% of the players missed the advertisement. To those who noticed, most of them did indeed remember the companies advertised. This shows that, messages in the application can be used to pass marketing information. Figure 4 a) shows data relative to how bothered were users by the advertisement used. Half of the players said that they were not bothered at all by this advertisement and, even, that they were expecting some, since most of the android's free applications contain advertisement. Just one user



a) (b)

Fig. 4. Survey answers: a) bothered by advertising; b) "Will use this application in the future?"

showed to be extremely bothered by the advertisement, having said that any form of advertisement is extremely annoying to him. The others were just slightly bothered, as they prefer applications without any form of marketing.

All the users found the application interesting to use. Some stated that the games made them perform actions that they wouldn't normally do, being the most used example the searching for a random point located in the area (the "Quest" game).

Most said that they felt fun in the games and that they would feel even more if they could do, somehow, some of the tasks in a group environment. Still, they felt motivated by the ranking system, and that they were doing tasks, sometimes, only to go up in the rankings chart. They wanted to finish the week as the winners.

Some of the users were happy to realize, as well, that they learned some general culture from playing the "Questions/Answers" game. They didn't realize that they had learned until they were asked to answer the same questions again. As they performed considerably better, they felt achievement. The game "Circuit" was basically used only by the FOREIGNERs, since the other users' level of interest in a tour around the city in which they live in, was low.

When asked about their likelihood to keep using the application after the testing period, the results were considerably positive. As shown in fig. 4 b), 55% of the testers answered positively to this question, having 20% of them stated that they would use it for sure.

6 Conclusions and Future Work

Overall, it was proven that serious games are a good approach in achieving the objective of creating collaboration in the generation of a tourism database. While having fun, and by having the ranking system creating competition, users were open to contribute with their knowledge, so as to achieve a better place in the rankings. At the same time, they said that they would be further encouraged to do it, if the rewards would be something more substantial than only social recognition. In addition, users, in general, confirmed that they would like to continue to use the application in their travels, since it fulfilled the objective of aiding them in the exploration of a new place,

and in teaching them facts about their whereabouts, at the same time that provided a fun approach in doing so.

Summarizing, the test group utilized in this project contributed with their knowledge, and would like to continue using the application, showing that serious games are an effective way of engaging users in a collaborative environment, possibly in a long-term basis.

Future work includes widening the number of users, and increasing the testing period, to test and evaluate the rewarding system in the long-term. Expand use and PoI to more cities and places, is also another issue to explore.

Future development includes: Multilanguage support, the "Treasure Hunt Game" an extension to the "Quest" game; improve the interface and buttons design; improved route generation algorithm, to indicate how much time the user has available and permit the games to continue per multiple days; instead of asking users' their preferences, infer them from social networks and by the users' behavior in PoI selections and ratings when using the game.

References

1. Abt, C.: Serious Games. Viking Press (1970)
2. Prensky, M.: The digital game-based learning revolution. Digital Game-based Learning (2001)
3. Kirriemuir, J., McFarlane, A.: Literature review in games and learning: a report for NESTA Futurelab. NESTA Futurelab Series Report 8, 5 (2004)
4. Boyd, D., Ellison, N.: Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication 13, 210–230 (2007)
5. Kramer, R., Modsching, M., ten Hagen, K., Gretzel, U.: Behavioural impacts of mobile tour guides. In: Information and Communication Technologies in Tourism, pp. 109–118 (2007)
6. Maruyama, A., Shibata, N., Murata, Y., Yasumoto, K., Ito, M.: A personal tourism navigation system to support traveling multiple destinations with time restrictions. Advanced Information Networking and Applications 2, 18–21 (2004)
7. Zenker, B., Ludwig, B.: Rose assisting pedestrians to find preferred events and comfortable public transport connections. In: ACM Mobility Conference (2009)
8. Garcia, A., Arbelaitz, O., Linaza, M.T., Vansteenwegen, P., Souffriau, W.: Personalized Tourist Route Generation. In: Daniel, F., Facca, F.M. (eds.) ICWE 2010. LNCS, vol. 6385, pp. 486–497. Springer, Heidelberg (2010)
9. Google Developers. Google distance matrix api, `https://developers.google.com/maps/documentation/distancematrix/`
10. Boy Scout Trail. Hot and cold game, `http://www.boyscouttrail.com/content/game/game-417.asp`

SOA – Based Authentication System for Dynamic Handwritten Signature

Andreea Salinca, Sorin Mircea Rusu, and Ana-Maria Pricochi

Research & Development, SOFTWIN, Bucharest, Romania
{asalinca,srusu,apricochi}@softwin.ro

Abstract. The increased frequency and damage caused by virtual attacks, reveals a need for stronger, easily deployable authentication solutions. This paper describes the architecture of a biometric authentication system based on dynamic handwritten signatures. Our design uses the principles of service-oriented architecture (SOA) and consists of non-intrusive, easy-to-integrate services and software components. Its features are enhanced security, reduced development time for the applications that use it, and modularity with respect to the underlying biometric system. Our experiments demonstrate that the system is a cost-efficient solution capable of handling thousands of authentication requests per minute.

Keywords: Authentication System, Service Oriented Architecture, Biometry, Dynamic Signature, Information Security.

1 Introduction

The spectacular growth of the Internet-based economy naturally gives rise to a host of security concerns. Impersonation attacks against on-line services occur with increased frequency and ingenuity. According to the Internet Crime Report Center Complaint Center (IC3) [1], in 2011 the most common reported crimes were of FBI-scams identity theft and advance fee fraud, with total losses of trillions of dollars. During the last year alone, there were over 300 000 reported identity theft cases [1] and the scale of the problem is only bound to increase. In Europe, FPEG (The EU Fraud Prevention Group), an expert fraud prevention group established by the European Commission, under an action plan (2004-2007) dealt with identity theft and payment.

Biometric technologies represent a promising defense against identity theft that have the potential to be adopted for day-to-day activities including for example, shopping or voting. Incorporating such authentication methods within practical applications poses significant challenges that range from ease of implementation for the developers to providing transparent, non-intrusive user experience.

In this paper we present a system that aims to meet these requirements. We rely on handwritten signatures as the underlying biometric authentication method. Our system is suitable for on-line authentication services, or more generally, for authentication to Internet-connected services.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 735–744.
DOI: 10.1007/978-3-642-36981-0_68

One of the major goals of the developers of security systems is the inclusion of biometric authentication technologies in web applications. This is very important because most of the transactions that carry essential information are done using a computer connected to the Internet.

An increasing number of providers of on-line security solutions are expected in the following years and therefore standardization is required in implementations. At the moment, OASIS (Organization for the Advancement of Structured Information Standards) Biometric Identity Assurance Services (BIAS) works on such a standard.

National Institute for Science and Technology strengthens the inclusion of web technologies in the biometric authentication process within the project Biometric Web Services (BWS) which aims to define a set of recommendations in the development of biometric devices that will exchange data through web services [2].

The following is the outline of the paper: in the introduction of this paper we give the motivation of constructing biometric authentication systems in protecting sensible asset given the high number of virtual attacks.

The second section presents the background on biometric verification systems based on Service-oriented architecture (SOA) principles. The next section presents the usage scenario of the proposed dynamic handwritten SOA-based authentication system which can be used by any type of application that needs an enhanced security.

Further we described the architecture of the proposed dynamic handwritten authentication system that establishes the user's identity in real time. Section 5 addresses implementation issues and experimental results on the service-based authentication system. Also we demonstrate that our system complies with the service-level-agreement (SLA) qualities.

In the last section we given conclusion remarks about the novel system which is a large-scale handwritten authentication system aiming to respect the qualities defined for SOA services.

2 Related Work

SOA (Service-oriented architecture) design principles are: standardized service contracts, service loose coupling, service abstraction, service reusability, service discoverability and composability, service autonomy, service statelessness and service-orientation, interoperability [3]. Along with the goals of service-oriented architecture these form a set of best practices and policies important in designing large scale biometric systems.

Previous work on architecture of systems based on SOA principles describe authentication systems which include various types of biometric technologies such as fingerprint, vein, iris, face, voice and signature.

Paper [4] presents a biometric authentication system containing multiple logical units: an enrolment terminal which extracts the features of the captured biometric data for the user enrolment and stores them into an authentication server, an authentication terminal that converts the captured biometric data for authentication and sends the information to an authentication server that matches the content with the stored

template. The system can use any of the following types of psychical biometrics: fingerprint, vein, iris or behavioral biometrics: voice.

The research of Yu et al. describes a web-based authentication system and method containing multiple server stations receiving the biometric data from multiple web client stations and an authorization server which makes the proper comparison by using a web cloud as a communication environment [5].

Another biometric authentication system that allows multiple clients using different biometric technologies such as iris, face or fingerprint contains a plurality of biometric machines specialized in a particular verification type of biometric data [6].

A large-scale biometric system based on a service - oriented architecture which handles multiple authentication transactions from users focuses on main problems like security, accuracy and efficiency. The authors argue that a large-scale biometric system handles 100 000 transactions/day and tends to have a large number of diverse clients. The performance of the face recognition system is evaluated using equal-error rate and proves that the method using neural networks is robust [7].

In [8] the authors present a remote handwritten signature recognition system comprising multiple clients and one verification server that uses cryptographic methods such as encryption and digital signing of the signature in order to secure the communication between components.

A variety of commercial biometric service oriented applications are in the market, such as BioSP (Biometric Services Platform) which is a configurable and programmable platform that uses data from multiple biometric technologies (fingerprint, face, iris, hand) in a distributed network for analyzing and processing the verification.

While most of the SOA-based biometric systems focus on using various biometric technologies, the system we propose is a large-scale system capable of handling millions of handwritten signature authentication requests a day. Moreover, the system complies with the qualities defined in the service-level agreement such as accuracy, availability, capacity, scalability, modifiability, interoperability and security.

3 Usage Scenario of Authentication System

Any type of application for which security plays a crucial role can use the proposed authentication systems. Some market segments that fit with the system are*: enterprise, consumer, government*.

Enterprise applications include e-Business transactions, business-to-business (B2B) transactions, e-Commerce, enterprise authentication or remote access. Applications regarding consumers include e-Banking or e-Commerce transactions. Also, for the government segment, application may involve electronic votes or biometric passports.

Using the SOA-based authentication system, each authentication requires an acquisition device that captures the dynamic handwritten signature. The acquisition device is an intelligent electronic pen which captures the graphic form of the signature by using an embedded optical navigation sensor. Also, by using two MEMS (micro-electro-mechanical-systems) acceleration sensors it captures a person's hand

movements. The system that analyzes the bio-kinetic information captured from each signature using the acquisition device is presented in the patent [9].

The SOA-based dynamic handwritten signature authentication system presented in this paper can be used by any type of application which demands the verification of the user's identity. The usage scenario of the system is the following:

- An owner of an application that requires authentication uses the provided SDK (software development kit) and implements his custom authentication module;
- The owner of the application registers his application in the system by using a special administration website of the system;
- All users from registered applications receive the biometric acquisition device that capture their individual signatures;
- Users from application enrolled in the system register their signatures in the system using the acquisition device. They introduce a number of signatures (for instance 5) considered specimens, upon which will be made the authentication. The acquisition process does not require many computational resources and it is made on the client machines;
- The client application uses a dedicated web-service of the authentication system to send the signature packed and encrypted over the Internet to the entry-point of the authentication system;
- The entry-point of the authentication system validates the request received by verifying the certificate information and sends the authentication requests to a load balancing service which chooses an authentication web-server from a network of servers capable of handling large number of requests;
- The evaluation of signatures - proper verification - is made on one of the authentication servers and the result will be sent back to the client.

4 Architecture of the SOA-Based Authentication System

Service-oriented Architecture (SOA) paradigm is widely adopted for the development of software systems through the dynamic composition of network accessible loosely-coupled services.

The proposed SOA-based authentication system for dynamic handwritten signature recognition architecture is web-services based and comprehends multiple logical components figured in Fig. 1. In the following, we present the authentication system's components, the relationship between them and the flow of the signature verification process.

Proxies are web-services components located on the clients that implement their own security modules using the provided SDK. The proxy encrypts and signs the requests containing biometric handwritten signatures and sends them to the reverse-proxy.

A web-service representing a *Reverse-proxy* enhances the security of the entry-point of the authentication system by allowing only authorized requests that are received from *Proxies.* It also assures data validation for the received request and caches some information such as the application's identifier, in order to speed up the time response of the request.

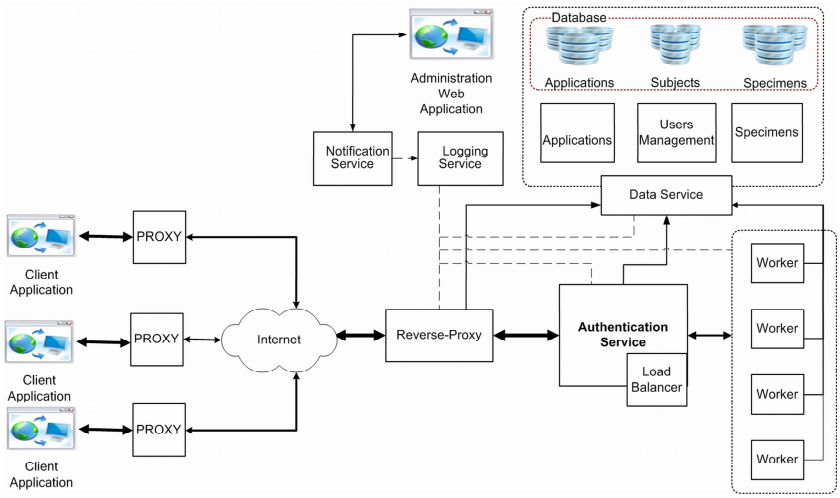


Fig. 1. SOA-based authentication system's architecture

An *Authentication web-service* comprising a load balancing module receives authentication requests from the *Reverse-proxy* and sends the requests to a *Worker* by the means of a multi-class dynamic algorithm running in batch mode. The task assignment uses a classifier taking into consideration worker node's information and task class proportionality. The load balancing algorithm is patent pending.

Data service is a web-service managing databases storing information about the system's application and users and the associated handwritten signatures for the registered users that will be used in a verification request.

Another component is *Logging service* which collects data about events occurred in the authentication system and triggers event's specific actions.

The *Notification service* sends e-mails to administrators and clients when error occurs. The Notification service can also dispatch periodic activity reports. Users can configure the notification and reports to be sent.

A plurality of *Workers* – represented by a plurality of web-services – each one containing multiple *authentication processes* for running the verification request of the handwritten dynamic signature in parallel on multi core and many core processors. Also, the authentication process can solve registering requests in the authentication system of the specimens (a number of signatures used to enroll in the system).

The *Authentication processes* use for verification of dynamic handwritten signature the multimodal recognition methods described in the system [9]. The authors define the term of invariants which represent the biometric information extracted from the signature signals captured using the acquisition device. A set of eight signature recognition algorithms (SRA) use invariants series to compute similarity distances between the captured signature and the user's specimens, such as Levenshtein distance [9]. To establish authenticity of the handwritten signature, the distances are combined using different decision methods like computing a statistical threshold per user or using artificial neural networks [10].

An *Administration web-application* offers management facilities to register, to update and to delete a client application along with his associated users. It allows data visualization of the client's signatures, activity reports generation in the authentication system and other user information.

5 Results

The SOA-based handwritten authentication system for handwritten signature recognition is a large-scale authentication system which respects both measurable and immeasurable defined service-level-agreement (SLA) qualities regarding accuracy, availability, quality requirements, scalability, modifiability, interoperability and security [11]. Below, we present and discuss the results corresponding of each of these quality measures.

We implement the components of the proposed authentication system as web-services using Windows Communication Foundation (WCF). Each component has a service interface based on the Web Service Description Language (WSDL) and uses XML (Extensible Markup Language) and SOAP (Simple Object Access Protocol) to transport the biometric request.

Security. Security is an important quality in the service-level-agreement qualities defined for a SOA-based system, providing confidentiality, integrity, non-repudiation and assurance [11]. In order to enhance the authentication system's security we used the Transport Layer Security (TLS) protocol. The package containing the handwritten signature is encrypted and digital signed and has associated a time stamping by using the SOAP extension: WS-Security which uses XML Signature and XML Encryption specifications. The encryption method used is DSA (Digital Signature Algorithm).

Availability. The handwritten authentication system has a high availability rate of 99%, being operational 24 hours a day, 7 days a week due to the specific mechanism which quickly threats the failures occurred. Logging *service* implements the mechanism of detecting errors and uses the *Notification services* to notify the administrators for handling failures.

Accuracy. The accuracy of the handwritten authentication system is another quality defined in the service level agreement. The accuracy of the authentication system may be defined as the accuracy of the signature recognition method using the performance coefficients: FRR (false reject rate) – computed using all genuine signatures for each person after enrollment in the system and FAR (false accept rate) – computed using all captured forgeries for a person. The performance coefficients FRR of 14.59% and FAR of 1.24% obtained using signatures of a database collected from more than 100 persons [12] and the authentication methods for recognizing the authenticity of a person described in [9], reach the accuracy of a real-world application.

Capacity. A great challenge in the performance of a large-biometric system is to maintain the system's efficiency by handling a large number of requests in real time using limited computational resources. SLA must mention the capacity of the SOA authentication system which is represented by number of concurrent verification tasks

in a given period of time. Therefore we evaluate the authentication system's capacity on multiple hardware configurations presented in Table 1. We construct SOAP messages containing signature verification requests for the authentication system and we use Apache JMeter to manipulate simultaneous requests in order to measure the capacity performances. An overview of the experiments is presented in Table 2. The results are computed on a desktop computer using multi core processors (type A) or many core processors (type B and type C). Table 2 measures the average amount of time it takes to process each signature verification request in milliseconds and provided a rate of signature verification requests in requests per minute. The results prove the efficiency of the authentication system which is able to process more than 650 signature verification requests per minute, with an average response time under 1 second, only by using one *Worker* machine of any type mentioned below.

Table 1. Hardware configuration types of Workers

Processing unit type	Frequency	Type
2xGPU (240 CUDA cores / GPU)	633 MHz	A
CPU (4 cores x 4 threads)	2.13 GHz	B
3xGPU (480 CUDA cores / GPU)	732 MHz	C

Table 2. Capacity of the handwritten authentication system using various configuration types of workers. The results of the last line correspond to using type A and B worker machines.

Worker Configuration type	Number of signature verification requests / minute	Average response time (ms)
A	700	981
B	650	953
C	4000	495
A and B	1200	1053

The SOA-based authentication system we propose is a large scale system able to handle more than 40000 verification requests per hour using only one low-cost hardware Worker. Therefore, millions of verification requests can be handled per day and the performance is comparable to a real-world project's requirements. EU (European Union) Biometric Matching System is a project supporting Visa application from members of 27 EU states and uses Daon's large scale multimodal biometric authentication solution based on SOA principles, being able of handling over 100000 fingerprint verifications per day [13].

Modifiability. If it is necessary to handle a higher number of requests coming from clients concerted in different geographical regions a possible option is to replicate the entire handwritten authentication system in multiple locations using a database mirroring solution. The client applications use the *proxies* to send authentication requests to another component *Localizer* which is designed to forward the requests to the appropriate *reverse-proxy* of an authentication system. The forward decision is

based on the location and the load characteristics of the authentication system. The modifiability of the handwritten authentication system is possible because the architecture design complies with the principles of a SOA system.

Interoperability. The architecture of the dynamic handwritten authentication system presented in this paper is modifiable and, by creating new *Authentication processes,* it can use multiple biometric technologies. Hence, the interoperability between various biometric technologies is assured by the usage of standard ISO / IEC 19784-1 in the implementation of the system's components. The standard defines a framework which manages multiple BSPs (Biometric Service Providers) of different manufacturers [14]. We implement such a framework and the BSP for dynamic handwritten signature recognition.

Scalability. To benchmark the scalability of the dynamic handwritten authentication system we implement a MPI-based application that simulates its behavior. Since the architecture of the system is service-oriented, we use different processes for simulating the *Reverse Proxy*, the *Authentication web-service* and each *Worker* service nodes. The simulation includes an implementation version of the load balancing algorithm. Meanwhile, we establish the additional processing times (e.g.: data access time, communication time between services) by real measurements of the authentication system's deployment. We design the performance and behavior of the *Workers* and the *Authentication processes* using the benchmark of various machines, in terms of computational time for processing each signature verification request (Table 2).

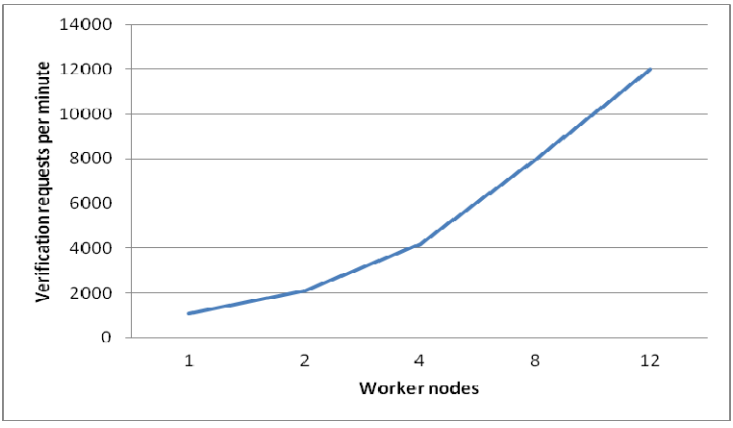


Fig. 2. SOA-based handwritten authentication system's scalability

We test the system on a 2.13 GHz (4 cores X 4 threads) and 4.0 GB RAM machine. To prove the scalability of the handwritten signature verification system, we gradually add several *Worker* nodes in the system and measure the number of authentication requests per minute while keeping an average response time for each verification request of 1 second. Strong scaling for the authentication requests shows good scalability up to 12 Worker nodes; the bottleneck is due to the hardware limitation of the test machine. Fig. 2 shows the results of the experiment above.

6 Conclusions

This paper presents a large-scalable architecture of a biometric authentication system based on SOA principles, providing a high level of modularity. The system uses handwritten signature verification to establish the identity of a person using specialized web-services (*Workers*). It provides administration and user management of applications using the authentication system in order to protect sensible information. It also has a configurable component of reporting and logging for maintaining a high level of availability.

In the previous section we prove that the proposed system respects the attributes defined by the service-level-agreement which ensure the quality of a system over the Internet. In this case is considered security, accuracy, capacity, interoperability, modifiability, scalability issues simultaneously, unlike a common biometric system.

The handwritten authentication system is a large-scale biometric system capable of handling millions of verification request in a day. It is a challenge to maintain a high level of system's efficiency considering the trade-off between the high number of authentication requests and the response time, which must be as short as possible.

Experimental results obtained show that dynamic handwritten authentication system achieves this goal, being able to solve 4000 verification requests per minute with an average response time of 495 ms using one many core Worker machine. Also, the proposed service-based system is scalable; the number of verification requests solved per minute is significantly increasing, proportionally to the capacity of each added Worker machine in the system.

Although we focus on describing the architecture of a biometric system using handwritten signature, the architecture is generic and modifiable to multiple biometric technologies allowing the implementation of dedicated *Authentication processes* for any type of biometrics according to the interoperability principle descried in the previous section.

Acknowledgments. The work from this paper is a result of the research activities from the project "Automated system of authenticating through biometric signature/ATHOS", co-funded from SOP IEC PA2: Research, Technological Development and Innovation for Competitiveness, Operation 2.1.2: "Complex research projects fostering the participation of high-level international experts".

References

1. Internet Crime Complaint Center, Bureau of Justice Assistance – U.S. Bureau of Justice. 2011 Internet Crime Report (2011), `http://www.ic3.gov/media/annualreport/2011_IC3Report.pdf` (accessed October 25, 2012)
2. Michaels, R.J., Aronoff, M., Kwong, K., Mangold, K.C., Marshall, K.: Specification for WS-Biometric Devices (WS-BD), pp. 1–123, NIST SP - 500-288 (2012)
3. Erl, T.: Service-Oriented Architecture: Concepts, Technology, and Design. Prentice Hall (2005)

4. Takahashi, K., Hirata, S., Hino, H.: Biometrical authentication system, enrolment terminal, authentication terminal and authentication server. US 2008/0178008 A1 (July 24, 2008)
5. Yu, Y., Wong, S., Hoffberg, M.B.: Web-based biometric authentication system and method. US 5930804 (July 27, 1999)
6. Miller Jr., S.J., Willis, W.F., Lau, J.H., Dlab, D.: Multimodal biometric platform. US 7298873 (November 20, 2007)
7. Choi, K., Toh, K., Uh, Y., Byun, H.: Service-oriented architecture based on biometric using random features and incremental neural networks. Soft Comput. 16(9), 1539–1553 (2012)
8. Trevathan, J., McCabe, A.: Remote handwritten signature authentication. In: Proceedings of the 2nd International Conference on e-Businees and Telecommunications Engineers (2005)
9. Rusu, S., Dinescu, A., Diaconescu, S.: Systems and methods for assessing the authenticity of dynamic handwritten signature. World Intellectual Property Organization WO/2011/112113 (September 15, 2011)
10. Andrei, V., Rusu, M.S., Diaconescu, S., Dinescu, A.: Securing On-line Payment using Dynamic Signature Verification. LISS (3), 58–65 (2011)
11. Bianco, P., Lewis, G.A., Merson, P.F.: Service Level Agreements in Service-Oriented Architecture Environments. Software Engineering Institute, Paper 294 (2008)
12. Salinca, A., Rusu, S.M., Diaconescu, Ş.: An Approach to data collection in an online signature verification system. In: WEBIST 2012, pp. 251–254 (2012)
13. European Union - Biometric Matching System, `http://www.nws-sa.com/biometrics/EU_Matching_CS.pdf` (accessed January 3, 2013)
14. ISO: ISO/IEC 19784-1:2006 Information technology – biometric application programming interface – part 1: BioAPI specification. International Standards Organization (2006a), `http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=33922` (accessed November 1, 2012)

Extending the Groovy Language Using AST Transformations to Monitor Variables and Methods

Carlos Cortinhas and Fernando Barros

Department of Informatics Engineering, University of Coimbra
3030 Coimbra, Portugal
ccort@student.dei.uc.pt, barros@dei.uc.pt

Abstract. Groovy is a dynamic language based on Java. Groovy enables the development of new operators in order to overcome the language's own limitations. The extensions to Groovy can be made through Abstract Syntax Tree (AST) Transformations. In this paper, we present two AST Transformations to monitor variables and methods in an AspectJ inspired manner. These transformations make it possible to execute custom closures (lambda expressions) at specific points of the application, enabling the improvement of software development.

Keywords: Abstract Syntax Tree, Groovy, Object Oriented Programming, Transformations, Monitors.

1 Introduction

Groovy [1] is a dynamic and flexible language [2] based on Java. In particular, Groovy enables the creation of new operators that extend the language in a user-friendly manner, in order to overcome Groovy's own limitations. The extensions to Groovy can be made through Abstract Syntax Tree (AST) Transformations [3].

In this paper, we use the AST transformations to develop variable and method monitors inspired on the AspectJ concepts of *pointcut* and *aspect* [4] [5]. Variable monitor, allows the user to insert and execute closures (lambda expressions), when the variable is read or written. This can be helpful, for example, for debugging and instrumentation purposes. Method monitor provides the user with the possibility to execute closures before and after the monitored method, enabling, for example, the verification of pre and post-conditions.

2 AST Transformations

One of the major features of Groovy is the Abstract Syntax Tree (AST) Transformations that we can be used to modify the compilation process. These transformations analyze the AST generated by the compiler and change it with custom code.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 745–752.
DOI: 10.1007/978-3-642-36981-0_69

Along the Groovy's compilation process, the source-code is represented in memory by a Concrete Syntax Tree (CST). This CST is transformed into an AST. It is only after the AST is available to the compiler that it produces the byte-code to be run by the Java Virtual Machine. Our transformations are executed after the AST is generated by the compiler.

These transformations allow to access, analyze and change the AST's generated during the compilation process. Therefore it is possible to insert custom code in this tree by adding new nodes. Removing code is also possible by removing nodes from the tree.

The Groovy compilation process has nine phases, which allows the transformations to choose in which phase they will execute. The phases are described in Table 1.

Table 1. Groovy compilation phases

Compilation Phase	Description
Initialization	Source-code is opened and the environment is configured
Parsing	Groovy grammar is used to detect syntax errors and produce the tree of Tokens
Conversion	The AST is created using the tree of tokens
Semantic Analysis	Viability and consistency tests are executed
Canonicalization	Finishes AST construction
Instruction Selection	It's chosen the instructions set to use, Java 5 and newer or prior to Java 5
Class Generation	Creates the byte-code and puts it to memory
Output	Writes the byte-code to the file system
Finalization	Final cleaning and finalization processes are undertaken

Groovy allows two kinds of transformations, described next.

Global Transformation: This kind of transformation is applied to every source-code file compiled. In order to use the transformation it needs to be created and compiled into a .jar file. This file needs to be added to Groovy compiler *classpath*. This ensures that the transformation will analyze and execute on all source files compiled by Groovy.

Local Transformation: Local transformations are quite similar to global transformations. However, there are two main differences. First, a local transformation does not need to be compiled into a jar file. Second, a local transformation requires an annotation in the source code.

The monitors described here are defined as local transformations, and therefore they require an annotation. This choice enables monitors to be easily turned on and off. An advantage of annotations is that they can have parameters, improving their expressiveness. From our experience, Groovy offers a very good support for annotation development when compared to Java capabilities.

3 Monitors

In this section we describe `VarMonitor` and `MethodMonitor`. These transformations can monitor read/write accesses to variables, and method invocation.

3.1 VarMonitor

This AST Transformation monitors a variable and executes closures (lambda expressions) every time a variable is read and/or written. `VarMonitor` offers two options: execute a closure when the variable is read, or execute a closure when the variable is written. In order to use a transformation we need place the corresponding annotation above the variable to monitor, and to configure the respective parameters. For example, if we want to execute the transformation when the variable is read we can use the code of Fig. 1.

```
@VarMonitor (read = {user's code})
Integer var
```

Fig. 1. Template of *read* example for VarMonitor

To execute the transformation when the variable is written we can use the code of Fig. 2.

```
@VarMonitor (write = {user's code})
Integer var
```

Fig. 2. Template of *write* example for VarMonitor

The last possibility is to execute different closures for the read and write parts, which can be done by the annotation shown in Fig. 3.

```
@VarMonitor (read = {user's code}, write = "user's code")
Integer var
```

Fig. 3. Template of *read* and *write* example for VarMonitor

Inside the user's code we can use several special variables:

- `target` – Identifies the object where the variable was created, in both read and write closure.
- `currentValue` – In the read closure, it represents the current value of the variable. In the write closure, it represents the old value of the variable.
- `newValue` –It can only be applied in the write closure, where it represents the new value of the annotated variable.

Inside the user's closure, the value of the annotated variable can be modified. In the example presented in Fig. 4, the transformation executes the closure when the variable is written.

```
class V {
  @VarMonitor(write = {target, oldValue, newValue->
  println "Target: " + target + " Old Value: " +
  oldValue + " New Value: " + newValue})
  Integer k = 20
  void foo(){
    k = k-1
  }
}
new V().foo()
```

Fig. 4. Example use of VarMonitor with *read* and *write*

The transformation is only executed when the variable is decremented. In this example, the closure prints the instance owning variable `k`, and variable `k` previous and new values. The output generated is in Fig. 5.

```
Target: aV Old Value: 20 New Value: 19
```

Fig. 5. Output of VarMonitor example

Next is presented a more practical example. Let's consider a variable that represents the temperature and we want to control its value whenever it's changed anywhere in the code. Instead of adjusting the code everywhere the variable can be changed, or creating specific custom setters and getters, we can create a monitor to check the variable. This monitor will detect where the variable is changed enabling its verification with custom code. Fig. 6 illustrates this example.

```
@VarMonitor(write = {target, oldValue, newValue->
  if (newValue > 100 ||(newValue -oldValue) > 40)
  shutdown()})
Integer temperature
//…
while(1){
  sleep(1000)
  temperature = updateTemp()
}
```

Fig. 6. VarMonitor temperature example

Custom code triggers the `shutdown` function when the temperature rises above 100°, or when the temperature rises abruptly (more than 40°), within a one second interval.

3.2 MethodMonitor

The goal of this annotation is to monitor methods. To accomplish this, the annotation allows the user to specify the closures that can be executed before and/or after the target method is called. In order to use it, the method needs to be annotated as illustrated in Fig. 7.

```
@MethodMonitor(before = {target, par -> User's code},
    after = {target, par-> User's code})
int method(){
      (...)
}
```

Fig. 7. Template of MethodMonitor usage

The transformation takes two parameters: `before` and `after`, referring to the closures that will run before and after the method call, respectively. The parameters of the closures are: `target`, representing the object owning the method, and `par`, the list of the monitored method parameters.

Fig. 8 describes of a simple use of this AST. In the example, all changes in the *x*, *y* coordinates of a moving entity made through method `move` are logged into a file.

```
class M {
  int x = 0, y = 0, limX = 20, limY = 20
  @MethodMonitor(
    before = {Object target, List par->
      new File("log.txt").append("Moving to:
      ["+(x + par[0])+", "+(y + par[1])+"]")},
    after = {Object target, List par ->
      new File("log.txt").append(" Real position:
      ["+x+", "+y+"]\n")})
  void move(dx, dy){
      x += dx; y += dy;
      if (x > limX) x = limX; if (x < 0) x = 0;
      if (y > limY) x = limY; if (y < 0) y = 0;
  }
}
//…
M m = new M()
m.move(10, 5)
m.move(155, -10)
```

Fig. 8. Concrete example of MethodMonitor

In this application we use a coordinate system ranging from 0 to 20 for both `x` and `y` coordinates. The `before` closure logs to a file the user's intention to change the variables `x` and `y` by `dx` and `dy` and stores the values these variables would take.

Afterwards, the original method executes and computes the new values of `x` and `y`. Finally, the `after` closure logs the final coordinates into the file. Fig. 9 shows the log's content.

```
Moving to: [10, 5] Real position: [10, 5]
Moving to: [165, -5] Real position: [20, 0]
```

Fig. 9. Content of the log file after the execution of MethodMonitor example

Method `move` is invoked twice. In the first execution, we observe in the log that no readjustment was needed to the coordinates as both were within the limits. In the second execution, the coordinates would assume values outside the allowed range. However, the `move` method has corrected the coordinates and the `after` closure has logged the final values.

4 Related Work

Groovy comes packaged with some AST transformations. These are usually simple transformations but very useful. Some of them are:

- `@Grab` - Allows the automatic download of the required libraries to the project *classpath*. With this annotation, the user does not have to search and download the libraries needed to execute the project.
- `@EqualsAndHashCode` - Generates `equals` and `hashcode` methods for the annotated class.
- `@ToString` - Generates a `toString` method for the annotated class, and prints out the formatted content of class's properties.
- `@Immutable` – With this transformation, annotated objects cannot change after their creation. This is an easy way to guarantee that, for example, a field does not change its value after creation.

An example of the approach we took in this paper is present in AspectJ [6], a framework supporting Aspect Oriented Programming (AOP). In AspectJ we can define three types of *aspects*:

- Inter-type declarations - Allows the user to add methods and fields to classes. This can also be achieved with AST Transformations.
- Pointcuts - Describe points in the application that can execute, for example, custom code specified by the *advice* (described next).
- Advices - Contain custom code that will be executed in the points defined by *pointcuts*.

The `VarMonitor` and `MethodMonitor` transformations were inspired in AspectJ features. However, AspectJ is essentially non-modular while AST Transformations are modular, meaning they are applied only where the annotation is declared. This can

be useful, for example, to enable and disable them quickly. AspectJ code can be applied to many methods at once. In our approach, we find more likely that each method requires a unique treatment. Additionally, AspectJ can only be used with Java and we haven't managed to make it work with Groovy. The lack of this functionality has motivated the development of this work.

GContracts [7] is a popular module of Groovy that enables the use of contract annotations to specify class-invariants on Groovy classes, since it allows the specification of pre and post conditions to the method or class annotated. GContracts makes use of AST Transformations to develop the module.

The available annotations in GContracts are:

- `@Invariant` - An invariant is a condition that must stay true during object lifetime.
- `@Requires` - Specifies a pre-condition for the annotated method.
- `@Ensures` - Specifies a post-condition for the annotated method. It ensures that the condition sent as a parameter is true when the method finishes. The annotation also enables the access to the old values of the method's parameters and to the return the value of the method.

The parameter passed to these annotations is a closure that specifies a assertion/condition. If the assertion fails the compilation will fail and informs the user where it failed.

The main difference between GContracts and the monitors described here is that we enable the user to insert, not only assertions, but general custom code into the annotations.

5 Conclusion and Future Work

AST Transformations provide a simple and effective mechanism to extend the Groovy language. In this paper, we have described two new Groovy Transformations to enable the monitoring of methods and variables. These transformations are very useful since they allow code to be easily added/removed to/from an application enabling the easy modification of software. As future work we plan to increase the capabilities of the current transformations in order to support, for example, annotation inheritance. Another goal is to enhance `MethodMonitor` providing access to the return value of the annotated method. The introduction of additional operators inspired by AspectJ is also planned.

Acknowledgments. This work was partially supported by the Portuguese Foundation for Science and Technology under project PTDC/EIA-EIA/100752/2008.

References

1. Groovy, `http://groovy.codehaus.org/`
2. Konig, D., King, P., Laforge, G., Skeet, J.: Groovy in Action, 2nd edn. Manning (2011)

3. Groovy - Compile-time Metaprogramming - AST Transformations, http://groovy.codehaus.org/Compile-time+Metaprogramming+-+AST+Transformations
4. Colyer, A., Clement, A., Harley, G., Webster, M.: Eclipse AspectJ: Aspect-Oriented Programming with AspectJ and the Eclipse AspectJ Development Tools, 1st edn. Addison-Wesley Professional (2004)
5. Laddad, R.: AspectJ in Action: Enterprise AOP with Spring, 2nd edn. Manning Publications (2009)
6. Kiczales, G., Lamping, J., Mendhekar, A.: Aspect-oriented programming (June 1997)
7. GContracts – A Programming by Contract Extension for Groovy! https://github.com/-andresteingress/gcontracts/wiki

Network Based Analysis of Intertextual Relations

Ioan Cristian Ghiban[1] and Ştefan Trăuşan-Matu[2]

[1] University "Politehnica" of Bucharest,
313 Splaiul Independenţei, 060032 Bucharest, Romania
[2] University "Politehnica" of Bucharest,
313 Splaiul Independenţei, 060032 Bucharest, Romania,
Institutul de Cercetări în Inteligenţă Artificială,
Calea 13 Septembrie nr. 13, 050711 Bucharest, Romania
{ioan.ghiban,stefan.trausan}@cs.pub.ro

Abstract. We present an approach of intertextuality in terms of graph theories, statistics, and bakhtinian polyphony, the latter perspective considering the way in which discourse threads are influencing each other. This paper presents theoretical models, processing techniques with their applications and results that acknowledge as important the approach based on networks of intertextuality. In the end are introduced two original applications of supervised and unsupervised analysis of antique texts of philosophical and religious nature.

Keywords: intertextuality, discourse, polyphony, text mining, natural language processing.

1 Introduction

In linguistics, because structuralist approaches focus on signs and at most on the relationships between them, there was a drawback when speaking in terms of texts and their interdependencies. The texts were considered as discrete and closed-off entities loosing the overall view even when the surveys were done on corpora. Having this in mind, the notion of intertextuality was introduced by Julia Kristeva [18] probably influenced by Bakhtin's polyphonic theory [6] (that each discourse is a dialog of the discourses that underlie at his composition).

Intertextuality represents the aggregation of the ways in which a text infers the knowledge of other texts [9]. Actually, intertextuality regards all forms of discourse, not only written text, and has applications in various semiotic systems for knowledge representation. Each author of a discourse uses a certain context of knowledge retrieved from other authors. His result will be a "mosaic" of quotations, of discourse sequences, enriched by the author's new knowledge.

The paper firstly presents theoretical and practical aspects regarding intertextual networks (i.e. analysis networks of intertextual relations). The third section describes applications that were developed for research in language dynamics, starting from these theories. Here are also presented the projects' results, together with the more important ones from other studies. In the end, we describe two software implementations of text analysis, related to intertextual networks and configured for humanistic texts.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 753–762.
DOI: 10.1007/978-3-642-36981-0_70

2 Theoretical Apparatus

Intertextuality is underlined by cohesion links and, based on them, we can classify it as referential or typological. The referential cohesions are based on explicitly marked references among texts (e.g. hyperlinks, citations).The typological ones are based on a common typology (e.g. vocabulary, genre, theme, URL address) that categorizes the texts into classes [16].

Studies on referential intertextuality have delineated several cases, starting from the approaches they take on surveying the linkages amongst texts. Regarding the bibliographical linkages, four main branches may be considered [16]: Informetrics, Bilbliometrics, Scientometrics, Webometrics.

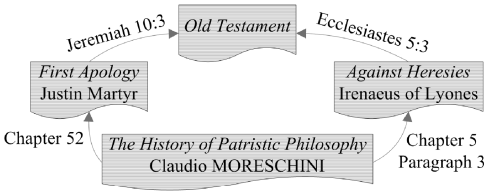


Fig. 1. Bibliometric Intertextuality

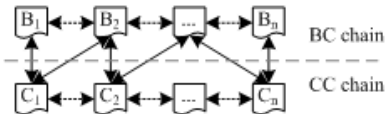


Fig. 2. BC and CC networks reduced to a bipartite graph

Scientometrics and other related fields use the term of bibliographic coupling if two documents have at least one common document that are citing together (e.g. in Fig. 1 "First Apology" and "Against Heresies" are bibliographically coupled), and if they both are cited in the same document then they are called co-cited (in Fig. 1 "First Apology" and "Against Heresies" are co-cited because "The History of Patristic Philosophy" is citing both). These two relations induce bibliographic coupling (BC) networks and co-citation (CC) networks respectively (Fig. 2). Weights can be attached to the links according to the number of common citations. Both types of networks can be mixed in a single network and if all documents in co-citation relations are distinct to the documents in bibliographical relations a bipartite graph can be inferred.

Analysis can stress the time windows comprising the publication date of the cited documents or of the citing ones. In this way, Redner and Raan [4] studied the relations between the co-citations from a certain period and the citing documents in a latter one, or reversely the bibliographically coupled documents in a certain period and the cited documents from an earlier one.

In typological intertextuality, graphs are used to describe the relations between distinct lexical or conceptual units. Unsupervised processing is used for building lexical systems like collocation graphs and, with the help of lexicographer's expert knowledge, supervised processing is used for building terminological ontologies (e.g. WordNet), thesauri (e.g. Roget) and other graph related systems that use sense relations (e.g. synonymy or polysemy) or conceptual relations (e.g. hyponymy, meronymy etc.). Therefore, three types of supervised lexical graphs emerge [16]:

1. Conceptual graphs using conceptual relations as edges and vertices as concepts (e.g. synsets in WordNet)
2. Thesaurus type graphs with sense relations as edges and words as vertices
3. Association graphs using association or word priming for representing edges between the words they describe.

Different types of graphs are built in an unsupervised manner: co-occurrence graphs, collocational graphs, sentence graphs and other types. The last one refers both to a lexical perspective and to another one with regard to sentences. In the former perspective the lexical units are linked if they co-occur at least in one sentence. In the second perspective the graph links the sentences if they share at least two lexically cohesive words [7].

Regarding the generation of collocational graphs, firstly, the interesting co-occurrences are found through some specific metrics [8]. Then, these metrics are also used to link the co-occurrences that share terms with semantic proximity. Afterwards, distinct threshold filters are applied to reduce the co-occurrences to collocations. Frequently used in this kind of graph based analysis, Latent Semantic Analysis (LSA) starts from the premise that lexical units with similar topics, themes or fields, share, with a big probability, related or common terms [16]. Here may be used Salton's vector space [5] model representing texts as term vectors. The results of processing this space are lexical classes or clusters and associative networks.

2.1 Models of SmallWorlds

The topologies and statistical characteristics of the intertextual networks are sensitive to the discourse genres of the surveyed documents (e.g. the citations in a scientific community link the documents distinctively to how the hypertext links do it in wikis). Studies were made trying to unify the properties of intertextual networks under the so-called Small World (SW) property [3] (e.g. this is manifested also by social networks). It gives measurement to what differentiate random graphs and the natural, nonartificial intertextual ones by means of (Eq. 1) considerably higher level of clusterizations and (Eq. 2) considerably shorter distance between any two random vertices in the graph.

Some researchers [17] proposed a SW model, shortly named WS model, that maximizes the clusterization coefficient CWS (Eq. 1) and minimizes the average geodesic distance L (Eq. 2):

$$C_{WS}(G) = \frac{1}{n}\sum_{i=1}^{n} C_{v_i}(G) \in [0,1]\,. \tag{1}$$

$$L(G) = \frac{1}{\binom{|V(G)|}{2}} \sum_{\{v,w\}\in[V]} \delta(v,w)\,. \tag{2}$$

$$C_{v_i}(G) = \frac{adj(v_i)}{\binom{d_G(v)}{2}} = \frac{adj(v_i)}{d(v_i)(d(v_i)-1)/2} \in [0,1] \tag{3}$$

Equation 3 gives the clusterization of node vi (used in Eq. 1), adj(vi) is the number of edges between vi's neighbors, d(vi) is the degree of vi node and δ is the distance between v and w. As a result, CWS indicates the probability that the neighbors of a randomly selected node v are also neighbors to each other and L indicates how fast is changing a variable like document topic or genre when following the links (e.g. Wikipedia has a small L and the topic is changing fast when following the links).

For the automatic generation of a Small World, a totally random approach is not working. Therefore, other statistical models are suited for better fitting the intertextual networks. For instance, Barabási and Albert [1] take the Zipf law as a paradigm, proposing the preferential attachment model (BA model) for describing the dynamics of intertextual network growth.

$$P(k) \sim k^{-\gamma} . \tag{4}$$

The distribution states that a randomly selected node will have a k degree with a probability of $k^{-\gamma}$ where usually $\gamma \in [1.5,3.5]$.

3 A Survey on Language Dynamics and on Results

There is much research on the fields of referential or typological intertextuality and, in the interest of this paper, a sample of software tools for network based analysis on intertextuality will be presented.

3.1 HSCM and Linguistic Networks

Aiming to find the dynamics of language use in a certain society along periods of time that span even to millennia, as a consequence of the researches in the field mainly focused on probabilistic methods [14] (e.g. probabilistic grammar or statistical collocation analyses) disregarding instead the long-term processes implied by language dynamics, a German team (including historians and software engineers) implemented in 2007 a system named HSCM (Historical Semantics Corpus Management System). It takes as input chronologically layered corpora (researches were made on "Patrologia Latina") and finds the semantic changes of (rather statistics through co-occurrences) words in diachronic texts giving hints with regard to the social changes that may have took place between the years in which the texts were written.

HSCM combines a corpus retrieval system with a list-oriented lemmatization of Latin [14] and generates highly customizable statistics on words' usages along texts (e.g. charts on term-document matrix, various types of filters or of output display). The system finds the co-occurrences determined by the specified word or phrase query and gives through frequency-charts indications of stabilization, marginalization, and transformation of language patterns useful to historical field.

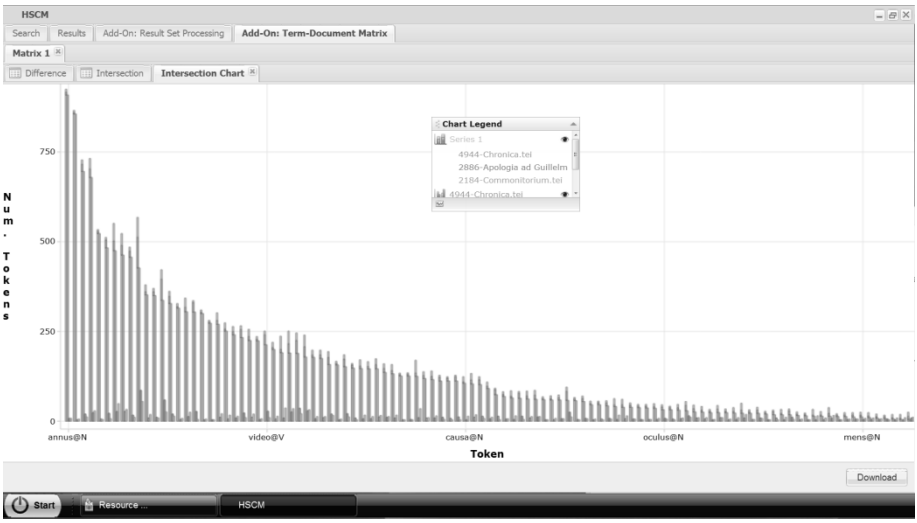


Fig. 3. Words' frequency in three texts viewed with HSCM (http://www.hucompute.org/ressourcen/historical-semantics)

Another system, called "Linguistic Networks" (http://www.linguistic-networks.net/) provides web tools for inducing lexical and sentence networks on corpora. Their nodes and the links between them are built by means of syntactic (i.e. part-of-speech), semantic (i.e. meaning) or pragmatic [15] (i.e. meaning in context) approaches.

The system is designed to work on distinct corpora (e.g. "Patrologia Latina") with their specific languages and delivers various types of networks (e.g. word, lemma, sentence or text). It provides other details like word's lemma in dictionary, in what forms (the one before stemming) or with what PoS was the word encountered, most frequent co-occurrences, left or right neighbors (i.e. the words positioned before or after the focused word in the same collocation with it) filtered or not by their frequency ranking. Also, the network display can be customized by weight labels on links, distinct node colors depending on PoS or various types of structuring (e.g. circular, organic, hierarchic). Other graphics with the neighbors are available: word cloud and neighbor distribution (rank vs. significance).

3.2 Some Results

Interesting outcomes were obtained with researches done on referential intertextuality, mainly on BC networks [13]. For instance, the researchers identified publications called "sleeping beauties", unnoticed for a longer period of time and suddenly cited to a high degree. Other studies were made on the connectivity in wiki hypertexts [2] and in blogosphere (i.e. the whole network of blogs) [12].

Approaches on typological intertextuality applied on British National Corpus by means of co-occurrence and collocation graphs were taken by some researchers [11]. As results, they observed that both types of graphs match the SW property in terms of

WS and BA models. Studies were also conducted on sentence graphs [10]. Besides the fitting with the WS and BA model, a particularity named disassortative mixing was observed: the words with a high connectivity degree (e.g. common nouns, function words etc.) are more likely linked with the poorly connected ones.

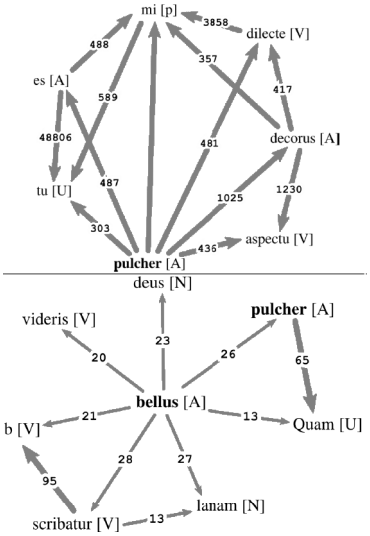


Fig. 4. Example of language dynamic: adjective "bellus" (lower graph) gradually replacing "pulcher" (upper graph) for representing "pretty" [15]

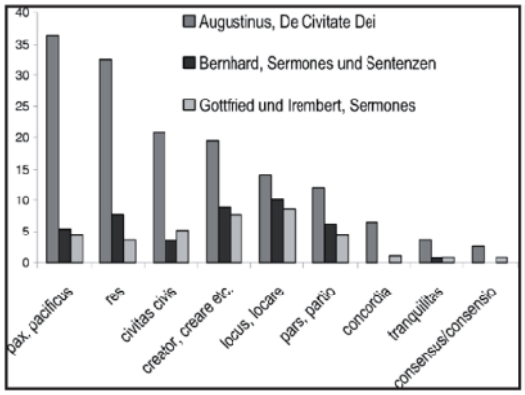


Fig. 5. Analysis of ordo's usage in three corpora (one in the 5th and two in the 12th century) by means of collocations [14]

The "Linguistic Networks" project, as mentioned above, induces lexical and sentence networks on corpora. The results are useful to determine new semantic artifacts replacing in time the old ones thereby getting the lexical dynamic of the represented concepts. In the induced networks the artifacts associated to a certain concept will appear as neighbor nodes and, as time passes by, the new ones may get more important (see Fig. 4).

Useful outputs regarding language dynamic can also be computed with the HSCM system. With this, history scientists can compare humanistic works written in distinct centuries to see how lexical artifacts' usage evolves in various configurations.

For instance, there is a widely shared hypothesis [14] that the Latin term "ordo" (i.e. rank, class, order) was central in medieval period to express concepts about world and its social structure. Also, there are works, like Augustine's "De Civitate Dei", considered to be seminal due their strong influence they had on the conceptualization of "ordo" in subsequent works. HSCM can give cues on the usage of "ordo" along Augustine's 5th century work and other latter religious works, like those written in the 12th century.

For Fig. 5, HSCM founds in Augustine's work the two words collocations containing "ordo". Then, takes from these collocations the other words as targets (e.g. "pax", "res") and finds all their collocations. After a ranking according to the percentage of collocations containing "ordo" in the overall collocations of the target words the same percentage computing is done also on the other latter works. One can notice that "ordo" doesn't appear so frequently near "pax, pacificus" or "res", and historians can infer conclusions concerning these details.

4 Tools for Analyzing Intertextuality on Humanistic Texts

For studying the intertextuality we have implemented an application using an unsupervised technique (LSA) and another one for offering support to the human experts in their supervised process of getting the main concepts behind the text. Due the online availability, we have used, as those from HSCM with their "Patrologia Latina", corpora representing antique writings of philosophical nature, namely Plato, Aristotle, Epicurus and other philosophers of antiquity, and also writings of religious nature, confining only to the Christian patristic (starting with the 2nd century until the 8th century). The choice was also reasoned by the presence of a long examination and criticism concerning the antique philosophical influences on the patristic writings. Thereby, we should have supervised information regarding the intertextual relationships in the framework of the corpus being analyzed.

We have accessed 17 sites that are publishing, for free, antique texts (e.g. gutenberg.org, tertullian.org, newadvent etc.). We have downloaded nearly 1800 files corresponding to the various sections from the writings, and it was build up a corpus of texts (books, epistles, discourses, hymns) for which the critics have expressed interest.

Regarding the local corpus, since the text formats were distinct, for reaching to a common format, and also for keeping the texts' structure on sections (these being useful due the indication of text sequences with the same subject), we have implemented a conversion module. Of the 17 sites, for allowing a further expanding of the local corpus, we have chosen 4 for which the application allows the direct conversion of the html files, recognizing through tags the texts' structure. For the files from the other sites the txt format was used, the writings' structure being kept through the recognition of the titles with the help of regular expressions. In the end, we have got 280 books in TEI P5 format.

4.1 Text Processing Based on LSA

We have implemented in Python an application that finds the most frequent words, the most significant ones and also the similarities amongst documents using a computed conceptual space of these documents. The implementation is processing files contained in one or more directories. The user can group files in these directories according to a common property (e.g. the same author). Each file contains paragraphs and one can consider as document the entire content of the file, or a single paragraph (this is useful for the term frequency matrix).

The processing of a document (written in English) is made with the help of the NLTK environment. It starts with a word recognition with regular expressions followed by stop words filtering, tokenization with the Brown corpus (religious section), stemming using the Porter stemmer and WordNet. In tokenizing the words there were used three taggers: trigram (finds the POS based on the current word and the preceding two), bigram, and unigram. If no result is found the word is tagged as "Noun". The application will keep only the words that are interesting: nouns, adverbs, verbs and adjectives.

As a word filtering, after the term frequency matrix was computed, the application is keeping only the words that have at least 5 occurrences in any of the documents. Next, TF-IDF is computed and eventually the singular value decomposition is performed using the svd function available in the Numpy library.

To save time in processing texts by the user, the application is saving the results in four steps: the parsed documents, the lexical space - after the computing of the term frequency matrix, the semantic space - after the decomposition in singular values and the last step given by the saving of the final statistics under the form desired by the user. Thus, the subsequent processing can begin from a step already performed.

The final results are ordered and displayed by the score in a listed form of statistical element, score pairs (i.e. the statistic element can be a word or a text). For example, comparing with texts from Plato, the first two similar texts written by other authors are as follows:

[Plato\TheApology, Justin | TheSecondApology-(0.6475); Plato | TheRepublic.7, Irenaeus | AgainstHeresies.6-(0.6095)]

This means that Plato's apology with Justin's "Second Apology" have a score of 0.6475, being the most similar two works, and chapter 7 from Plato's Republic with chapter 6 from Irenaeus' "Against Heresies" are on the second place. The results are also graphically available.

The testing was done on works of St. Justin Martyr (103-105), St. Irenaeus (second half of the 2nd century), Origen (185-254) and Plato's works. Comparing the St. Justin's documents with Irenaeus and Plato's works, in the top 10 most similar works there were found eight wrote by St. Irenaeus and two by Plato: "Cratylus", similar to Justin's "The Second Apology" and "Lysis" similar to Justin's "Discourse to the Greeks".

In the first and the second Apology, St. Justin wrote to the pagans and in the second (being addressed to the Greeks) is using many terms typical to the Hellenic philosophical environment: "truth", "vice", "virtue", "good", "soul", "philosopher", "Gods". Also, Justin is trying in "Discourse to the Greeks" to persuade that the Christian religion is for the entire humanity and he is again using specific Greek terms.

4.2 Text Annotator

Many patristic writings are to be found on http://www.newadvent.org/fathers/. The resources hosted at this site are hypertexts that embody different types of comments (tagged) on the raw texts. These comments are usually labeling key-words or are linking to details and may prove useful in getting the concepts behind the text.

Having this in mind, it was needed an application to extract the comments from the above mentioned site. "Text Annotator" (Fig. 6) automatically annotates the newadvent.org pages based on the encountered html tags and also helps users to make annotations on html or txt files using a given set of comment types. With further development, our final purpose is to offer statistics about the supervised concepts inside a text and compare to those found with the previous application.

The input texts are usually structured in books, chapters, paragraphs and other elements of users' interests and, using this inheritance format, the application is building a tree of annotations. After this, one can insert, move or delete an annotation and collapse the substructures to easily manage the resulting tree.

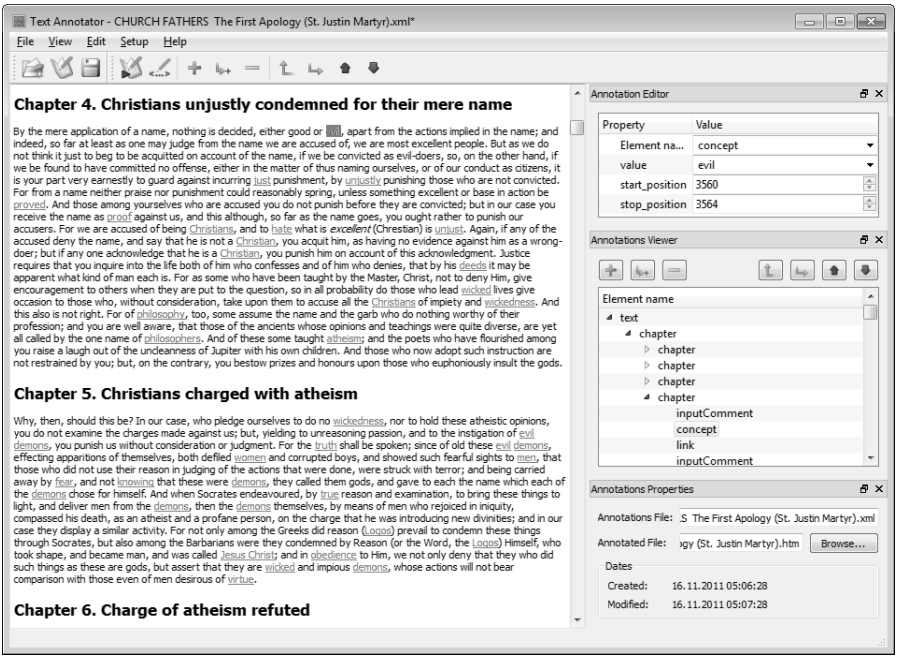


Fig. 6. Results in "Text Annotator" after processing "The First Apology" wrote by St. Justin Martyr

Each annotation can be edited manually by changing the values of the attributes. Also the attributes themselves can be modified (in terms of name and type) through changes in the Elements.xml configuration file. From here one can also add new types of annotations with their attributes and possible values.

5 Conclusions

As we have seen, both referential and typological intertextuality can be described through concepts related to graph theory. More specific, they can be described by the various models proposed for characterizing the so called concept of Small World.

Having these models, one can make forecasts regarding the dynamic of the social networks, of citations, sitations (i.e. the hyperliking between documents), the blogosphear or of other areas that demand interactions similar to the social ones.

Implementing such theories has proved the utility of networks to analyze the relationships between texts. In this manner, we benefit from results that reveal linguistic dynamics or philosophical influences over the writings, gains useful at least to the historians or linguists.

References

1. Barabási, A.-L., et al.: Scale-free characteristics of random networks: The topology of the World Wide Web. Physica A 281, 69–77 (1999)
2. Bianconi, G., Barbási, A.-L.: Bose-einstein condensation in complex networks. Physical Review Letters 86(24), 5632–5635
3. Milgram, S.: The small-world problem. Psychology Today 2, 60–67 (1967)
4. Redner, S.: How popular is your paper? An empirical study of the citation distribution. European Physical Journal B(4), 131–134 (1998)
5. Mehler, M.A.: Text minning with the help of cohesion trees. In: Gaul, W., et al. (eds.) Classification, Automation, and New Media: Proceedings of the 24th Annual. Springer (2002)
6. Bakhtin, M.: Problems of Dostoevsky's Poetics. University of Minnesota Press (1984)
7. Hoey, M.: Patterns of Lexis in Text. Oxford University Press, Oxford
8. Manning, C.D., Schütze, H.: Foundations of Statistical Natural Language Processing. MIT Press, Cambridge (1999)
9. Tanskanen, S.-K.: Collaborating towards Coherence. John Benjamin, Amsterdam (2006)
10. Ferrer i Cancho, R., et al.: Patterns in syntactic dependency-networks. Physical Review E(69), 051915 (2004)
11. Ferrer i Cancho, R., Solé, R.V.: The small-world of human language. Royal Society of London. Series B, Biological Sciences 268(1482), 2261–2265 (2001)
12. Glance, N., et al.: Blogpulse: Automated trend discovery for weblogs. In: Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, WWW 2004 (2004)
13. Glänzel, W., Czerwon, H.-J.: A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level. Scientometrics 37(2), 195–221 (1996)
14. Jussen, B., et al.: A corpus management system for historical semantics, sprache und datenverarbeitung. International Journal for Language Data Processing 31(1-2) (2007)
15. Mehler, M.A., et al.: Time Series of Linguistic Networks by Example of the Patrologia Latina. Goethe-Universität Frankfurt am Main. Course presentation
16. Mehler, M.A.: Large text networks as an object of corpus linguistic studies. Technical report, Bielefeld University, D-33615 Bielefeld, Germany (2007)
17. Watts, D.J., Strogatz, S.H.: Collective dynamics of "small-world" networks. Nature 393, 440–442 (1998)
18. Chandler, D.: Semiotics for Beginners, `http://www.aber.ac.uk/media/Documents/S4B/`

Building an Integrated System for the Management of Scientific Nature Events through Open Source Software Integration

Carlos Serrão and Miguel Gamito

ISCTE-IUL/ADETTI-IUL, Ed. ISCTE, Av. das Forças Armadas,
1649-026 Lisboa, Portugal
carlos.serrao@iscte.pt, miggamito@gmail.com

Abstract. This article presents the development of a specific software system through the integration of several open source applications. This integration makes use of several synchronous Enterprise Application Integration (EAI) techniques that were grouped as a way to keep the applications the most independent possible. The integration process was conducted directly on the source code of the applications, making usage of open source applications as an asset. This article presents and approach that allows the development of new software solutions through the integration of several open source applications.

Keywords: Open source, Software Integration, web services, API, cURL, Source code, Database, EAI.

1 Introduction

The main objective of this article is to present an approach to build a system that covers all the phases related to the management of scientific events, like scientific conferences, workshops, exhibitions and others. Most of the current open-source scientific event management tools are too focused on a specific part of the event, paper submission and management, while the other aspects are not covered [14][15][16]. There are important and specific aspects of these types of events that are not covered by any open-source tool (even in the commercial panorama, these tools do not exist)[17]. All the efforts that have been made so far are always focused on improving those specific aspects instead of covering new ones[18]. The purpose of this work was to build an open-source scientific event management system, through the integration of multiple third-party open-source solutions that offer an end-to-end solution for the organization, promotion and management of these types of events. The organization of these events involves a different set of phases that need to be covered by this tool. These phases are displayed as workflows in the following image (Fig. 1). On the right workflow are described the common steps to all event types (workshops, fairs, and others). On the left workflow are shown the phases that are specific to a scientific event.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 763–772.
DOI: 10.1007/978-3-642-36981-0_71

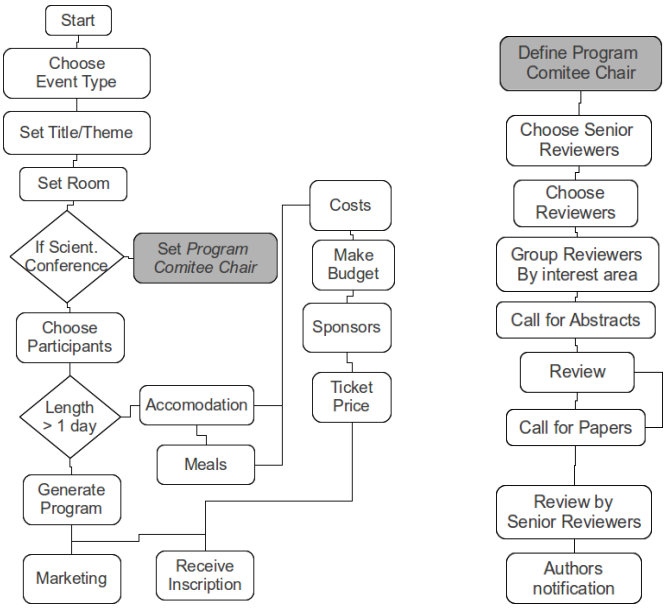


Fig. 1. Workflow of System to develop

The need to build this system emerged from the fact that all the main open-source applications that manage scientific events are more focused on the management of the papers (submission, evaluation, and other) and neglect other aspects such as the financial/budget management, the event marketing and communication and more important the participants inscriptions. In a previously conducted analysis it was possible to conclude that most of the available software solutions, some of them open-source, in fact, do not cover such aspects. Among the considered solutions were EasyChair, ConfTool and OpenConf [10].

To build a new and integrated system one of the usual approaches consists in building it from scratch, which is time consuming, expensive, and error prone. In this article the approach proposed consisted in the integration of several existing open source software applications that match the necessary requirements of the final solution.

Over the past few years open source has passed from a cheap alternative to commercial software to become one of the major references for some software categories [1][13]. Examples of such success include Apache Web Server and Firefox Web Browser [1][9]. The usage of open source software to build a new system is a viable solution.

Currently, open source software is considered to be cost-efficient and high quality [12][13].

Software integration, or more specifically Enterprise Application Integration (EAI), has emerged as one of the solutions for the poor data exchange between the several applications present on the organizations. This lack of integration was caused by the evolution of the information systems as isolated islands along the years. Also

the merge and acquisitions of companies by other companies, has contributed to this problem [3][7][4].

This concept is divided on synchronous and asynchronous techniques. While synchronous techniques consist on the direct communication between the integrated applications, resulting on a point-to-point architecture, the asynchronous techniques require a middleware solution that is used to administer the communication between the integrated applications [3][4].

This work has combined both concepts to build an end-to-end scientific event management system. Open source concedes the integrator the possibility to add or alter the source code of the different applications to integrate, despite the specific openness and interoperability facilities offered by the open-source application to integrate in the system.

The following section of this paper explains the generic approach that was used to build the integrated system, using the EAI concepts applied to open source applications. In the next section it will be presented the open-source applications that were selected to integrate in the system. The following section presents the integration of the applications and finally some conclusions are exposed.

2 Integration Methodology Proposed

The methodology that was adopted to conduct this work uses an incremental method and must be applied using a top to bottom approach. This approach allows the application present on the integration to be changed by newer versions or by other similar applications in an easy manner. The following section describes the methodology that was used and presents the different steps conducted.

2.1 Methodology

This section describes the methodology that was applied for the integration of open-source applications (Fig. 2) and the development of the desired system. The integration process starts by conducting an in depth requirements identification of the solution to be developed. This requirements identification is an important selection criterion that will be used to select the appropriate open-source applications that will be integrated to build the system.

After this initial requirements identification, it is necessary to select an application to be used as base system for all the integration. The application chose should be the one that covers most of the identified criteria and the one that meets the major number of identified requirements.

For each of the requirements not entirely covered by the base application, it is necessary to select an open source solution capable of implementing the missing requirements. After the selection of the solution has been concluded, the following steps need to be executed:

1. Test the application to integrate within the system. In this step it is necessary to make an inventory of the variables used by the forms of the application. To be able to do this it is recommended the usage of an IDE with debugging capabilities.
2. Integrate the Application into the system. It is necessary to select one of the following techniques to use:

2.1 Web Services: If the application to integrate has a web services based interface, it is recommended to use them to integrate in the application integration in the system.

2.2 API: If the application exposes a public API, than it can be used to perform the integration.

2.3 Application Forms: Use the application forms to enter the SQL-DML data into the database. Through the usage of the applications forms the risk of messing the application integrity is lower. For this, it can be used a tool like PHP-cURL. For this integration, the variable inventory built on step 1, will be used.

2.4 Direct use of SQL-DML: This option should be used with extreme caution, since it is the most dangerous one. The integrity of the data on the database can easily be compromised by this technique.

3. After the open source application has been integrated, the system should be tested in order to assess its functionality.

Each of the times it is necessary to alter directly the source-code it is recommended to document such alterations through the usage of comments in the source-code.

This way it will always be possible to easily discover the code created or altered, in the event of system problems occur, or when upgrading a specific application part of the integrated applications pool.

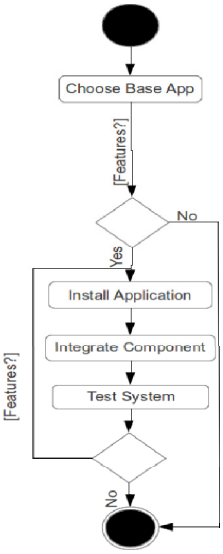


Fig. 2. UML Activity diagram of the methodology

The methodology resorts mainly to synchronous integration techniques. From an architectural point of view it is a point-to-point architecture. Although authors, along the years, have criticized this kind of architecture, it is still considered as a viable solution for the integration of a small number of applications [7][3]. Talend [8] says that the modification of the source code is a viable solution for small integration projects, over the use of an open source integration application like Jboss[1] or OpenESB[2]. The biggest problem with this methodology arises from the fact that it is synchronous, and therefore has no transactions support mechanism. In order to minimize this problem it was created a specific table for the log. Every time an operation integration is made, a row is inserted on that table [12].

3 Applications Selected for the Integration

The different open-source applications that were selected to conduct the integration and build a new system are presented in this section. These applications have been selected from two of the major open-source project repositories: Freecode and Sourceforge. The selection of the open-source applications to build the new system has been guided by a set of well-defined criteria. This criteria is further detailed in the following sections.

3.1 Selected Open-Source Applications

The first application selected was considered to be the "base application". This base application should offer the largest requirements coverage as they were defined for the final system to develop – the end-to-end scientific event management system. At the same time, the application source code, and the database schema should be easily understandable and modifiable. The selected open source application, to assume the base application role, was **REM**[3] (Related Event Management). Despite this application has just a few downloads and recommendations on the repository, its source code covers a vast number of phases on the workflow and requirements. Also, its source code is well structured and easy to modify, a major requirement for the integration. Moreover REM allows the establishment of relations between different events.

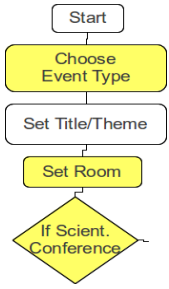


Fig. 3. REM on the Workflow

[1] http://www.jboss.org/
[2] http://www.open-esb.net/
[3] http://sourceforge.net/projects/relatedeventsma/

Scientific conference events, which are presented on the right workflow (Fig. 1), shall be managed by **OpenConf**[4] and **SMF**[5] (Simple Machines Forum). OpenConf is one of the most used open source applications for this purpose (openconfportfolio, 2011). It is at the same time a very active and well-documented project. The SMF was used to build a discussion forum for the conference, creating a communication channel for all the participants on the event. This idea is not original - one of the open-source applications analyzed [10], **MyReview**[6], uses a similar approach.

For the financial/budget management, it is recommend that the integration is performed with the financial software that is used by the organization (if it already uses one). In order to keep the system as simple as possible the integration was made with the **WMMG**[7] (Where My Money Go) open source application, which is a very simple financial/budget management application.

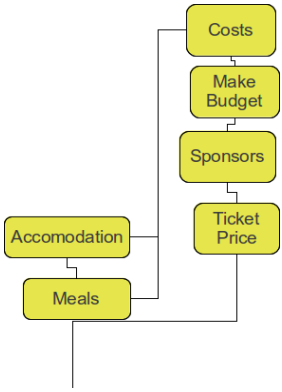


Fig. 4. WMMG on the Workflow

The last application selected was **CiviCRM**[8]. This application is build as a CMS module, and it can be used with all the main open source CMS platforms [11]. Drupal was used as CMS. This application is responsible for the remaining phases of the workflow, marketing and for receiving online inscriptions. The bundle (Drupal + CiviCRM) is also used as the system frontend.

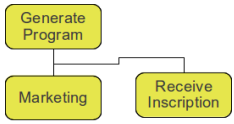


Fig. 5. CiviCRM on the Workflow

[4] http://www.openconf.com/
[5] http://www.simplemachines.org/
[6] http://myreview.sourceforge.net/
[7] http://sourceforge.net/projects/wheremymoneygo/
[8] http://civicrm.org/

The remaining phases that were not entirely covered by these applications were developed as external modules, in an effort to keep the number of applications to integrate as low as possible.

4 The End-to-End Scientific Event Management Application Integration

In this section, the end-to-end scientific event management solution implementation is explained, using the proposed open-source application integration approach. This integrated solution would be capable of supporting the different aspects of organizing any type of events, but with a special focus on scientific events. The integration is going to be contextualized on the EAI concept techniques and will use the approach described in the previous section.

4.1 Integration

The first integration step was conducted between the base application, **REM** and **OpenConf** and **SMF**. The selected integration method was the integration through the applications forms (**php-cURL**) for **SMF.** In the specific case of **OpenConf**, the OpenConf installer was invoked with information passed as **session variables**. This integration was conducted at the interface layer.

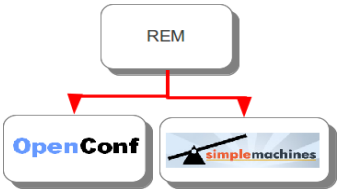


Fig. 6. Integration REM to OpenConf and SMF

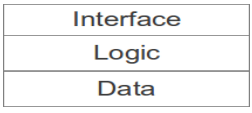


Fig. 7. Typical information system layers

The second integration conducted uses the same approach as SMF. The integration was also conducted at the interface layer and the **applications forms** were used. This was performed between **REM** and **WMMG**.

The final integration step was conducted between **REM** and **CiviCRM**. This integration occurred at the application's logic layer. It is in this layer that the referential integrity of data is guaranteed, therefore making it the appropriate place to conduct all the integrations. If the integration were performed at the interface layer,

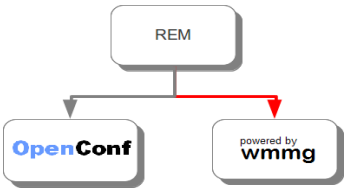


Fig. 8. Integration REM to WMMG

although the referential integrity would have been maintained, there would be a loss of application performance – mostly due to the fact that this interface was developed mostly for human interaction which takes more time than application interactivity. This integration makes use of the **API** integration (step 2.2 of the presented methodology). CiviCrm has a very complete API, which allowed ease integration. Integration from **CiviCrm** to **REM** also occurred, but in this case the integration methodology selected was **SQL-DML** (using the step 2.4 of the methodology). This integration shows why the source code on the base application should be easy to understand. Normally all integrations to the base application will be made using point 2.4, SQL-DML. In this particular case this integration allowed the same functionality on both the back office and the frontend applications.

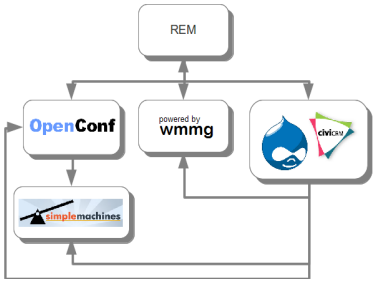


Fig. 9. Full System integration

5 Conclusions

This article described the process of integrating several open source applications to build a new integrated system, capable of managing all the different aspects of scientific events. The resulting system can manage all the procedures involved with scientific events management, including financial/budget management, event marketing/promotion and even accept online participant inscriptions.

The methodology used for the integration was designed to keep the applications as independent as possible from each another. This will allow to easily changing one application for another.

The use of open source applications on an integration project was proved to be an asset, because it allowed the integrator to adjust the source code facilitating the integration.

Using an integration approach instead of the development of a new application has several advantages:

- Lower costs: due to a reduced development time.
- Applications already tested: by integrating applications that already have been tested for that purposed, the only testing that will have to be done is of the integrated system. The application by itself has already been tested. That's why it's important to choose application with high number of downloads and comments.

On evaluating and choosing the open source application to use, the documentation assumed a primordial role. Open source applications tend to have a problem of insufficient documentation, that can be minimized by recurring to a commercial company that gives support to that open source application [12].

If companies share their experiences on the open source integration, as suggested by Janne Merilinna and Mari Matinlassi article [12], it will make the choice of application to use easier.

The developed system is available to download, as a virtual machine (.img), on http://sourceforge.net/vicispromus.

References

1. Web Server Survey | Netcraft, http://news.netcraft.com/archives/category/web-server-survey/
2. Fauscette, M., Shirer, M.: Open Source Software Market Accelerated by Economy and Increased Acceptance From Enterprise Buyers, IDC Finds | Business Wire, http://www.businesswire.com/news/home/20090729005107/en/Open-Source-Software-Market-Accelerated-Economy-Increased
3. Johannesson, P., Perjons, E.: Design principles for process modelling in enterprise application integration 26, 165–184 (2001)
4. Silva, M.M.: Integação de Sistemas de Informação. FCA - Editora de Informática LDA, Lisboa (2003)
5. Bagheri, O.R., Nasiri, R., Peyravi, M.H., Dehkordi, P.K.: Toward an elastic service based framework for Enterprise Application Integration. Presented at the Fifth International Conference on Software Engineering Research, Management and Applications (2007)
6. OpenConf Portfolio, http://www.openconf.com/portfolio/
7. Erasala, N., Yen, D.C., Rajkumar, T.M.: Enterprise Application Integration in the electronic commerce world. Computer Standards & Interfaces 25, 69–82 (2003)
8. Talend: The Return on Investment of Open Source Data Integration (2010)
9. Top 5 Browsers from December 2011 to December 2012 | StatCounter Global Stats, http://gs.statcounter.com/
10. Serrão, C., Gamito, M.: Comparação de software Open-Source para Gestão de Eventos de natureza Cientifica. Presented at the CAPSI (2011)
11. CMS Market Share – OpenSourceCMS, http://www.opensourcecms.com/general/cms-marketshare.php

12. Merilinna, J., Matinlassi, M.: State of the art and practice of open source component integration. Presented at the 32nd EUROMICRO Conference on Software Engineering and Advanced Applications (2006)
13. Allen, J.P.: Open Source Deployment at the City and County of San Francisco: From Cost Reduction to Rapid Innovation. Presented at the 43rd Hawaii International Conference on System Sciences (2010)
14. Papagelis, M., Plexousakis, D., Nikolaou, P.N.: CONFIOUS: Managing the Electronic Submission and Reviewing Process of Scientific Conferences. In: Ngu, A.H.H., Kitsuregawa, M., Neuhold, E.J., Chung, J.-Y., Sheng, Q.Z. (eds.) WISE 2005. LNCS, vol. 3806, pp. 711–720. Springer, Heidelberg (2005)
15. OpenConf Editions, `http://www.openconf.com/editions/`
16. EasyChair: frequently asked questions, `http://www.easychair.org/faq.cgi`
17. Jain, M., Tewari, T.K., Singh, S.K.: Survey of Conference Management Systems. International Journal of Computer Applications 2 (2010)
18. Pesenhofer, A., Mayer, R., Rauber, A.: Improving Scientific Conferences by enhancing Conference Management System with information mining capabilities

A Tool for Fractal Component Based Applications Performance Modelling Using Stochastic Well Formed Nets

Nabila Salmi[1,2], Malika Ioualalen[1], Smail Lallali[1], and Hamza Zerguine[1]

[1] LSI - USTHB - BP 32, El-Alia, Bab-Ezzouar, 16111 - Alger, Algérie
{salmi,ioualalen}@lsi-usthb.dz
[2] LISTIC - Université de Savoie - BP 80449, 74944 - Annecy le Vieux, France

Abstract. Today, performance prediction of component-based systems is important to help software engineers to analyze their applications in early stages of the development life-cycle, so that performance problems are avoided. To achieve performance prediction, modelling is a crucial step. It would be interesting if component performance models can be derived automatically. To this aim, we describe in this paper a software toolset which allows component designers of specific systems, that are Fractal systems, to generate performance models, starting from the Fractal architectural description of their system and component behaviours. These models consist of Stochastic Well formed Nets (SWN) and Stochastic Petri nets (SPN), and can be analyzed using SPN/SWN analysis tools. A case study illustrates the effectiveness of our approach.

Keywords: Fractal model, Component Based Systems, performance modelling, SPN, SWN.

1 Introduction

Since the last decades, software engineers are more and more interested in non-functional properties, such as performance, quality of service or scalability, at an early life-cycle stage. This allows to avoid costly changes if a problem is detected in late development stages. Especially, predicting performance of component based systems gets more and more important to build trustworthy systems. In this direction, our long-term goal is to provide component designers with tools allowing automatic generation of performance models for a component based system (CBS), to be used for its automatic quantitative analysis.

In this context of CBS performance prediction based on performance modelling, several approaches were proposed, in the last years, summarized in [1]. Two classes of techniques were described: those based on the Unified Modelling Language (UML) and others based on proprietary meta-models. In UML-based approaches, several work proposed to derive from UML statecharts and diagrams various models, such as queueing networks [2], High level Petri nets [3], stochastic Petri nets [4], etc. However, these proposals didn't address CBS. In

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 773–784.
DOI: 10.1007/978-3-642-36981-0_72

the context of CBS, [5] proposed to translate UML component diagrams into Layered Queuing Networks (LQN). In the second class of approaches, we cite the Component-Based Modelling Language (CBML), defined as an extension to LQNs by [6]; the Prediction Enabled Component Technology (PECT), the Palladio component Model (PCM) [7], a domain-specific modeling language for CBS, enabling performance predictions [8]. However, these technics are not generic as they are restricted to specific meta-models. In this paper, we introduce a new tool allowing formal modelling of specific component based applications, FRACTAL systems, to be used later in automatic a priori performance analysis. The FRACTAL model [9] has been chosen to be studied for its genericity: It is an open academic model which offers a hierarchical and reflective component model with dynamic configuration, component composition, management and sharing capabilities.

Formal modelling of FRACTAL systems has been already studied in our previous work [10,11], where the objective was to provide an efficient "a priori" component based systems (CBS) analysis method for performance prediction. In this paper, we provide a tool implementing an improved version of our modelling method with automatic generation of performance models. The benefits of the tool is to allow conducting automatic performance analysis of FRACTAL component based applications, thanks to generated performance models.

We start from the FRACTAL architecture description of the CBS under study provided in an ADL and component behaviours, generate models for each component, and finally deduce a global model for the CBS by composing obtained component models. The obtained models are Stochastic Petri nets (SPN) and Stochastic Well-formed Nets (SWN), a special class of Stochastic coloured Petri Nets, widely used for performance analysis of complex symmetrical systems. These formalisms are motivated by several benefits: first, we need a state based model to carry out performance evaluation; Stochastic Petri Nets are state based models, well known for being able to model complex systems with concurrency and conflicts, in contrast with Queuing networks or process algebras for instance. Besides, nowadays software systems, particularly CBS, may be characterized by symmetrical component behaviors (several threads or resources of the same component). In this case, the SWN model is well suited as it can take advantage of behavioral symmetries of system's entities It is also a well studied class of high level stochastic Petri nets and benefit from a large set of analysis algorithms and tools. Finally, it allows scalability and analysis of a huge number of states. SPN and SWN models are provided by the tool in XML format and in the proprietary format proper to the GreatSPN tool [12], for allowing direct analysis using this latter.

The structure of the paper is as follows. We present in section 2 the main features of the FRACTAL model. We also recall in section 3 our approach for generating component formal models and building the global model of a FRACTAL application. The tool is presented in section 4 and illustrated in section 5. We conclude and give future works in section 6.

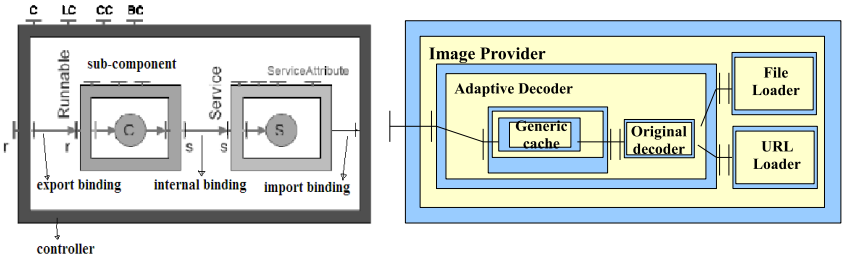


Fig. 1. A FRACTAL composite component(left) and a FRACTAL application (right)

2 The FRACTAL Component Model

FRACTAL [9] is a general component model developed within the consortium ObjectWeb, for implementing, monitoring and dynamically configuring complex software systems. In this model, a component is a *runtime* entity that interacts with its environment (i.e. other components) through well-defined *interfaces* (figure 1). An interface is an access point to a component, that specifies provided services or required services exposed by other components. There are two kinds of interfaces: *server* interfaces correspond to points accepting incoming operation invocations, and *client* interfaces which support outgoing operation invocations.

A FRACTAL component possesses two parts : a *content part* and a *controller part*. The content part consists of a finite number of other components, called *sub-components*, making the model recursive and allowing components to be nested at an arbitrary level. At the lowest level, a FRACTAL component is a black box, called *base* or *primitive* component, that doesn't provide introspection or intercession capabilities. A FRACTAL component whose content is not empty is said *composite*. The controller part, termed *the membrane*, provides a set of *control* interfaces, supporting introspection (monitoring) and reconfiguration of internal features of the component, such as suspending and resuming activities of a sub-component.

To define the architecture of an application, FRACTAL components are connected by *bindings* (see figure 1, bottom). To define component architectures for the FRACTAL model, an open and extensible language has been developed : the FRACTAL Architecture Description Language (ADL), which is XML based.

3 Modelling Fractal CBS with SPN/SWN

From the main characteristics of the FRACTAL model, we derived a generic approach for automatic model generation of FRACTAL applications [10,11]. We recall in this section this approach. The proposed modelling is based on the Stochastic Petri Net (SPN) and Stochastic Well-Formed Petri Net (SWN) models; the SWN model being a high level (coloured) model of Petri net with probabilistic extensions for performance analysis.

3.1 The Well-Formed and Stochastic Well-Formed Petri Nets

A well-formed (WN) net [13] is a high level (coloured) Petri net model, where places and transitions are provided with a structured type of tokens. In this model, tokens are grouped into basic classes called *colour classes*. These classes are grouped together to form a colour domain, which is associated to places and transitions. Colours of a place label its tokens, whereas colours of a transition define possible firings of the transition. Thus, an initial marking of a place is defined as a multiset (bag) of coloured tokens. A colour function is attached to each arc: its role is to define for, a given colour of the associated transition, the number of coloured tokens to add or to remove from the attached place. A colour domain (denoted by *D(N)*, where N is a place or transition) is a Cartesian product of colour classes.

A colour function is built from standard operations (linear combination, composition, etc) of basic functions. The projection selects an element of a tuple; it is represented by a typed variable or by $< X >$ if no confusion is possible. The synchronization/diffusion returns the set of all colours of a class (Si) or a sub-class (Si, k). The successor function is defined for ordered classes only and returns the colour following a given colour. A transition or an arc function can be guarded by an expression, which is a linear combination of atomic predicates. A predicate is evaluated on colours of a transition firing.

The structured definition of a WN allows us to exploit automatically system symmetries, by compacting its reachability graph, leading to a *Symbolic Reachability Graph* (SRG) [13]. An SRG is composed of *symbolic markings*, where each symbolic marking represents a set of ordinary (coloured) markings having equivalent behaviours. Several qualitative properties can be checked on the SRG (reachability of a marking, deadlock freeness, etc.) From WNs was derived the *stochastic well-formed* (SWN) model, which associates to each transition an exponentially distributed delay. The SRG of an SWN, augmented with stochastic firings information, results in an aggregated Markov chain of the chain derived from the coloured net. Thus, we can study performances of a system directly on this aggregated chain. (for formal definitions of WN and SWN, see [13]).

3.2 Generating SPN and SWN Models

To model a Fractal application, we assume that it is defined through an ADL description and a set of Java classes corresponding to primitive components. We need to define how to model the main elements composing a Fractal component. These elements are deduced from the following alternatives :

- A component can have a local behavior (set of internal actions), or
- It can invoke services from other components through a client interface, or
- It can process a client request submitted to one of its server interfaces.

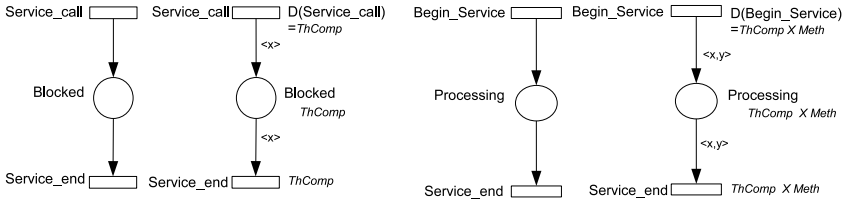
Thus, basic models for interfaces and internal component behavior were proposed in our approach. Then, an SPN/SWN model is built for each primitive component. Finally, a general algorithm generates the Fractal application model, using previously generated SPN/SWN component models.

To model with the SWN formalism, the following colour classes were defined:

- A component is characterized by a set of threads giving its execution flow. To represent them, we use a colour class denoted *ThComp* : $ThComp = \bigcup_{i=1}^{n_c} c_i$ where c_i is a thread colour.
- If the component has server interfaces, we need to specify which service or method is invoked. Thus, a second colour class, denoted *Meth* is used : $Meth = \bigcup_{i=1}^{n_m} m_i$, where m_i is a method colour.

The n_c cardinality of a thread colour class represents the maximum number of threads of the corresponding component and are is defined manually in the tool. The n_m cardinality of a method colour class is the number of different methods available on a service interface. This cardinality is deduced from the method definition given by either the Interface definition language (IDL) or the implementation code (Java, C or other) of the analyzed system.

Modelling Interfaces



(a) client SPN(left)/SWN(right) (b) Server SPN(left)/SWN(right)

Fig. 2. Client and Server interface SPN (left) and SWN (right) models

Client interface. A component with a client interface, in the JULIA implementation, invokes a service offered by another component and waits for the end of service. Consequently, a client interface is modelled with one place *Blocked*, indicating the Waiting state of component and two transitions : *Service_call* for the service invocation and *Service_end* for the end of service. The obtained model is an SPN sketched in the left side of figure 2a.

In case of generating an SWN model, places and transitions should be coloured. As the client interface is triggered by a thread, we colour each node N in the SPN model with the *ThComp* domain : $D(N) = ThComp$ such that $N \in \{Blocked, Service\_call, Service\_end\}$. Moreover, arcs are labelled with an arc function which is the projection X(c)=c (see the right side of figure 2a).

Server Interface. A component with a server interface offers a service for other components. Thus, a server interface is modelled with one place *Processing*, indicating the service processing state and two transitions : *Begin_Service* for

launching the service and *Service_end* for the end of service. The obtained model is an SPN sketched in figure 2b left. In case of generating an SWN model, as the server interface is triggered by a thread representing the server component and a method, we colour each node N in the SPN model as follows :

$$D(N) = ThComp \times Meth,\ N \in \{Processing,\ Begin\_Service, Service\_end\}\ .$$

We use as arc function the projection of a tuple $< c, m >$(see figure 2b right).

Note that these are high level abstracted models, where we don't care about service processing details. When the component developer need to evaluate such details impact on the component, we are faced to two cases:

- If the processing details don't invoke other component services, we replace the *Begin_Service* transition with a subnet modelling these details. Otherwise, we can model implicitly such case by estimating the time needed for service processing and using it to fix transition rates.
- However, when the service processing needs invocation of other component services, we update our model by linking component models to obtain the global application model.

Internal component behavior. A component with an internal behavior performs a set of serial actions. So, we model this set of actions with a model consisting of two transitions *Begin_int_Bloc*, *End_int_Bloc* modelling the beginning and end of this behavior, and one place *Execution* expressing execution of internal instructions. Note that we abstract internal details of the local behavior. When the designer is interested in internal details, each important activity in these details is modelled by a transition.

To obtain an SWN model, as for the client interface, we colour each node in the previous model with the *ThComp* domain :
D(Begin_int_Bloc)=D(End_int_Bloc)=D(Execution)= ThComp.

In some cases, a component may be without internal behavior. In this case, the corresponding SPN/SWN model is not added.

Modelling Primitive Components. During its lifecycle, a component can perform a local behavior, invoke client requests or process service requests. Depending on these cases, we model a component with several composed parts:

1. The 1*st* part models the component execution flow by a place *Component.*
2. If the component has a local behavior, we add, as a second part, the internal behavior model by connecting its two transitions to the *Component* place : this place is a precondition to the *Begin_int_Bloc* transition, and a postcondition to the *End_int_Bloc* transition.
3. The third part to add is the client interface model, whenever the component has a client interface. This part is being duplicated when several client interfaces are exhibited, with a renaming of the nodes.
4. Finally, whenever the component offers one or several server interfaces, we add first a place *Method* to model possible methods invoked by clients. Then, we add the server interface model, while linking the *Component* and *Method*

places as preconditions to the *Begin_service* transition, and postconditions to the *Service_end* transition. This interface model is also duplicated (and corresponding nodes renamed) if several server interfaces are exhibited.

We obtain an SPN model. Note that recursive interface invocation is not handled in this paper and should be modelled otherwise. To build an SWN, we colour the *Component* place with the *ThComp* class and the *Method* place with the *Meth* class. The other nodes are coloured as explained in section 3.2. An example of a component model having a local behavior, one client interface and one server interface is given in figure 3.

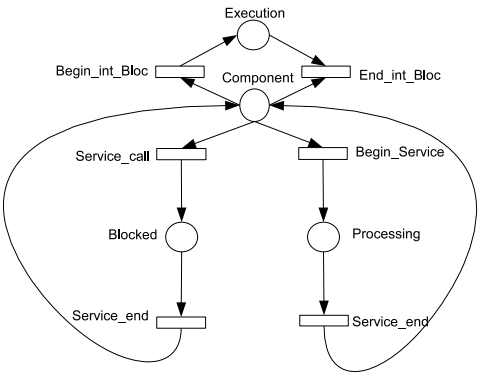


Fig. 3. A component SPN model (with local behavior, client/server interfaces)

Generating a CBS SPN/ SWN Global Model. To generate an SPN/SWN model of a FRACTAL application, our method starts from its ADL description. The FRACTAL application may be composed of a finite set of composite and primitive components. Modelling this application requires to know invoked activities at execution, which means going through composite components until reaching primitive components and retrieving basic actions to perform. Consequently, before modelling, we need to explore the ADL description, extract a view made only of primitive components and bindings between them. This step is rather delicate to realize, because bindings defined at a high level of the components hierarchy should be moved at the lowest level.

Hence, modelling a FRACTAL application is realized through three steps:

1. The first step consists of flattening the hierarchical ADL description, and keeping only primitive components and their bindings. This step is done only for modelling a given component system configuration, even if we loose hierarchy, as the goal is to analyze the model of the configuration and analysis with SPN/SWN tools does'nt need to keep hierarchy. We call this step the *flattening* step.

2. The second step consists of modelling each primitive component.
3. The final step generates a global model for the application by merging models of primitive components, following their bindings. The merging consists of fusing transitions corresponding to interfaces (for more details see [10]).

Note that the global net is completed with marked places modelling application threads so that the resulted state space is finite.

4 The *FRACMOD* Tool

The automatic model generation of SPN/SWN models for a Fractal application was implemented in a toolset *FRACMOD*, using the Java language.

The *FRACMOD* tool allows Fractal designers to :

- Read an ADL description of a Fractal application and deduce its primitive components with their bindings.
- Automatically generate, from the ADL description, SPN and SWN models for primitive components of the application.
- Automatically generate the global SPN/SWN model of the Fractal application described through the ADL description.

4.1 Software Architecture

In this section, we describe the architecture of our toolset. Five modules constitute the toolset (see Figure 4):

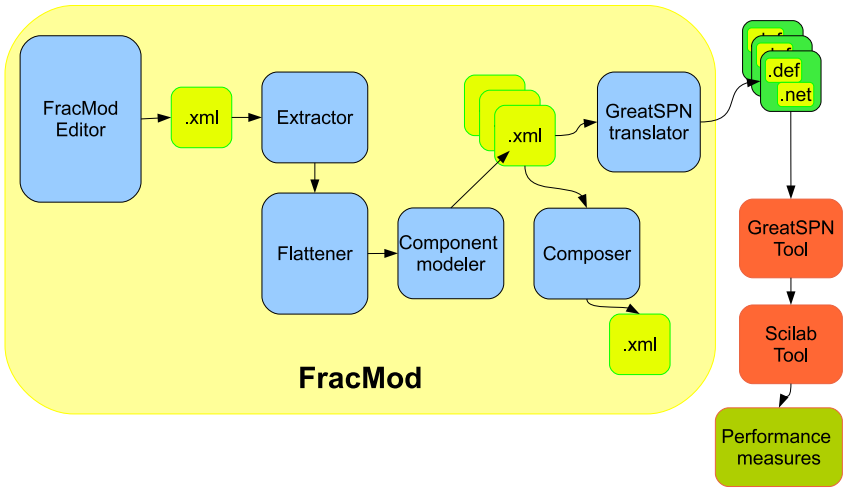


Fig. 4. Software architecture of the *FRACMOD* tool

- The FRACTAL *editor* allows users to introduce the architecture description of a FRACTAL application into an XML format.
- The *extractor* module reads the introduced XML description of the FRACTAL application, parses it and extracts composite and primitive components with their characteristics. It also extracts bindings between the different components.
- The *flattener* module flattenes the hierarchical application ADL description and finds out the list of primitive components and bindings linking them.
- The *component modeler* generates SPN and SWN models for each primitive component and displays them. As SPN/SWNs are timed models, we need to input timing information which should be associated to transitions of the generated models. We need also to input other specific data, mainly thread and services (methods) classes cardinalities. In fact, this information should be provided by the CBS developer which estimates processing rates (speed) of service methods and the number of threads and services described in component interfaces, and input them through this module. Processing rates can be estimated from preliminary performance information provided by hardware specifications and operating systems performance monitors.
- The *composer* module gets the SPN/SWN component models already generated and primitive component bindings, then builds a global SPN/SWN model for the FRACTAL application, by composing component models according to their bindings. Composition between models is done by fusing transitions modelling client and service bound interfaces.
- Finally, the *GreatSPN translator* translates SPN/SWN generated component models into the GreatSPN Petri net format, which is a proprietary format given by two files (.def and .net files). It also translate the global application model into the def and net formats. Generated models are then analyzed with the Greatspn tool [12] and the *Scilab* environment(http://www.scilab.org/).

5 A Simple Example

We illustrate our modelling using a concrete example, a minimal image viewer/ browser [14]. At its core, there is an adaptive image decoder component whose role is to interpret the content of an image file (JPEG, PNG...) into a bitmap which can be printed on a screen. Next, a file loader component is used. Another component is also used to load the content of the image files from a source location (local or remote). The adaptive decoder (the front end component) is composed of a generic caching component and an original decoder. Figure 1, right, shows the corresponding part of the application architecture, coded using FRACTALḂasic colour classes used for modelling components of the image viewer example model decoding requests, files, and even component threads.

To illustrate model generation, we used *FRACMOD* for the Image viewer application to generate SWN models. The application is a composite component, built from four components: the cache, the decoder, the file and url components, extracted by the tool from the application ADL description. We input necessary

information for SWN models (colour classes cardinalities and transition rates). So, the *component modeler* module generates the SWN model of each component, with two static colour classes, one for modelling component threads and the other for offered component methods (CMet and CThr for the Cache methods and threads, DMet and DThr for the decoder methods and threads, FMet and FThr for the file component methods and threads and UMet and UThr for the url component methods and threads). Some of the obtained SWN models are depicted in figures 5. Next, the global SWN of the whole application is built.

To compute performances of the image viewer/browser application, we used the GreatSPN environment on the generated models, with Scilab computations, on a Suse linux 11.0 workstation with 2 GB of RAM. We were interested in studying the response time for a decoding request submitted to the application. For that purpose, we take fixed rate values of a critical set of transitions, and vary the decoding service rate (rate of the *Beg_serv_decod* transition). Then, we study the evolution of response time from obtained steady-state probabilities. Main transitions rate values are given in table given in figure 6, left. Transitions not appearing in this table have rate 1 (i.e. faster than all other transitions, rates being given in the same unit). We present results for some colour cardinalities of a system configuration. We obtained the diagram of the right side of figure 6.

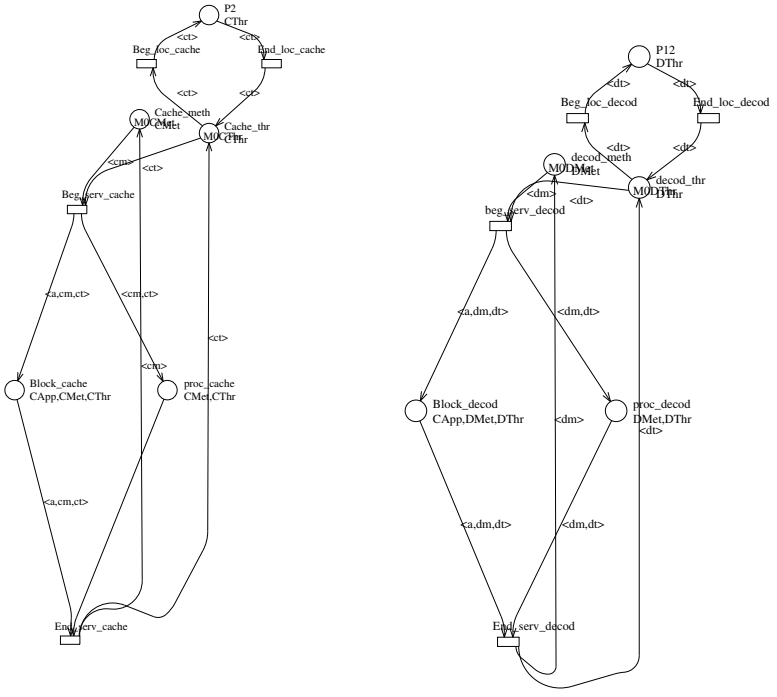


Fig. 5. The cache (left) and decoder (right) SWN models

Component	Transition	Rate
Cache	Beg_serv_cache	0.9
Decoder	Beg_loc_cache	0.6
Decoder	Beg_serv_decod	0.9
Decoder	Beg_loc_decod	0.6
File	Beg_serv_file	0.85
File	Beg_loc_file	0.5
Url	Beg_serv_url	0.85
Url	Beg_loc_url	0.5

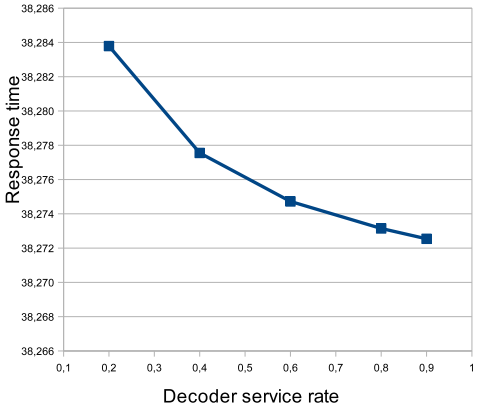


Fig. 6. Transition rates (left) and Response time versus decoder service rate (right)

In this diagram, we can see that the image viewer application presents a slightly better response time as far as the decoder service rate increases. This is expected since the system becomes faster with higher decoder service rate.

6 Conclusion

This paper prsents the *FRACMOD* tool for automated modelling of a Fractal application, using stochastic well formed Petri net models. The goal of automating modelling is to allow designers, not experts in modelling, to get performance models for their applications, and enable them to analyze and predict their performances, basing on the knowledge of its ADL description and component behaviors. We experimented the *FRACMOD* tool on several Fractal applications and the first results are promising. What remains is to plug *FRACMOD* to an analysis tool (GreatSPN) to provide automatically performance indices. However, a minor drawback is that flatenning the CBS architecture makes somewhat difficult to get feedback to the final user. Besides, still more research work is required in several directions: integrating the *FRACMOD* tool in a global modelling and analysis tool, starting from the ADL description and automatic modelling, and resulting in performance computations, given specification of performance indices of interest. We also target to use the automated modelling of primitive components in a compositional analysis step, based on components models. This can be done thanks to our previous work [10], which defined a structured performance analysis method for CBS, allowing to reduce computation times and memory usage, basing on component models. Finally, we work on modelling and analysis of reconfiguration features of Fractal CBSs.

References

1. Koziolek, H.: Performance evaluation of component-based software systems: A survey. Performance Evaluation, Special Issue on Soft. and Perf (August 2010)
2. Gu, G.P., Petriu, D.C.: From uml to lqn by xml algebra-based model transformations. In: WOSP, pp. 99–110 (2005)
3. Baresi, L., Pezzé, M.: On Formalizing UML with High-Level Petri Nets. In: Agha, G., De Cindio, F., Rozenberg, G. (eds.) APN 2001. LNCS, vol. 2001, pp. 276–304. Springer, Heidelberg (2001)
4. Bernardi, S., Donatelli, S., Merseguer, J.: From UML sequence diagrams and statecharts to analysable petri net models. In: Proc. of 3rd WOSP 2002, Rome, Italy, pp. 35–45 (July 2002)
5. Grassi, V., Mirandola, R., Sabetta, A.: Filling the gap between design and performance/reliability models of component-based systems: A model-driven approach. J. Syst. Softw. 80(4), 528–558 (2007)
6. Wu, X., Woodside, M.: Performance modeling from software components. SIGSOFT Softw. Eng. Notes 29(1), 290–301 (2004)
7. Becker, S., Koziolek, H., Reussner, R.: Model-based Performance Prediction with the Palladio Component Model. In: WOSP 2007. ACM, Sigsoft (2007)
8. Meier, P., Kounev, S., Koziolek, H.: Automated transformation of cbs architecture models to queueing petri nets. In: MASCOTS, pp. 339–348 (2011)
9. Bruneton, E., Coupaye, T., Stefani, J.: The fractal component model, version 2.0-3. Technical report, Fractal team (February 2004)
10. Salmi, N., Moreaux, P., Ioualalen, M.: Performance evaluation of fractal component based systems. Annals of Telecommunications. Special Issue: Software Component: The Fractal Initiative 64(1), 81–101 (2009)
11. Salmi, N., Moreaux, P., Ioualalen, M.: Structured performance analysis for component based systems (IJCCBS) - Special Issue on Verification and Evaluation of Computer and Communication Systems - Part II - Issue 1/2 3(1), 96–131 (2012)
12. Perf. Eval. Group: GreatSPN home page, `http://www.di.unito.it/~greatspn`
13. Chiola, G., Dutheillet, C., Franceschinis, G., Haddad, S.: Stochastic well-formed colored nets and symmetric modeling applications. IEEE Trans. on Comp. 42(11), 1343–1360 (1993)
14. David, P.-C., Ledoux, T.: Towards a Framework for Self-adaptive Component-Based Applications. In: Stefani, J.-B., Demeure, I., Zhang, J. (eds.) DAIS 2003. LNCS, vol. 2893, pp. 1–14. Springer, Heidelberg (2003)

Open Source Software Documentation Mining for Quality Assessment

Nuno Ramos Carvalho[1], Alberto Simões[2], and José João Almeida[1]

[1] Departamento de Informática, Universidade do Minho
{narcarvalho,jj}@di.uminho.pt
[2] Centro de Estudos Humanísticos, Universidade do Minho
ambs@ilch.uminho.pt

Abstract. Besides source code, the fundamental source of information about Open Source Software lies in documentation, and other non source code files, like *README*, *INSTALL*, or *HowTo* files, commonly available in the software ecosystem. These documents, written in natural language, provide valuable information during the software development stage, but also in future maintenance and evolution tasks.

DMOSS[1] is a toolkit designed to systematically assess the quality of non source code text found in software packages. The toolkit handles a package as an attribute tree, and performs several tree traverse algorithms through a set of plugins, specialized in retrieving specific metrics from text, gathering information about the software. These metrics are later used to infer knowledge about the software, and composed together to build reports that assess the quality of specific features of the software.

This paper discusses the motivations for this work, continues with a description of the toolkit implementation and design goals. Follows an example of its usage to process a software package, and the produced report. Finally some final remarks and trends for future work are presented.

1 Introduction

Open Source Software (OOS) wide spread adoption, including in the industry, has raised increased concerns related with software quality and certification [3]. In this context, the CROSS research project[2] aims at developing software analysis techniques that can be combined to assess open source software projects. Although most of the effort is spent analyzing source code, non-source code content found in packages can have a direct impact on the overall quality of the software. For example documentation, installation procedures, practical information available in *README* files etc. The goal of the DMOSS toolkit is to provide a systematic approach to gather metrics about this content and assess

[1] Documentation Mining Open Source Software.
[2] An Infrastructure for Certification and Re-engineering of Open Source Software: http://twiki.di.uminho.pt/twiki/bin/view/Research/CROSS/WebHome

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 785–794.
DOI: 10.1007/978-3-642-36981-0_73

its quality. It starts by gathering content in the package written in natural language, processing this content to compute metrics, and finally reasoning about these metrics to draw conclusions about the software quality.

Documentation analysis is also relevant in other research areas. Program Comprehension (PC) is an area of Software Engineering concerned with gathering information and provide knowledge about software to help programmers understand how a program works in order to ease software evolution and maintenance tasks [6]. Many of the techniques and methods used rely on mappings between program elements and the real world concepts these elements are addressing [7]. Non-source code content included in software packages can provide clues and valuable information to enhance the creation of these mappings. Program maintainers often rely on documentation to understand some key aspects of the software [9].

Assessing software quality for any given definition of quality is not easy [5] mainly due to subjectivity. The toolkit described in this works evaluates the non-source code files included in a software package. This set of files can include *README* files, *INSTALL* files, HTML (HyperText Markup Language) documentation pages, or even UNIX man(ual) pages. Instead of trying to come up with a definition for quality, we select three main traits that we are concerned about. We envisage that these characteristics have a direct impact in the overall documentation quality regardless of the degree of individual subjectivity.

- Readability: text readability can be subjective, but there are linguistic characteristics that generally make it harder to read. Some of them can even be measured, as for example, the number of syntax errors or the excessive use of abbreviations;
- Actuality: this is an important feature of documentation and other textual files, they should be up-to-date, and refer to the latest version of the software;
- Completeness: this trait tells us how much the documentation is complete, and if it addressees all the required topics.

DMOSS processes a software package to gather information about specific metrics that are related with these traits. Reasoning about these metrics helps drawing conclusions relevant to assess the described traits. Based on these conclusions a quantitative measure can be calculated about the quality of the non-source code content.

The next section of this article discusses some related work in this area. Section 3 introduces the DMOSS toolkit and gives an overview about its implementation details. This section also illustrates the major algorithms used. Section 4 presents a quick tour about using the toolkit, and example of generated reports. Finally, Section 5 concludes with some final remarks and discusses some trends for future work.

During the remainder of this paper the software package *tree*[3] (version 1.5.3) will be used for illustration purposes, mainly because it is small and produces outputs that can fit in the paper size without jeopardize reading.

[3] Available from `http://mama.indstate.edu/users/ice/tree/`

2 Related Work

Forward *et al* [2], in their survey about the general opinion of software professionals regarding the relevance of documentation and related tools, highlight the general consensus that documentation content is relevant and important. They also highlight a set of concerns that software documentation technologies should be more aware of professionals' requirements, opposed to blindly enforce documentation formats or tools.

Scacchi [8], in his work about the requirements for open source software development, highlights not just the relevance of system documentation, but also the relevance of informal documents (for example, *How-Tos*). They are significant not only for documenting the system itself, but also to communicate important information for other people in the community (for example, how to contribute for the project).

There is a substantial body of work which illustrates the relevance of documentation quality in the context of software development and maintenance. Chen *et al* [1] have identified documentation quality problems is a dimension by itself, and a key problem factor that affects software maintenance phase. Nevertheless, the literature is sparse when describing metrics and methods for evaluating non-source code content for software quality assessment. This work focus on addressing this problem.

3 DMOSS Toolkit

The DMOSS toolkit main goal is to provide a set of tools that systematically process a software package and produce a final report with conclusions about the quality of the non-source content found in the package. This includes analyzing all the natural language text stored in the documentation, comments in the code, and other non-source code files typically found in packages.

The main design goals for DMOSS are:

- Develop small tools that can be used by themselves, so that they can be useful in other contexts or environments. Higher order applications in the toolkit (applications that use the smaller tools) need to be modular so that new tools can be added without any effort (just like typical plugins).
- Many tools in DMOSS take advantage of known algorithms and techniques (for example the file *processors*). The main engine in the toolkit needs to be based on the usage of plugins, so that new processors and similar utilities can be added and improved easily.
- Look at the software package being analyzed as a tree. This allows the implementation of the analysis algorithms as a set of tree traversals. This keeps the implementation of the specific analysis algorithms self-contained.

Let us stress again the importance regarding the way DMOSS represents a software package: an annotated tree. In this tree, nodes represent files and directories, and edges describe the hierarchical structure of the package. An example tree is illustrated in Fig. 1, for the `tree` software package.

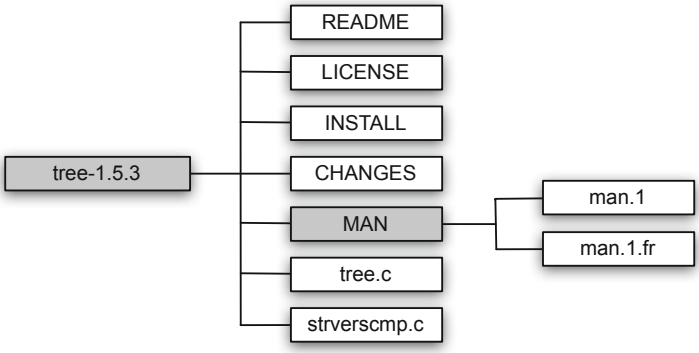


Fig. 1. Package tree like structure

Once this tree is automatically generated, the task of processing a software package is divided in two tree traversals:

1. During the first pass the goal is to gather informations about what files exist and what is their content. Each plugin is also run for each file individually, and the computed metrics are stored in the tree as node attributes;
2. In the second pass, results are aggregated. For each directory node the available metrics are reduced to a single result, and in the end, the root node (the package top directory) will contain the results of processing the entire package.

The plugins that perform the actual analysis and build conclusions need to implement three functions to be used in both tree traversals:

1. A *processor*, which is responsible for gathering information about a file and produce a set of features (a metric can be measured using one or more features) about its content. These features are stored in the tree as node attributes:
$$processor :: Node \longrightarrow [Feature]$$
2. A *reducer*, which is responsible for reducing features to produce either intermediate or final results. Results can be a single feature or a set of features:
$$reducer :: [Feature] \longrightarrow [Feature]$$
3. Finally, a *reporter*, which is responsible for building the final report given a set of features:
$$reporter :: [Feature] \longrightarrow Report$$

The only strictly required function is the *processor*, as there are default implementations for the other two functions, which are used when a plugin does not provide them. The default *reducer* reduces attributes using string concatenation or arithmetic sum depending on value type. The default *reporter* uses a pre-defined template to produce a simple report.

A feature is defined as a pair, consisting of a name and a value:

$$Feature = Name \times Value$$

where, $Name$ is the attribute identifier — a string — and $Value$ can be an atomic value (a string or number for example), or a structured set of more $Feature$s for storing complex data structures.

A node in the tree is defined as:

$$Node = Path \times isFile \times Text$$

where, $Path$ stores the file name and its path, $isFile$ is a boolean value stating if this node is a file or a directory, and $Text$ stores the natural language text found in the file. The $Text$ value is computed before starting the tree traversal stages.

3.1 First Pass: Gathering Information

When traversing the tree, each file node is processed, i.e., the files represented by each node are processed . These nodes are processed in two steps:

1. Determine the file type, either using its full media type [4], or using heuristics, like the file header or extension. The result of this step is the creation of an attribute named *type* with the corresponding file type (for example *plain/text*, *text/xml* or *text/html*) as its value.
2. Given the node *type* and a list of available *processors* for each file type[4] the next step is to process the current file with all the available processors that support it, and store each processor result as a new node attribute.

This workflow is executed in every single node that represents a file, Algorithm 1 illustrates it. The final result is a tree with a set of metrics calculated for each node file and stored as attributes (including the file type).

Algorithm 1. Decorate tree with *processor*s results

```
Input: tree : Tree representing package content.
Input: processors : Set of processors indexed by type.
Result: Tree after adding processors resulting features.
for node ← tree : node.isFile == True do
    for proc ← processors(node.type) do
        // Add processor resulting feature list to node
        node.push(proc(node))
return tree
```

[4] The toolkit provides a set of plugins that implement several *processors*, and new plugins can be easily added.

Processors. In order to compute attributes values for file nodes, the toolkit provides a heterogeneous set of processors. Each processor typically handles a single file, and produces a result that is stored as an attribute in the tree. For example, the spell checker processor computes the total number of words in a text file, and the total number of words found in the dictionary (see Algorithm 2).

Algorithm 2. Processor Example: Spell Checker

```
Input: node : Node
Result: New feature list to be added to the node.
total ← 0
found ← 0
for word ← splitWords(node.text) do
    if dictionary.valid(word) then
        found++
    total++
f = Feature(attr = "spellChecker", value = (total, found))
return [f]
```

New processors can be added or plugged in at anytime. Built in the toolkit there is also a table that states which file types can be handled by each processor. This helps to keep the traversing tree engine agnostic to which processors are available, and which files to process.

3.2 Second Pass: Reducing Results

The goal of the second tree traversal is to produce the final metrics results. This is achieved by combining the intermediate results for every level of the package tree, and adding new attributes (typically to the directories nodes) that store the result of combining the results for each sub-tree (and for each specific metrics). Every plugin may provide a specific function to combine results. The default method for combining intermediate results is plain string concatenation, or arithmetic addition (depending on value type).

For example, the combining function for the spell checker processor is to add the total number of words, and the total number of words not found for the files on each directory. This means that after this pass, the `MAN` node (illustrated in figure 1, which represents the filesystem `man/` directory) will have an attribute that stores the result of combining the spell checker processor result for files `man.1` and `man.1.fr`[5]. Later, this attribute value will be used to calculate the total result for the package, stored in the top level directory.

Figure 2 illustrates this process for an arbitrary metric. The algorithm is also described in algorithm 3.

[5] Although the two files are written in different languages the plugin uses a language identification algorithm before the spell checking task.

Algorithm 3. Reduce Package Tree

Input: tree : Tree representing package content.
Input: reducers : Set of available reducers.
Result: Tree after adding reducers results to nodes as features.
for $node \leftarrow tree : node \in Directories$ **do**
  for $reduce \leftarrow reducers$ **do**
    $node.push(reduce(node.children))$
return *tree*

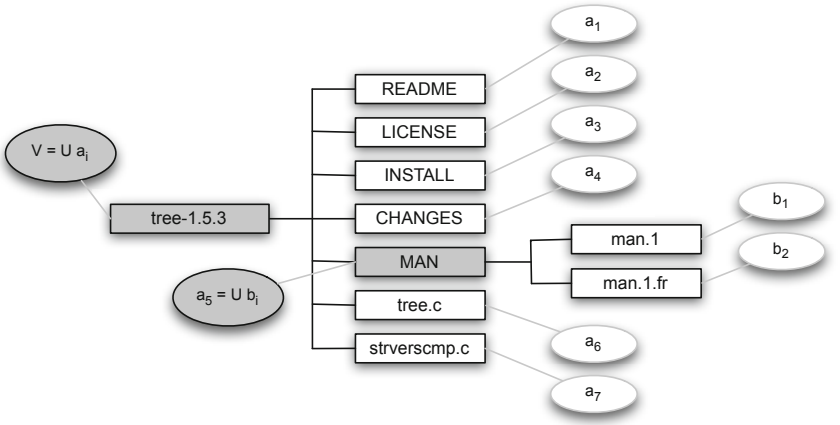


Fig. 2. Calculate final values recursively

Reducers. These functions are used to reduce intermediate results, *i.e.*, combine the results found by the processors in the sub-tree of the node currently being processed, and add this reduced result to the current node as a new attribute. Algorithm 4 illustrates the reducer for the spell checker example.

3.3 Building Reports

After the package is processed a tree representing the package is available. This tree is decorated with a set of attributes per node, that confine all the results gathered from processing each file node, and also the conclusions taken for each processor. This information is stored in the tree using attributes. The toolkit provides a tool that can build reports in several formats including HTML, and ontology style graphs in GraphViz[6] notation. An example of a HTML formatted report is illustrated in Figure 3.

In the end of the trees transversals the set of reporters functions can be used to produce a final report. In this step all the reporters functions are executed, and the results are aggregated to build the final report (algorithm 5). Besides these structured reports, the full tree is available as an associative array to be further processed by any other tool or application.

[6] `http://www.graphviz.org/`

Algorithm 4. Reducer Example: Spell Checker

Input: children : list of Nodes
Result: New set of features to be added to the node.
$totalAcc \leftarrow 0$
$foundAcc \leftarrow 0$
for $child \leftarrow children$ **do**
  $attr = child.getAttr(\text{"SpellChecker"})$
  $totalAcc \leftarrow totalAcc + fst(attr.value)$
  $foundAcc \leftarrow foundAcc + snd(attr.value)$
$new = Feature(name = \text{"spellChecker"}, value = (totalAcc, foundAcc))$
return $[new]$

Algorithm 5. Build Final Report

Input: tree : Tree representing package content.
Input: reporters : Set of available reporters.
Result: Final HTML report.
for $r \leftarrow reporters$ **do**
  $slice = sliceTree(r.features)$
  $final = final + r(slice)$
return $final$

Reporters. The reporters process a specific set of features about the package and produce custom reports. They are mainly used for producing reports that require post processing computations to achieve the intended result in the report (averages computations, for example). Reporters' usually compute a final grade for a specific analyzed feature (the formula for computing the grade is another responsibility of a reporter function). Reporters' output is usually a snippet of HTML, built using a default set of templates.

4 DMOSS Quick Tour

This section illustrates a step-by-step usage of the toolkit applied to the `tree` software package.

The first step is to process the software package, this is done using the `dmoss-process` tool, which has a mandatory argument, either the file, or the complete URL for the package. The result of processing the given package is a tree, decorated with attributes storing the computed features, by default this tree is stored in a filename called `dmoss.data`. An example of execution is:

```
$ dmoss-process -file tree-1.5.3.tgz
Data saved as dmoss.data
```

This builds the tree representing the package, and executes all the tree transverses described in Section 3. This information can now be used by other tools, including the tool that builds a final report about the package, using all the defined reporters functions. An example of execution of this tool is:

```
$ dmoss-report dmoss.data > report.html
```

The result `report.html`, illustrated in Fig. 3, shows metrics that are used to grade key features about the package. For example, many documents in software packages contain links to official websites or discussion forums, one of the plugins included in the toolkit validates that these link are still working. If all links included in the documentation are working this feature is graded *A*. Another example is the number of comment lines in order to the total source code lines. In this specific case the percentage of comment lines per number of line codes is below 20%, which graded this feature of documentation with grade *F*. Averaging all the features the final grade for the documentation in the package is *C*. Some of these features are based on thresholds, that can be configured and adapted to specific contexts or packages. By clicking on each specific feature in the HTML report, more information is shown regarding each specific metric. A final grade is given to the package (*C* in this report), which is the features' grade average.

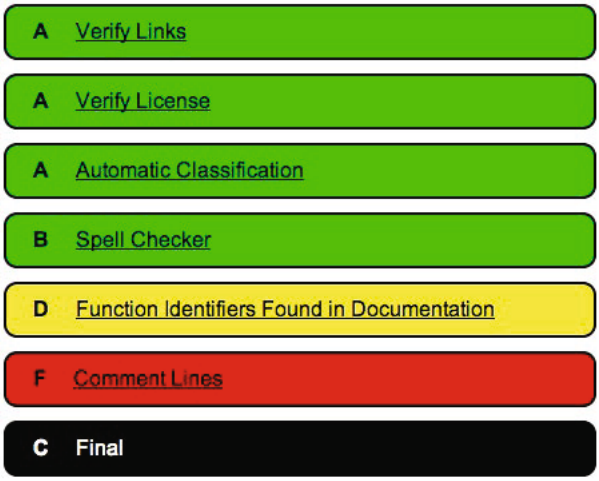


Fig. 3. Screenshot of a HTML report produced using DMOSS

5 Conclusion

Non-source code content found in software packages provides useful insights and information about the application. This information can be used in many distinct areas: software certification, or source code understanding for software maintenance or evolution.

The DMOSS toolkit is able to process a software package as an attribute decorated tree, and has proven most valuable. Since once all the major engines (algorithms described in Section 3) were implemented, adding features to the analysis workflow is just a matter of adding a new plugin. This approach has allowed the development of a modular and pluggable toolkit, easy to maintain,

and extend. The toolkit can process any package, regardless of programming language used, but the text extracting tool (from files) can require update for some specific archiving technologies.

Regarding the obtained results from software package analysis, we noticed that there is quite a lot of concern about overall package natural text information content. Nowadays, communities spend time making sure that information for users and developers is available, and up-to-date. There is also a concern with information related with licenses and other less software engineering content. There are still some features that are more prone to have lower grades, for example the number of comment lines per lines of code.

Some tasks that can be performed in the future to improve this work:

- increase the number of available plugins, and thus increase the number of analyzed features;
- implement tools that provide other views of the decorated tree, for example browsable graphs;
- some key features require a more detailed investigation because they are prone to less grades, and maybe the evaluation process needs to be relaxed.

Acknowledgments. This work was partly supported by project CROSS (PTDC/EIA-CCO/108995/2008), funded by the Portuguese Foundation for Science and Technology.

We would like to thank the reviewers for their valuable insight and detailed comments, which aided in improving this paper.

References

1. Chen, J.C., Huang, S.J.: An empirical analysis of the impact of software development problem factors on software maintainability. Journal of Systems and Software 82(6), 981–992 (2009)
2. Forward, A., Lethbridge, T.C.: The relevance of software documentation, tools and technologies: a survey. In: Proceedings of the 2002 ACM Symposium on Document Engineering, pp. 26–33. ACM (2002)
3. Hauge, Ø., Ayala, C., Conradi, R.: Adoption of open source software in software-intensive organizations–a systematic literature review. nformation and Software Technology 52(11), 1133–1154 (2010)
4. IANA. MIME Media Types. Web site, `http://www.iana.org/assignments/media-types/index.html` (last accessed: November 27, 2012)
5. Kitchenham, B., Pfleeger, S.L.: Software quality: the elusive target (special issues section). IEEE Software 13(1), 12–21 (1996)
6. Nelson, M.L.: A survey of reverse engineering and program comprehension. arXiv preprint cs/0503068 (2005)
7. Rajlich, V., Wilde, N.: The role of concepts in program comprehension. In: Proceedings of 10th International Workshop on Program Comprehension, pp. 271–278. IEEE (2002)
8. Scacchi, W.: Understanding the requirements for developing open source software systems. In: IEE Proceedings Software, vol. 149, pp. 24–39. IET (2002)
9. Thomas, B., Tilley, S.: Documentation for software engineers: what is needed to aid system understanding? In: Proceedings of the 19th Annual International Conference on Computer Documentation, pp. 235–236. ACM (2001)

A Networked Application to Support the Learning of Electronic Marketing Based on e-Learning and a Portfolio of Mediating Tools

Luis Vaz and Nuno David

Dinamia-CET/ISCTE
ISCTE – Lisbon University Institute, Lisboa, Portugal
luis.m.vaz@mail.telepac.pt, nuno.david@iscte.pt

Abstract. Transition from traditional teacher-centered models, focusing on curricula description and content plans, to student-centered models, in which learning activities are supported in e-learning platforms, is not an easy task to resolve. This issue is especially important when the knowledge is emergent, i.e. when the activities that this knowledge enables are intensively supported by information and communications technologies, subjected to rapid obsolescence. A method for mapping a course content plan into technological mediating tools that form a portfolio is proposed. A mediating tool is a set of learning activities organized according to a specific structure. The portfolio enables synergetic relations between mediating tools, promoting a good content organization and high-order learning activities. The characteristics of the resulting application are described for a course of electronic marketing, in which a content plan is mapped into a set of mediating tools organized in a portfolio. The application was evaluated with surveys and benchmarked with traditional teaching and other e-learning solutions. Benchmarking positioned our e-learning solution as having the best features in team building, achieving scores similar to traditional teaching.

Keywords: collaborative learning, e-learning platform, electronic marketing, emergent knowledge, practice field, puzzles, simulation.

1 The Difficulties of e-Learning Instructional Models

The use of Information and Communication Technologies (ICTs) has been changing the way we build and deliver instruction, enabling learning in new forms and to newcomers in an unprecedented way. Nevertheless, this type of delivery implies pedagogical innovation, which leads to a higher risk of failure and falling reputation of teachers and educational institutions. Failure risks and the frustrations they may cause are well illustrated in cases described in the bibliography related to the application of information systems and technologies to education, usually called e-learning [1,2]. Risks that are especially salient in disciplines where the knowledge is emergent, i.e. when the activities that this knowledge enables are intensively supported by ICTs and subjected to rapid obsolescence.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 795–806.
DOI: 10.1007/978-3-642-36981-0_74

We use the word emergent behind knowledge to give the sense of something that has just arrived but that evolves fast, like nations with social or business activity in the process of rapid growth and industrialization. We can find this type of knowledge in relation with the fast development of some disciplines, like information systems and technologies, management, some life sciences. In particular, we consider the field of Electronic Marketing (EM) as emergent knowledge.

The application of the traditional teaching model to emergent knowledge presents instructional deficiencies that prevent students from acquiring recent and highly deprecated knowledge, as well as skills and abilities necessary for the effective exercise of their professions. We use the term traditional teaching to refer to that kind of learning heavily based on the appropriation of encyclopedic knowledge, and transmitted by a teacher in a classroom. It has several shortcomings, such as the difficulty in promoting high-order learning, including the difficulty in promoting practices within the domain of knowledge to be transmitted, especially when the application of that knowledge entails specific practices. Deficiencies significantly more salient when the traditional model is applied to subjects where students will come to use knowledge acquired in the exercise of professions that require the use of information systems and high-technology environments [3]. The gap between traditional teaching methods and the needs of students is also a concern in the discipline of marketing. More and more professionals are using ICTs in data research, market studies, price competition and market driven competitive models [4,5].

Aware of these difficulties and of the increasing level of demand in the teaching of emergent knowledge, teachers are required to have a solution for this instructional problem, whose relevance becomes increasingly pressing [6,7]. The construction of courses of this nature is however struggling with difficulties, including those relating to the ability to implement a specific plan of content in a learning model intensively supported in ICTs, and where learning should be faced by students in a motivating and engaging way [8,9]. Some authors advise the use of case studies or patterns that can be a form of guided support or scaffolding to the process of learning design. Some explain the passage of specific content to e-learning, in which the content is mapped to mediating tools that cluster learning activities [10]. A learning activity is any form of organized activity that promotes learning. To some extent, this is also our approach. Mediating tools (MTs), such as puzzles, simulation, practice fields and text/hypermedia are used to deal with a content plan for learning EM. But a new value for this kind of learning design must be endorsed, once the interaction between MTs is considered. Interaction requires the organization of MTs in a portfolio, allowing the establishment of practical rules to map the content into learning activities, and enabling new learning capabilities in the e-learning solution.

In this paper we put forward a method for mapping learning content into MTs organized in a portfolio, and that can be implemented in an e-learning platform. The paper is organized as follows. Section 2 identifies the purpose and the characteristics of each MT. In section 3 MTs are organized in a portfolio that produces synergetic effects. Section 4 explains the way content is segmented into MTs. In section 5 a concrete example is described in the field of electronic marketing. Results are reported in section 6.

2 Which Mediating Tools?

We see an increasing trend to include MTs that provide support for certain learning approaches, such as problem-based learning and resource-based learning.

From literature we can identify the development of patterns for: i) person centered e-learning [11]; ii) collaborative learning [12]; iii) designing adaptive hypermedia for learning styles [13]. Specifically for EM we cannot find in the literature learning design patterns for devising learning activities in an e-learning solution. Notwithstanding, adopting the correct MTs is the fundamental task of a teacher when he designs learning activities, which should be aligned with the results that the teacher intends to obtain, and should provide a way for checking that the results were indeed achieved by students [14]. As part of our proposal, we put forward a specific solution of MTs: i) puzzles; ii) simulations; iii) practice fields; and iv) hypermedia content. All these MTs will cluster sets of learning activities supported in an e-learning system in which students learn EM.

Puzzle-based learning approaches enthusiasms students as i) they promote edutainment and ii) engagement with learning; iii) can be applied to solve real-world problems; iv) conclusions taken from puzzles can be applied to a broader context [15].

But the framing of the link between learning through problem solving and simulations has attracted more attention. The simulation in this context appears at different roles, allowing to keep the requirements of a good learning through problem solving, namely: i) the management of large amounts of information; ii) the connection of multiple domains in a structured way; iii) the setting up of the need for creating real situations [16].

A practice field is also a MT where learning activities can be clustered. The implementation of practice fields allows the construction of instructional activities that implement the principle of situated learning. It also makes learning an active process, on a student-centered learning model, where collaboration and cooperation are promoted through the implementation of practices where students play roles that will be daily present when exercising the profession they are studying for.

Finally, with a complementary role to these MTs, we also use learning activities based on the analysis and study of hypermedia and text content.

3 Organizing Mediating Tools for Synergistic Effects

To organize MTs in a portfolio we consider two main characteristics of content that could be mapped in MTs. The first concerns the extent to which the content can be modeled. A second dimension characterizes the kind of interaction that learning activities clustered inside MTs require in the relationship of the student with the learning environment. In this dimension we consider asynchronous learning as learning activities that could be carried out without the use of the e-learning platform, and synchronous learning as the activities that can only be carried out using the e-learning platform in a synchronous mode. Learning design has been implemented mainly based in asynchronous situations, where the content in a learning management

system will complement learning activities in a learning room. This means that learning in a synchronous environment introduces special characteristics in the learning solution, which the learning designer should have in mind when considering how to map content into MTs.

3.1 Modeling and Non-modeling Contents

Content plans seem to have two main tendencies. Content that is easy to be modeled, such as describing facts, concepts or processes, whose origin is based on certain theories, possibly where mathematical functions play a role, that is, functions that can be easily coded and transformed into learning activities framed in games, puzzles and simulations. On the other hand, content not liable to being modeled should be transformed into activities in which learning is based in an-hands on approach, such as a practice field. Practice fields are, however, not always easy to build. In such cases, learning activities should be based on the study of content in the format of text and hypermedia.

3.2 Synchronous and Asynchronous Learning

Two kinds of interactions with the system are considered: i) simple interactions, where students develop learning activities asynchronously; ii) complex interactions, where students develop their learning activities in a synchronous environment with the system.

Upon constructing the learning environment, learning activities grouped in a simulation and practice field use characteristics synchronous with the system. In highly ICT supported professions the only way to implement a practice field is in a synchronous mode with the e-learning system. Moreover, a simulation can only be effective if its results are given in a short period of time, maintaining the students' interest.

Learning activities organized in puzzles and passive hypermedia elements should be asynchronous with the system. This allows greater flexibility in the organization of the students' time. Puzzles are usually implemented based on information search in different sources, external to the e-learning system, harder to implement in a synchronous mode with the e-learning system.

In the context of the method proposed, the organization of a content plan in learning activities, as well as the interactions among such learning activities, is presented in figure 1.

3.3 Interactions among MTs

In the portfolio MTs interact with each other. Interactions, identified in Fig. 1, are a key success factor for building a learning model in which its value is higher than each one of the MTs alone. Moreover, the identification of such relations will facilitate the establishment of requirements about the architecture of the learning management system.

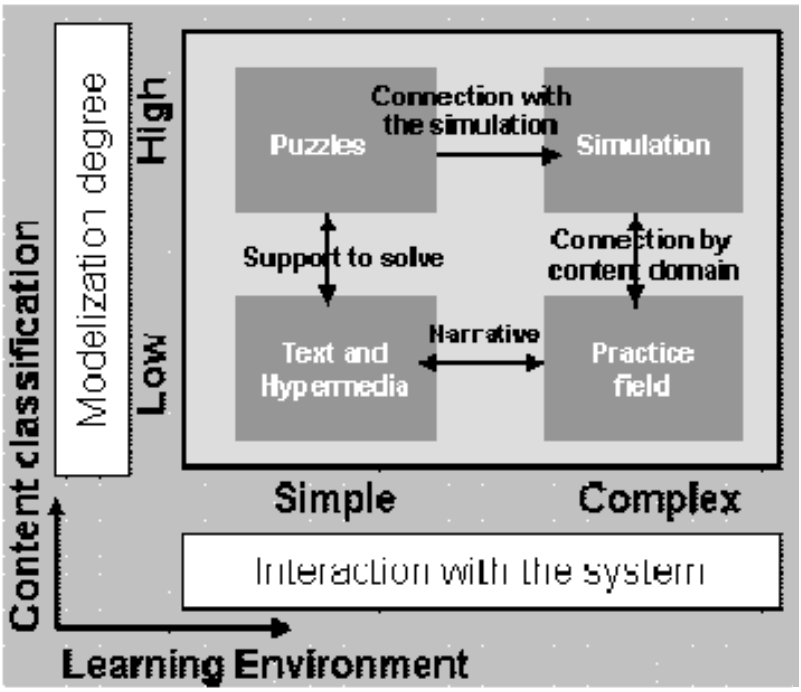


Fig. 1. A portfolio of Mediating Tools

Solving puzzles allow students to obtain information later used for calculating the model of the simulation. While solving a puzzle, students get "an entrance" to the black box of the simulation. The entrance appears as the knowledge necessary for solving one of the functions that support the simulation. A number of puzzles should be the solution for a set of formula that forms the black box of the simulation. In sum, for each interaction with the puzzle students acquire additional pieces of information used in the calculation model of the simulation.

A relationship between simulation and the practice field is established through the description of a common base of knowledge characterizing a simulation and a practice field. For instance, EM has a common knowledge base, theoretical and practical, on the use of media that supports online advertising.

The passive learning activities connected with the study of content are related with puzzles, simulations and practice field, having a relationship with puzzles that we can characterize as an explanation of how puzzles should be analyzed in order to be solved; a relationship with the practice field through the explanation and the framing of the history and narrative that support it; and with the simulation insofar as they enable the learning of the context and history of the simulation, the simulation rules, the kind of data entry and how results are presented.

4 Segmenting Content for Mediating Tools

Segmenting content into MTs requires an understanding of how learning activities may be clustered inside each MT. In the following sections we explain some activities that can be developed to properly map content into MTs.

4.1 From Content to Puzzles

Solving puzzles is a way to implement learning through problem solving and discovery. Its application can be found in the learning of programming languages and in learning related to operating systems [17,18]. Puzzles are often used in a dynamic of game [19]. A puzzle applied to learning is an exercise of discovering and information integration to construct a coherent whole; in doing these activities a person will be able to learn. We see puzzles as a way to build interest in simulations by open windows in the simulation black box. In the case of EM, we also foresee the use of puzzles in learning situations where it is necessary to gain knowledge on the use of certain management reports.

In such context, the first activity to map content into puzzles is the characterization and segmentation of content, being particularly suitable for modeling those parts who better characterize the key performance indicators of the management model intended to be learned. For example, in internet advertising, reports on the characterization of visits to a website may be particularly suitable to be transformed into puzzles. For instance, a perceptual map about the nature of a website could be generated, give it to the student, relate it with the response of the market to an email campaign, and ask the student to find such relation.

4.2 From Content to Simulations

More and more different management disciplines are being studied and modeled [20]. This trend has been reinforced by the decreasing costs of ICTs. It enables practitioners and theorists to proceed in the implementation of mathematical models to explain and study management issues. Marketing phenomena has also been intensively modeled in different domains. This growing tendency is an opportunity for building simulations.

In the case of EM we have identified advertisement as a type of content that has been highly subject to modeling [21]. A feature that arises from the nature of advertisement, which can be modeled as the result of a given stimulus to actual or potential customers who buy the product. In the case of EM, we choose an advertising model, ADBUG [22], and decide to implement this model in a discrete event based simulation.

After identifying the content to transform in one simulation, the designer should choose the type of approach to build the simulation model. Many approaches use discrete events to build the mathematical model of the simulation.

This enables to manage time in periods where the state of simulation changes at discrete points that we can associate to points in time of management decisions. In each of these discrete event points in time, we can make changes in the model that reflect changes in the market.

4.3 From Content to a Practice Field

Instructional designers can model a practice field by observing the roles that their students will play in their professional lives and how they apply what they learned as students. For example, if these practices are related with actions that have impact on a market, the students in the class could play all together the role of consumers. Each student could be identified as a professional of the practices modeled in the practice field, e.g. in a practice field for EM students can do practices related with the profession of a marketeer. In this way students will have the opportunity to understand the impact that their practical activities have during the frequency of the class. In sum, it is essential to develop authentic activities in the practice field, in the same social and collaborative context that students will find after concluding their studies.

4.4 From Content to Text or Hypermedia Format

Transferring content to text and hypermedia format is a well established practice. Learning designers should follow the best practices in this domain, namely in what respects to content organization, interface design and the tutorial system [23].

5 An Effective Example in the Field of Electronic Marketing

Professionals of electronic marketing make intensive use of information and communication technologies. An intensity that makes the teaching of electronic marketing somewhat particular when compared with other disciplines, and quite challenge as well. Teaching electronic marketing can only be effective once the learning environment reaches similar levels of technological intensity as the discipline itself. A demand that suggest us that electronic marketing might be particularly appropriate to the use of e-learning as a teaching instructional model.

For teaching EM through e-learning we suggest the passage of a content plan based course to an e-learning solution, whose implementation methodology should be based on the grouping of learning activities as outlined above, and according to Figure 1.

The content plan we propose relies mainly on management activities for promoting products and services through media channels supported intensively in ICT. The topics can be related to each one of the quadrants of the portfolio, according to its characteristics as well as the strategy for implementing the learning activities, according to Table 1. In the remainder of this section we briefly describe the characteristics of the MTs deployed in our platform; see [24, 25] for more details.

Table 1. Mapping content plan for teaching EM

	Modeling Degree	**Level of Interaction with the e-learning system**	**Type of grouping**
Advertising and business intelligence on internet	High	Low (Solution to develop strongly asynchronous)	Puzzles
Management plan media on internet	High	High	Simulation
Activities of viral marketing on internet	Low	High	Practice field
Learning activities that support other groups	Low	Low (Content in text for printing).	Text/ Hypermedia

Activities supported by text reading and study of hypermedia elements – the last line of Table 1 – are intended to support the more complex learning activities, including simulation, puzzles and practice fields, as well as supporting autonomous learning activities such as site placing, introduction to EM and conversion strategies.

5.1 A Simulation for e-Learning in Electronic Marketing

When doing advertising campaigns companies expect to increase sales. In order to be effective and efficient those responsible for marketing have to: i) analyze the situation; ii) identify the objectives of advertising; iii) identify a strategy for the message to be achieved with the advertising; iv) develop a media plan; v) make the budget allocation; vi) implement and evaluate [26].

The planning of a media campaign has as objective the alignment between media where advertisement will be placed and the segments the product we want to promote. Marketing managers should develop advertising in media that better correspond to efficiency, i.e. in which the targets hit by advertising are maximized at minimum cost [27]. In our platform, students, when using simulation, intend to maximize the achievement of market share through the implementation of a previously defined media plan. The simulation will have a set of functions in its core (black box) which will simulate the response of a market when subjected to advertisement campaigns.

5.2 A Puzzle for e-Learning in Electronic Marketing

Puzzles were constructed as market studies. They are reports of the application of business intelligence to the traffic that visits the targets of the media campaign to be held, in this case Web sites, forums and mailing lists of electronic mail.

The introduction of puzzles in the learning model aimed at transforming a set of content related to advertising in order to provide students with tools for market analysis, that are nowadays part of the techniques adopted in any communication and advertising plan.

Solving puzzles has enabled students to better understand the simulation. This objective is achieved by providing students with openings in the black box of the simulation, giving students the opportunity to understand how the model inherent to the simulation works.

5.3 A Practice Field for e-Learning in Electronic Marketing

The practice field implemented in the e-learning application has an architecture that comprises three characteristics, important for teaching disciplines with high technology intensity: (1) it must be integrated in an e-learning environment extended with an intelligent tutoring system; (2) practices should be authentic and supported by the same execution and computer supported contexts as the ones students will find in their professional life; (3) dynamics created by the practice field at the cognitive, social and emotional dimension should be maintained by a narrative supported by hypermedia means of the e-learning platform.

In our practice field for EM the practices comprise activities of electronic mail, messages on forum, online advertising through banners, online purchasing and permission marketing.

6 Implementation and Results

The e-learning platform was tailor made, with the specific purpose of teaching EM and testing the instructional model. The need to introduce specific requirements led us to implement it from scratch, namely i) the need of a tutorial system, for coordination and synchronization of learning activities; ii) the specificity of the practice field, which must be integrated in the e-learning platform, as well as the need of functional integration between communication tools inside the e-learning platform and the practice field; iii) the specificity of the simulation, related to data validation in the input process, as well as to the output results, which integrate graphical representations; iv) the temporal sequence and the information exchange between learning activities clustered in different MTs.

In effect, in [25] we show that the functional requirements of a platform resulting from a learning model based on mediating tools, but implemented on traditional teacher-centered platforms, would eventually imply the implementation of a whole new solution given the depth of the changes that would have to be carried out.

Our technological solution is based in a networked application supported on internet technologies. The access is made through the browser present in any computer and demands a typical internet connection. The server side of the application is supported in an open source operating system, Linux, connected to a database.

Two hundred and ninety three students in eight post graduate courses attended the discipline of EM with the instructional approach implemented in the networked application. Each class had between twenty and thirty five students; ten classes were delivered with this instructional model. The discipline lasted for five weeks, in eighteen different modules.

A 6-page survey was administered to all students at the final of the discipline. The purpose of the survey was the evaluation of the instructional model as well as the IT supporting the e-learning platform.

Our discipline was the first one delivered with this method in each post-graduation course. One objective was the development of skills for building consistent teams that could proceed strong in the entire course. Achieving this objective, while delivering the EM discipline, was one of the main results of the instructional model.

To benchmark the team building capabilities we compared our results with other results collected from two other sources, namely, a survey applied to different online courses using the IDEA center survey [28] and a web assisted course in the field of marketing education [29]. The web assisted course has also been benchmarked with a teacher centered discipline delivered in a traditional class.

In the following tables we show the results of such benchmarking, with our discipline achieving a 3.7 score (Table 3), the highest score among the distance delivery approaches, and close to the traditional teaching model.

Table 2. Benchmarking team building skills

Traditional class	4,1
Our discipline – E-Marketing	3,7
Web assisted course in the field of marketing	3,4
Online courses rated with the IDEA survey	3,3

The students also give a high score to their satisfaction with the use of the application, well above the other web assisted course (Table 4).

Table 3. Benchmarking student satisfaction

Traditional class	4,2
Our discipline – E-Marketing	4,0
Web assisted course in the field of marketing	2,6

7 Final Remarks

This proposal has produced a guide for mapping learning content into MTs organized in a portfolio, and that can be implemented in an e-learning platform. The method has been proven robust, and a practical application to support learning in electronic marketing is provided.

In effect, we consider our benchmarking results particularly remarkable when benchmarked with other disciplines delivered with the same distance model. The instructional model has worked and the students reported a good perception of learning. As a main benefit from the delivering method, they also reported the freedom to study with more flexibility. The students evaluated positively the program of EM, achieving a score of 3.4, which is again a good classification in the context of

learning based on e-learning. Working in teams was also positively evaluated, in line with the results of team building skills. Other technological features, such as download times and system time of response, were much appreciated.

References

1. Law, N., et al.: Methodological approaches to comparing pedagogical innovations using technology. In: The First IEA International Research Conference, Lefkosia, Cyprus (2004)
2. Huk, T., Lipper, T., Steinke, M., Floto, C.: The role of navigation and motivation in e-learning – the crimp-approach within a swedish-german research cooperation. In: Proceedings of the Annual Conference 2002, pp. 364–369 (2002)
3. Bailey, D., Dessouky, M., Rickel, V., Verma, S.: A methodology for developing a web based factory simulator for manufacturing education. Department of Industrial and Systems Engineering. University of Southern California (2000)
4. Gillentine, A., Schulz, J.: Marketing the fantasy football league: Utilization of simulation to enhance sport marketing concepts. Journal of Marketing Education 23(3) (2001)
5. Castleberry, S.B.: Using secondary data in marketing research: A project that melds web and off-web sources. Journal of Marketing Education 23(3), 195–203 (2001)
6. Oliveira, E.D., Guimarães, I.C.: Employability through competencies and curricular innovation: a Portuguese account. Faculty of Economics and Management, Catholic University of Portugal (2010)
7. Conole, G., Thorpe, M., Weller, M., Wilson, P., Nixon, S., Grace, P.: Capturing practice and scaffolding learning design. The Open University, UK (2007)
8. Alonso, F., Lopez, G., Manrique, D., Vines, J.M.: An instructional model for web-based e-learning education with a blended learning process approach. British Journal of Educational Technology 36(2), 217–235 (2005)
9. Richards, C.: The design of effective ICT-supported learning activities: Exemplary models, changing requirements, and new possibilities. Language Learning &Technology 9(1), 60–79 (2005), http://llt.msu.edu/vol9num1/pdf/richards.pdf
10. Conole, G.: Using compendium as a tool to support the design of learning activities. In: Okada, A., Buckingham Shum, S., Sherborne, T. (eds.) Knowledge Cartography: Software Tools and Mapping Techniques, ch. 10. Springer, London (2008b)
11. Derntl, M.: Patterns for person-centered e-learning. PhD thesis, Faculty of Computer Science, University of Vienna, Austria (2005)
12. Leo, D.H., Perez, J.I.A., Dimitriadis, Y.A.: IMS learning design for the formalization of collaborative learning. In: Proceedings IEEE Int. Conference on Advanced Learning Technologies, pp. 350–354 (August 2004)
13. Brown, E., Cristea, A., Stewart, C., Brailsford, T.: Patterns in authoring of adaptive educational hypermedia: A taxonomy of learning styles. Education Technology & Society 8(3), 77–90 (2005)
14. Biggs, J.: Aligning teaching and assessing to course objectives. Paper presented at ICHEd Conference: Teaching and Learning in Higher Education: New Trends and Innovations, April 13-17. University of Aveiro, Portugal (2003)
15. Michalewicz, Z., Michalewicz, M.: Puzzle-based learning. In: Proceedings of the 2007 AaeE Conference, Melbourne. University of Adelaide, Australia (2007)
16. Tan, O.-S.: Problem-based learning approach for simulation and modeling. National Institute of Education. Nanyang Technological University, Singapore (2007)

17. Parsons, D., Haden, P.: Parson's programming puzzles: A fun and effective learning tool for first programming courses. In: Proceedings of the Eighth Australasian Computing Education Conference (ACE 2006), Hobart, Australia, pp. 157–163 (2006)
18. Hill, J.M., Ray, C.K., Blair, J.R.S., Carver, C.A.: Puzzles and games: addressing different learning styles in teaching operating systems concepts. In: Proceedings of the Technical Symposium on Computer Science Education (SIGCSE 2003), pp. 182–186 (2003)
19. Freitas, S.: Learning in immersive worlds. A review of game-based learning. Joint Information Systems Committee, London (2006)
20. Chandukala, S.R., Kik, J., Otter, T., Rossi, P.E., Allenby, G.M.: Choice models in marketing: Economic assumptions, challenges and trends. Foundations and Trends in Marketing 2(2), 97–184 (2008)
21. van Heerde, H.J., Neslin, S.A.: Sales promotion models. In: Wierenga, B. (ed.) Handbook of Marketing Decision Models, pp. 107–162. Springer, New York (2008)
22. Cannon, H.M., Leckenby John, D., Abernethy Avery, M.: Modeling advertising media effectiveness. Developments in Business Simulation & Experiential Exercises 23 (1996)
23. Caeliner, S.: Designing e-learning (2002)
24. Vaz, L., David, N.: A practice field for teaching electronic marketing. In: Proceedings of 4th International Conference on Computer Supported Education, CSEDU 2012 (2012)
25. Vaz, L., David, N.: Mapping content plans into learning activities: Organizing a portfolio of e-learning activities for teaching emergent knowledge. Accepted to IEEE EDUCON 2013 - Global Engineering Education Conference (2013)
26. Smith, S., Swinyard, W.: Advertising models. In: Introduction to Marketing Models, ch. 9, pp. 1–17 (1999)
27. Parente, D.: Advertising campaign strategy – a guide to marketing communication plans. South-Western College Pub., USA (2006)
28. Benthon, S., Webster, R., Gross, A., Pallett, W.: An analysis of IDEA student ratings of instruction in traditional versus online courses, IDEA Center. Technical Report Nº15 (2010)
29. Priluk, R.: Web-assisted courses for business education: An examination of two sections of principles of marketing. Journal of Marketing Education 26, 161 (2004)

A Platform-as-a-Service API Aggregator

David Cunha[1,2], Pedro Neves[1], and Pedro Sousa[2]

[1] Portugal Telecom Inovação, SA, 3810-106 Aveiro, Portugal
[2] Centro ALGORITMI / Department of Informatics,
University of Minho, Braga, Portugal
{david-g-cunha,pedro-m-neves}@ptinovacao.pt,
pns@di.uminho.pt

Abstract. Cloud Computing has emerged as a new paradigm for services delivering over the Internet. In this growing market, PaaS (Platform-as-a-Service) service model has been the subject of great interest from several organizations allowing a straightforward deployment of applications without the need for a dedicated infrastructure, dependencies installation or server configuration. However, each PaaS provider generates a lock-in to their proprietary features, technologies or APIs (Application Programming Interfaces). This paper focuses on the specification, development and test of an abstraction layer that aims to unify the management and information processes of applications and databases created in PaaS environments, in order to struggle the existing lock-in in the market. As result, PaaS users will have now the opportunity for selecting the most appropriate platform for an application, interacting seamlessly with any supported vendor, also having the opportunity to attain the portability of applications among distinct PaaS providers. As far as we know, this is the first work in this research area presenting a real implemented and tested system aggregating several PaaS offerings.

Keywords: Platform-as-a-Service (PaaS), Interoperability, API.

1 Introduction

With the growth of the Internet and processing technologies, computational resources' prices have reduced becoming accessible to the masses. These circumstances made the application delivering, now accessible by multiple terminals, a new attractive market for the IT industry. Besides a buzzword as the web is, cloud computing is an evolution of several paradigms transforming the dream of computing as a utility into a reality [1]. The potential of such a model advocates a major impact on the IT industry with computational resources and software delivered through a pay-per-use paradigm [2]. Likewise, the cloud computing analogy with traditional services such as electricity, gas or water is always present. This means that end users can access services when they wish, regardless of their locations and being charged by service providers only regarding the levels of use [3].

Over the past years, roughly since 2009, there have been several PaaS vendors and start-ups competing and focusing arduously on innovative aspects that may

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 807–818.
DOI: 10.1007/978-3-642-36981-0_75

differentiate them from the user point of view. In the platform market, there are polyglot PaaS with a broad range of runtimes, frameworks, databases, but also PaaS only specialized in specific programming languages or specific tools. The ecosystem covers from Java, Ruby, PHP, .NET environments, up to Ruby on Rails, JavaEE, MySQL or NoSQL databases. At the same time each vendor provides different business models, development tools and APIs. This mix of vendors favors the selection giving the opportunity for developers to evaluate the different offers with no installation, configuration or maintaining any dependency [4]. But, on the other hand, a developer may somehow get locked to a specific vendor and its proprietary features. Interoperability between cloud service providers has emerged as a topic of general concern associated with the inter-cloud term [5]. If this concept were applied in a business environment, it would decrease the reticence of many organizations in adopting cloud computing, namely PaaS solutions. Consequently, it would offer the ability to manage, monitor and migrate data seamlessly between heterogeneous vendors, in order to abstract the intrinsic differences from each PaaS provider [6].

Much of the effort that has been made in this area involves only the IaaS layer, such as OGF's OCCI or Apache DeltaCloud API. However recently, in late August 2012, some players such as Oracle, Rackspace, RedHat, CloudBees, Huawei, Cloudsoft Corporation and Software AG announced an initiative with the main objectives of specifying and developing an open-source API named CAMP[1] (Application Management for Cloud Platforms). This interface will allow developers to create, monitor and manage applications and databases across multiple PaaS vendors. CAMP leverages the inherent similarities in the application lifecycle and aims to create a generic management solution. With this initiative, the partners expect to revolutionize the platform market, having submitted CAMP specification to OASIS to be adopted in the future as a possible PaaS standard. However, it is not yet clear if such standards will be effectively adopted and implemented in the future by many relevant vendors in this competing market, resulting that heterogeneous PaaS solutions be always expected to exist in the market.

In this context, this paper aims to contribute to this research field presenting a novel abstraction layer, named as PaaSManager, which aggregates several relevant PaaS public offering based on their similarities exposing a common API for developers. The proposed architecture addresses the interoperability between providers and the portability of applications in order to reduce the vendor lock-in.

After introducing the topic covered in this paper, section 2 makes a briefly summary regarding what a Platform-as-a-Service is, and which PaaS offerings are available in the market. Section 3 describes the proposed framework features as well as the PaaSManager architecture. Section 4 details the several modules that form the PaaSManager and some supported operations are then tested in section 5. Finally, the paper ends with the conclusions in section 6.

2 Platform-as-a-Service

So far, the IaaS is the most used and successful service model in enterprise cloud computing. However, the PaaS service, being application oriented has the potential to

[1] `http://www.cloudspecs.org/paas/`

abstract organizations from all the configuration and maintenance processes [7]. A PaaS provider offers a simple and intuitive integrated environment for developing, testing, monitoring and hosting applications and databases. These platforms are intended to reduce the complexity and time allocated in the application lifecycle, supporting the auto-scalable features of cloud computing. Within days, a simple service prototype can be released to production without a large initial investment. Suddenly, innovation ceases to be only on the horizon for those who have economic resources to materialize it.

2.1 Platform Market Overview

Besides titans like Amazon, Google and Windows, other offers emerge such as Java-based PaaS CloudBees, and also open-source solutions, namely VMware's CloudFoundry and Red Hat OpenShift. In turn, the Salesforce.com's Heroku is also a leading vendor of open source languages, as for instance Ruby, supporting Git tool for the deployment process.

CloudBees[2] was founded in 2010 and is a PaaS entirely directed to the development of Java-based applications. It provides two services, one being directed to the application development and testing (DEV@Cloud), and other targeted only to the application deployment and execution (RUN@CLoud).

CloudFoundry[3] is a PaaS that has a dissimilar approach in the market by being fully open-source and multi-cloud. Launched in 2011 by VMware, CloudFoundry offers a polyglot environment without being attached to a single infrastructure vendor. Users have the opportunity to change the PaaS source code and seat it on any infrastructure service at their disposal, whether public or private. With this multi-cloud and open source paradigm, CloudFoundry has received extensive recognition. Behind this success, there is a strong community of users and organizations as well as PaaS solutions based on CloudFoundry like AppFog, Stackato and the .NET extension IronFoundry.

Heroku[4] came in 2007 as a Ruby oriented PaaS. After being acquired by Salesforce.com in 2010, Heroku supports much more technologies and it is becoming one of the most used PaaS in the market. Currently, it is estimated that more than 1.5 million applications are hosted in Heroku. Heroku has an extensive catalog of add-ons that allows users to add various types of services: logging, billing, testing, monitoring etc. Unlike the aforementioned vendors, Heroku only provides Git for source code deployment. Hence, Git is becoming a *de facto* in PaaS market being supported also by Red Hat OpenShift and recently by Amazon Beanstalk.

[2] `http://www.cloudbees.com/`
[3] `http://www.cloudfoundry.com/`
[4] `http://www.heroku.com/`

3 PaaSManager

The proposed PaaSManager is a layer supporting the fundamental operations implemented by the native APIs and abstracts the PaaS differences for developers. This common management approach is intended to fit the many developer's requirements in a PaaS environment, such as: **create and manage** applications and databases, **acquire information** concerning applications and databases, **monitor** applications in real-time, **migrate** applications between vendors. To fulfill the mentioned features, four providers were selected (CloudBees, CloudFoundry, IronFoundry, Heroku) to belong to an interoperable PaaS ecosystem. All vendors have different APIs, monitoring and deployments tools except CloudFoundry and IronFoundry that share the same API implementation but distinct technologies.

3.1 PaaSManager Functionalities Overview

As aforementioned the PaaSManager was designed for centralizing the different PaaS offerings in the market. For this to be feasible, the selected platforms' APIs and inherent management processes were broadly studied. As result, 20 key operations were specified based on the encountered similarities. In cases where the similarity factor was reduced, certain developments needed to be conducted in order to sustain a complete transparency for developers. In other cases, when some exposed method was not shared with the remaining providers from the ecosystem, an aggregation could not be made by the PaaSManager. By this means, the introduction of a new platform reduces the impact effect in the several system modules. Besides the similarity factor, the selection took into account the provision of fundamental application lifecycle processes.

In addition to simple tasks like create, deploy, start, stop, delete and monitor applications, developers have the possibility to restart, scale, update, migrate, create and delete databases in any supported vendor. In certain transactions transformations were necessary, including Git support for source code deployment due to Heroku, as equally aggregating various methods from the same vendor API to achieve the desired outcome. The scalability and taxonomy were unified now being possible to scale an application horizontally across more instances. Regarding the monitoring process each vendor supports different paradigms. CloudBees and Heroku only provide the collection of monitoring metrics through NewRelic API, which is an Application Performance Management widely used in cloud environments. On the other hand, CloudFoundry and IronFoundry expose a monitoring operation through their native API. To accomplish a seamless portability of applications, an algorithm was implemented enabling the migration to a new platform if it supports the dependencies needed to the application to run properly. From the wide range of supported operations, some fundamental processes will be later detailed over the section 4.

3.2 PaaSManager Logical Architecture

The PaaSManager architecture design kept in mind the prevailing modularity that allows the entire system to remains fully operational even if some vendor or

monitoring API is not operating correctly. Consequently, each API has been implemented by distinct modules and managed by single entities. Finally a REST interface exposes the specified operations to be invoked by any HTTP client application. Within Figure 1 is presented the PaaSManager architecture, as well as the several designed modules which support the operational processes interacting with the selected PaaS providers. Each module of the PaaSManager is described in section 4.

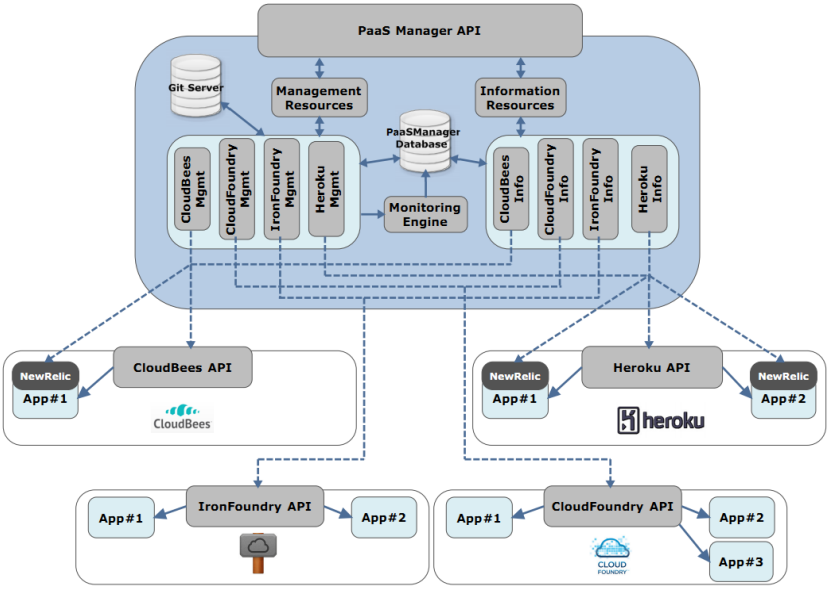


Fig. 1. PaaSManager Architecture

4 PaaSManager Modules

This section analyzes in detail the design and implementation of the main modules sustaining the PaaSManager architecture shown in Figure 1.

4.1 PaaSManager API

All the selected vendors provide a RESTFul API for developers, so it made perfect sense that the PaaSManager API was also REST compliant. This API is the element that exposes the supported features and abstracts occurring transformations in background. The interface can be easily implemented by any HTTP client leveraging a lightweight and web oriented approach with JSON representations. Each specified method has a respective URI which gives the possibility to manipulate resources through GET, POST, PUT and DELETE HTTP verbs. Within the addresses design,

REST good practices defined by Roy T.Fielding [8] were considered to keep the user interaction simple and intuitive as possible. As part of the authentication and authorization, all the requests to the PaaSManager API are secured via an *api-key* that identifies the logged user. On the other hand, the authentication and authorization with vendors are done through a unique account enabling the PaaSManager to acts as a mediator between users and platforms providers. Consequently the end user does not need to register in each vendor to use the corresponding cloud service.

4.2 Management Resources

The Management Resources module presented in Figure 1 is a decision module responsible for forwarding the message to the desired method implemented by a specific PaaS adapter. In terms of management services, arise four adapters: CloudBees Mgmt, CloudFoundry Mgmt, IronFoundry Mgmt and Heroku Mgmt. These adapters have all the logic that implements the methods related to management tasks (create, deploy, start, stop, etc.) and exposed by each PaaS API. Besides interacting directly with the vendors APIs and returning unified JSON responses, these adapters are integrated with other key elements, for instance, the PaaSManager database and a Git Server. The central database keeps state of created applications, storing the application framework identifier and the vendor name where the application is hosted. The Git server maintains repositories for each hosted application enabling users to keep multiple versions of source code. The Git repositories are essential for some of the crucial operations supported by PaaSManager in particular, create, deploy, update and migrate applications. Two illustrative examples of the deployment and migration processes are now described.

Deploy App- In order to unify the deployment process so the developer does not need to worry about the different tools offered by vendors, the PaaSManager only supports one tool: Git. Figure 2 illustrates the deployment operation.

The request received by the API is instantly sent to the Management Resources, which performs a search in the central database in order to acquire the platform identification where the application was created with a previous operation. Upon the outcome, the Management Resources invokes the specific PaaS adapter which in turn runs *git-add* and *git-commit* commands in the application repository. After this process is accomplished, the deployment is carried based on the PaaS supported paradigm. With CloudBees, the repository is examined until the adapter finds the application *web archive*. For CloudFoundry and IronFoundry the process is similar for Java-based applications, however, for other environments only a *.zip* file containing all the source code is sent to the platform API. Finally with Heroku, the deployment is accomplished running the *git-push* command to the remote repository created previously for the application. In case of success, a monitoring engine is started in order to extract real-time statistics to the central database. The monitoring engine adopted strategy will be analyzed in section 4.4.

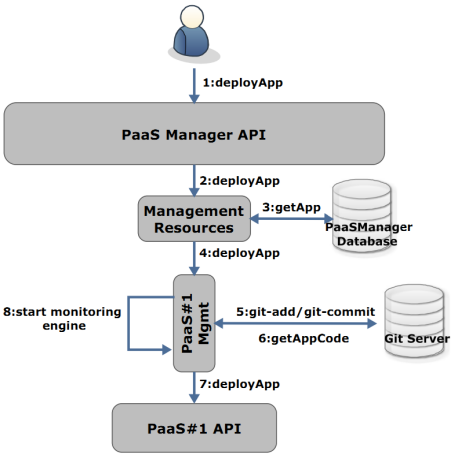


Fig. 2. Deployment Process

Migrate App- The portability of applications is becoming crucial in the developers' point of view. The decision of migrating may be based on seeking a PaaS environment that would give a better performance or a more advantageous business model for the planned activity. The scope of this work does not intend to define any standard model that would be shared by heterogeneous vendors, but rather aims to adapt a solution to what is already exposed by each platform API. Therefore, it requires a prior analysis if the platform to where the application will be migrated supports the required technologies for the application to run properly. From the defined ecosystem, only Heroku allows to retrieve the application source code via Git commands, particularly *git-pull*. However, the remaining PaaS don't support any form of access after the deployment process is accomplished. Hence it was essential to maintain all the applications source codes and respective versions in a central Git server. This approach may question the architecture scalability, although, the deletion of applications performed by the developers will enforce the removal of unnecessary repositories keeping the system clean and efficient. Figure 3 overviews the main steps that are performed during an application migration process.

As observed in Figure 3, the request done towards the PaaSManager API is forwarded to the Management Resources module. The central database is queried in order to return the PaaS identification where the application is hosted, as also the supported framework, e.g. Rails, Spring, CakePHP. After the specific adapter has this information, it analyzes, through the database, whether the new vendor to where the application will be migrated supports the required framework. In the case of correlation, the application code kept in the Git server is deployed in the new platform by invoking the needed operations. The application status in the new PaaS is analyzed and in case of success, triggers the removal of the same application in the previous provider. To complete the migration process, the state information in the central database is updated and the monitoring engine associated to the new platform is started. However, if the application is not running correctly, the adapter activates the removal process keeping the initial application running in the first provider.

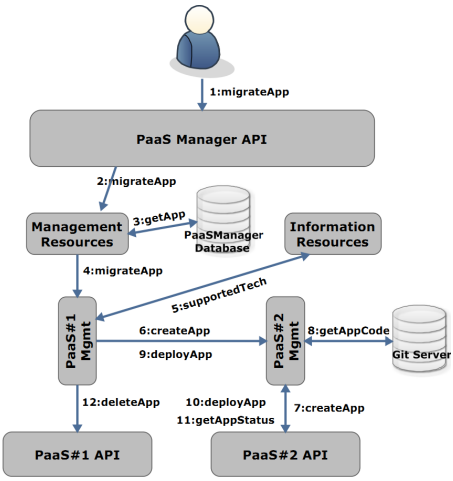


Fig. 3. Migration Process

At the moment, the PaaSManager only supports the migration of applications that not require persistence (databases) due to limitations that could arise in this case. These limitations include synchronizing the entire database data to a central location and the auto-reconfiguration of the application source code files in order to access the newly created database. Nevertheless, the PaaSManager supports an operation that returns the database's access credentials.

4.3 Information Resources

The Information Resources module presented in Figure 1 is a decision module that implements methods related to acquisition of information concerning applications and databases. Likewise the Management Resources module, four PaaS adapters were defined: CloudBees Info, CloudFoundry Info, IronFoundry Info and Heroku Info. These adapters have all the logic that implements the methods related to acquisition of information tasks (get app info, get database info, monitoring, logs, etc.) and exposed by each of the PaaS APIs. The operations specified give all the essential information about the behavior of applications deployed in different platforms environments. Such information can be used by developers to activate the scaling or migration processes.

Get App Status- One of the supported methods gives the possibility to acquire the status of an application that runs on any vendor of the ecosystem. Figure 4 illustrates this process. In order to give a simple understanding for developers, four states were specified: *running*, *stopped*, *crashed* and *unknown*. Firstly, the request made to the PaaSManager API is instantly routed to the Information Resources module, which directs the message to the specific PaaS adapter. Finally, the adapter's logic invokes the platform API operation, mapping the result to one of the states aforementioned.

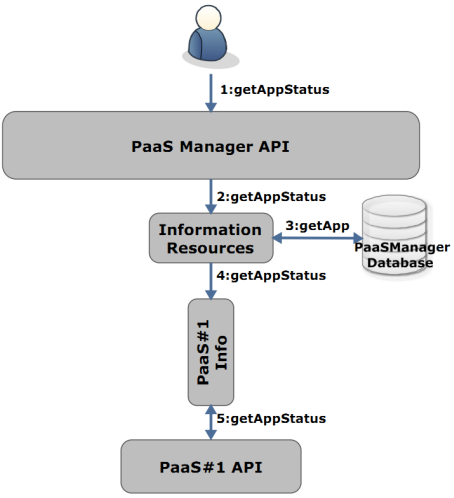


Fig. 4. Get App Status Process

4.4 Monitoring Engine

Monitoring applications give feedback about how a piece of software reacts and behaves in a specific PaaS environment. The module responsible for carrying this process is the Monitoring Engine depicted in Figure 1. So far there is no consensus and no standardized monitoring model shared among the largest providers. In recent years, the studies conducted in this area tried to define monitoring frameworks as well as metrics models which are recognized as appropriate to ensure an efficient management [9]. The metrics list is quite extensive, including availability, response time, RAM, CPU usage, database requests, web transactions, threads, and user's sessions, among others [10]. A monitoring standard would make SLAs more equivalent even in heterogeneous environments. However, the PaaS service model is still in its infancy and therefore each vendor provides different metrics and ways to monitor applications. As discussed in section 3.1, CloudBees and Heroku have partnered with NewRelic, on the other hand, CloudFoundry and IronFoundry provide a native monitoring support. Unlike the remaining methods provided by the PaaSManager, the defined monitoring process is vendor-oriented collecting real-time metrics exhibited by each platform. After the application has been deployed in one of the platforms, a background job is launched and kept alive until the application is stopped or until it is removed. This process is defined by a synchronous sampling performed every minute towards the NewRelic API or the native API according to the vendor where the application is hosted. The achieved information is then stored in the central database and can be queried through the PaaSManager API.

5 Performance Analysis

The PaaSManager system was submitted to several performance and validation tests. This section presents only some illustrative performance analysis tests conducted in order to evaluate some fundamental operations of the devised framework.

The Apache JMeter was the tool used for performing load tests simulating the simultaneously access of several users to the PaaSManager. The main metric analyzed was the response time. Obviously, this metric is highly variable and depends on several factors such as network conditions, the application server, computational resources, etc. However the acquired values should be seen as references of the system behavior and not for comparing directly the supported providers. The machine used for hosting the software and database was an Intel®DualCore™2GHz with 3GB of RAM and Ubuntu 12.04 as operating system. Due to space constraints, from the several performance results obtained only two illustrative cases are presented: the deployment and the get application status processes. Also, each test was divided into two series: *PaaSManager* and *PaaSManager+PaaS API*, which respectively isolate the time consumed only by the PaaSManager internal processing modules, from the time consumed by the internal processing plus the request to the specific PaaS API.

5.1 Deploy App

The deployment process, detailed in Figure 2, was tested with 30 developers inquiring this operation simultaneously. The application used for this evaluation was a SOAP web service with 8MB. In Figure 5, the *PaaSManager* series reveals response time values roughly between 800 and 1200ms which includes the uploading of the application source code and the commit to the created Git repository. On the other hand, the *PaaSManager+PaaS API* series showed a value around 94000ms for CloudBees, in which more than 90% of the response time was only used by the CloudBees native API processing. Moreover, Heroku possessed a smaller time of deployment, about 64000ms, while CloudFoundry and IronFoundry shown lower values, around 6500ms. Consequently, from the observed results the PaaSManager does not include a large overhead in the traditional deployment process.

5.2 Get App Status

The getting application status process, depicted in Figure 4, was also tested with 30 developers inquiring this operation simultaneously. In Figure 6 the *PaaSManager* series shows response time values approximately between 60 and 800ms. This series includes the acquisition of the platform identifier and the message forwarding to the respective adapter. In the *PaaSManager+PaaS API* series, that also includes the request to the vendor native API, CloudFoundry and IronFoundry reveal the highest values around 1300 and 1600ms. In such cases, the adapters perform a log-check for each instance where the application executes in order to detect operating errors. This procedure justifies the increasing of response time comparatively to the getting status process performed explicitly through the native vendor API. However, because this process is not time critical, the overhead does not degrade the user experience.

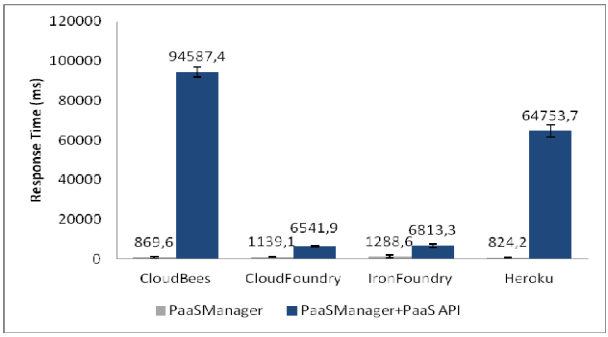


Fig. 5. Deployment Process – Response Time (ms)

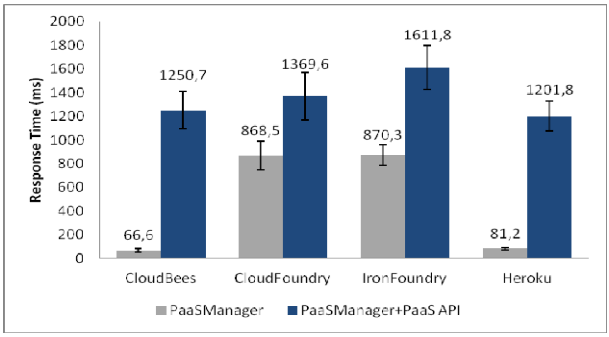


Fig. 6. Get App Status Process – Response Time (ms)

6 Conclusions and Future Work

This work emerges from the current concern in the community about providing interoperability among cloud service providers, namely PaaS providers. With the proposed approach, users would not be longer vulnerable to a particular vendor and their proprietary features getting the opportunity to compare, manage and migrate data between heterogeneous providers. This work provides the first prototype in this research area that aggregates several PaaS offerings in order to abstract the differences between them. Through the PaaSManager a developer has access to the main features of the application lifecycle as well as a seamless portability to another vendor. Also, this solution supports a monitoring vendor-oriented approach that gives feedback about how an application behaves in a PaaS environment. Regarding the performance, some behavioral tests were performed proving the efficiency and the reduced overhead introduced by the framework in real scenarios. At the moment, the PaaSManager is integrated with a Cloud Service Broker that has a rules engine for recommending the suitable platform provider based on each application profile. The architecture also provides a web interface allowing developers to have access to all

the supported features. Further steps include the development of operations for importing and exporting data from databases, as well as the study of techniques for migrating databases between PaaS providers.

Acknowledgments. The work of Pedro Sousa is partially funded by FEDER Funds through the Programa Operacional Fatores de Competitividade COMPETE and by National Funds through the FCT (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-022674.

References

1. Zhang, S., et al.: Cloud Computing Research and Development Trend. In: 2nd International Conference on Future Networks, pp. 93–97. IEEE Computer Society, Washington (2010)
2. Armbrust, M., et al.: A View Of Cloud Computing. Commun. 53, 50–58 (2010)
3. Buyya, R., et al.: Cloud Computing and Emerging IT Platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Gener. Comput. Syst. 25, 599–616 (2009)
4. Buyya, R., et al.: Cloud Computing Principles and Paradigms, pp. 58–62. Wiley Publishing (2011)
5. Loutas, N., et al.: Cloud Computing Interoperability: the state of play. In: 3rd IEEE International Conference on Cloud Computing Technology and Science, pp. 752–757. IEEE Computer Society, Washington (2011)
6. Petcu, D.: Portability and Interoperability between Clouds: Challenges and Case Study. In: Abramowicz, W., Llorente, I.M., Surridge, M., Zisman, A., Vayssière, J. (eds.) ServiceWave 2011. LNCS, vol. 6994, pp. 62–74. Springer, Heidelberg (2011)
7. Beimborn, D., Miletzki, T., Wenzel, S.: Platform-as-a-Service. Business & Information Systems Engineering 3, 381–384 (2011)
8. Fielding, R.T.: Architectural Styles and the Design of Network-based Software Architectures. PhD thesis, University of California (2000)
9. Shao, J., Wang, Q.: A Performance Guarantee Approach for Cloud Computing based on Monitoring. In: 35th IEEE Annual Computer Software and Applications Conference Workshops, pp. 25–30. IEEE Computer Society, Washington (2011)
10. Cheng, X., Yuliang, S., Qingzhong, L.: A Multi-Tenant Oriented Performance Monitoring, Detecting and Scheduling Architecture based on SLA. In: Joint Conferences on Pervasive Computing, pp. 599–604. IEEE Internet Computing, Taipei (2009)

MC64-Cluster: A Many-Core CPU Cluster for Bioinformatics Applications

Francisco J. Esteban[1,*], David Díaz[2], Pilar Hernández[3], Juan A. Caballero[4], Gabriel Dorado[5,**], and Sergio Gálvez[2,**]

[1] Servicio de Informática, Campus Rabanales, Universidad de Córdoba, 14071 Córdoba, Spain
[2] Dep. Lenguajes y Ciencias de la Computación, Boulevard Louis Pasteur 35, Universidad de Málaga, 29071 Málaga, Spain
[3] Instituto de Agricultura Sostenible (IAS – CSIC), Alameda del Obispo s/n, 14080 Córdoba, Spain
[4] Dep. Estadística, Campus Rabanales, Universidad de Córdoba, 14071 Córdoba, Spain
[5] Dep. Bioquímica y Biología Molecular, Campus Rabanales C6-1-E17, Campus de Excelencia Internacional Agroalimentario, Universidad de Córdoba, 14071 Córdoba, Spain
{fjesteban,ma1camoj,bb1dopeg}@uco.es, david.diaz@lcc.uma.es, phernandez@ias.csic.es, galvez@uma.es

Abstract. The current developments in life sciences face a big challenge, with the need of dealing with huge amounts of data and the increasing demand of computational resources, both in hardware and in software, pushing the limits of the available state-of-the-art at an affordable price. This paper introduces a computer cluster whose building blocks are the first commercially available many-core CPU systems: the Tile64 by Tilera Corporation, packed in PCIe cards (TILExpress-20G). We have developed the main software components of the cluster (resource manager and scheduler) and a communication library, in order to offer a high-performance general-purpose platform to speedup bioinformatics applications.

Keywords: Parallel programming, grid and cloud computing, biotechnology, genomics.

1 Introduction

Several network-computing-related paradigms have arisen through the years, including "parallel computing", "grid computing" or "cloud computing". The computer cluster concept was historically the first attempt to execute a program among a group of connected computers and, though it may be considered as a «group

* Corresponding author.
** Authors who contributed to the project leadership.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 819–825.
DOI: 10.1007/978-3-642-36981-0_76

of loosely coupled computers» [1], it is currently associated with a group of homogeneous computers joined by a high-performance private network, available at full-time and delivered under a unique administrative authority.

In order to execute programs, (usually named jobs), a cluster needs: i) a resource manager to gain the appropriate visibility and control on the computing components involved; ii) a task manager/scheduler to control the jobs launched to the system; and iii) other add-ons like user, security or billing managers, when such systems are opened to general users. There are many implementations of such tools, but the most widespread systems in the scientific literature are the Terascale Open-Source Resource and QUEue Manager (TORQUE) [2] as resource manager, as well as the Maui Cluster Scheduler [3] as advanced scheduler. Furthermore, a new software framework is needed for a job to get the most out of the parallel hardware. This includes the network storage and the communication libraries that provide better-distributed execution abilities, as the Message Passing Interface (MPI) [4].

In this work, we take advantage of our previous experiences with the Tile64 processors in the field of bioinformatics [5-9] to design a homogeneous general-purpose cluster, developing the software components for job control. We have been capable to achieve an almost linear scalability in computational power, opening the possibility to deal with a new range of bioinformatics problems that were previously not practical using single Personal Computers (PC).

2 Tile64 Architecture: TILExpress-20G Card as a Cluster Component

The starting point for our development has been the TILExpress-20G, a Peripheral Component Interconnect express (PCIe) card with an integrated Tile64 processor. These cards are inserted into PCIe x8 slots of a host PC and they are managed in a Linux environment. A Tile64 processor contains 64 general-purpose cores (named tiles by the manufacturer) connected by a low-latency network branded as “intelligent Mesh” (iMesh). Each tile can execute a customized Linux kernel, so it behaves as a miniaturized computer. Each TILExpress-20G card includes 8 GB of shared Random Access Memory (RAM) and two 10GBase-CX4 Ethernet connectors. A few dedicated tiles are needed for coordinating tasks, so the real number of available tiles to run applications is reduced to 59 [5], or even to 56 if the 10G Ethernet connectors are involved.

Regarding the software, the programs to be run on the Tile64 processor can be developed with the Eclipse-based Tilera’s Multicore Development Environment (MDE), which includes a C/C++ cross-compiler, so native code for the processor can be generated and later on deployed to the card via the PCIe interface. In order to exploit the parallelism, two main Application Programming Interfaces (API) are provided, named iLib and NetIO. The former allows the programmer to use the iMesh to provide common parallel functionality: task processing, message passing, channel communication and shared memory. The latter gives access to the Ethernet ports, and a higher level socket-based driver is also supplied. Unfortunately, none of these

libraries are based on a widely-used standard like the Message Passing Interface (MPI) [4] or the Parallel Virtual Machine (PVM) [10].

The built-in Ethernet ports of the TILExpress-20G cards allows to build a high-performance cluster, whose building blocks are such cards, exploiting the high bandwidth and low latency that a dual 10 gigabits per-second (Gbit/s) connection can provide. As stated before, the ports built in the cards use 10GBase-CX4 cables, internally composed by four coaxial wires. This was the first 10 Gbit Ethernet (GbE) implementation [11].

On the other hand, the built-in cross-over functionality allows the direct connections between the cards without requiring a dedicated and expensive 10GBase-CX4 switch. The two ports in each card allow two possible cluster connection topologies with 20 Gbit/s of aggregated speed: i) linear topology, with just two hosts connected to each other using the two ports in each card; and ii) ring topology, with each host connected to two neighbors via the two ports in each card.

3 MC64-Cluster General Scheme

Once the network topology was setup, we carried out the analyses of the software components needed to operate it as a cluster (MC64-Cluster). We grouped the management tasks required by the cluster into a single component: a Resource and Task Manager. We used the TORQUE/PBS [2] as a model, adopting a general client/server scheme with a master-server process running on a tile of a master card, and a slave-server process running on a tile of each of the remaining cards. In this scheme, the master-server daemon receives commands from the client utilities, issues them and replies with a message. This daemon is the focus of the section 3.1.

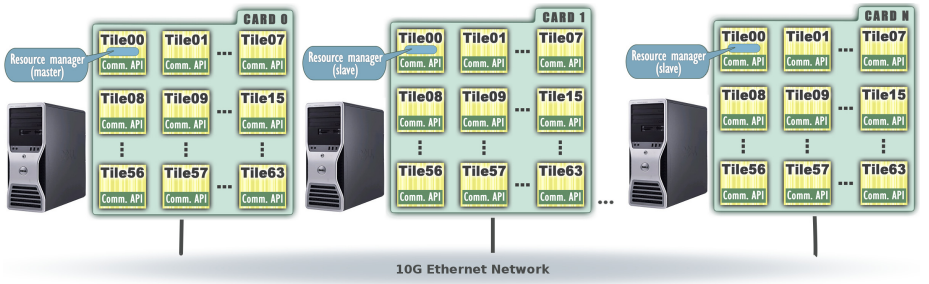


Fig. 1. Software component scheme. Distribution of resource manager and communication API in the Tile 64 cards.

The other major component evaluated during the analyses was the communication library, so that the programs running in the cluster can gain full advantage of the parallel execution. Our model for this component has been the MPI for a message-passing set of functions, and the iLib for a connection-oriented one (see section 3.2 for more detail). The Figure 1 shows the software components in a typical cluster implementation, as would be distributed through the different tiles of each TILExpress-20G card.

3.1 Resource Manager/Scheduler implementation

In a cluster environment, a Resource and Task Manager is in charge of defining the cluster's geometry, knowing the resources states, executing the jobs sent by the users, distributing them through the available processors and controlling the execution state and stages. In our design, these functions are achieved following a client/server approach, having one server process in each card that receives commands from the client tools and executes them. The server process in the first card acts as a master-server, as it receives requests from the client utilities and, if necessary, forwards secondary requests to the server processes in the remaining cards (acting as clients), in order to complete such requests. Furthermore, the master process provides coordination and information services to the jobs through the communication library. Finally, the master process also acts autonomously, inspecting the system-resource state and attending the previously scheduled tasks.

The communication between the clients and the server is accomplished by means of a Transmission Control Protocol (TCP) session, in which the first packet sent by the clients contains a command code, followed by its applicable options and, eventually, the required data for the command completion. In a simple command, the master replies to this initial packet with a straightforward result. In a more complex one, a dialog is established between the clients and the server, until the server has enough information to issue the requested command. Thus, in a job-execution command from a master server to a slave server, it is necessary to transfer the files associated with the job to the slave card before the job can be executed on it.

The main commands available in a MC64 Cluster and their functions are:

- `MC64_cstat`: View the status of the tiles in a MC64 Cluster.
- `MC64_jsub`: Submit a new job to a MC64 Cluster.
- `MC64_jstat`: View a job status in a MC64 Cluster.
- `MC64_jdel`: Cancel a job in a MC64 Cluster.
- `MC64_cshed`: Alter the scheduling strategy in a MC64 Cluster.

3.2 Communication Library implementation

Once the MC64-Cluster is setup, the tiles involved in a job must communicate between themselves. Although the Linux kernel executed by the tiles has some built-in standard capabilities like the UNIX sockets, TCP sockets and semaphores, we were interested in developing an MPI-like library layer supported by an enhancement of the Tile64-provided communication libraries. In this scenario, three major areas were identified: i) the connection-oriented communication; ii) the message-passing; and iii) the execution coordination. In the second and third areas, we followed the MPI syntax and interface, in order to give the programmers a chance to use the known functions and resources. In the first one, we resembled the Tilera's iLib interface. The Table 1 shows some functions of this API, grouped by category.

Table 1. Sample of the API functions of the MC64-Cluster communication

Group	Function	Description
Connection-oriented	`CCI_bufchan_connect()`	Define a buffered channel between a sender and a receiver
	`CCI_bufchan_open_sender()`	Open a buffered-channel send port
	`CCI_bufchan_send()`	Send a packet of data into a buffered channel
Message-passing	`MPI_Send()`	Perform a standard-mode blocking send
	`MPI_Recv()`	Perform a standard-mode blocking receive
	`MPI_Bcast()`	Broadcast a message to all the other processes
Execution coordination	`MPI_Init()`	Initialize the cluster-execution environment
	`MPI_Group_rank()`	Return the unique identifier of the calling process in a given group
	`MPI_Barrier()`	Stop the execution until all processes in a group reach a specified point

The MC64-Cluster API provides these functions independently of the number of cards interconnected and the location of each executing tile. However, regarding the implementation details, the native iLib functions have been used when the tiles involved in the communication processes were in the same physical card, extending this functionality with the TCP/IP communication socket when the tiles belong to different cards. To coordinate this process, an occasional communication with the master process was necessary, especially in connection-oriented and collective communications. The Figure 2 shows the general communication scheme between the processes and the strategies implemented in each case.

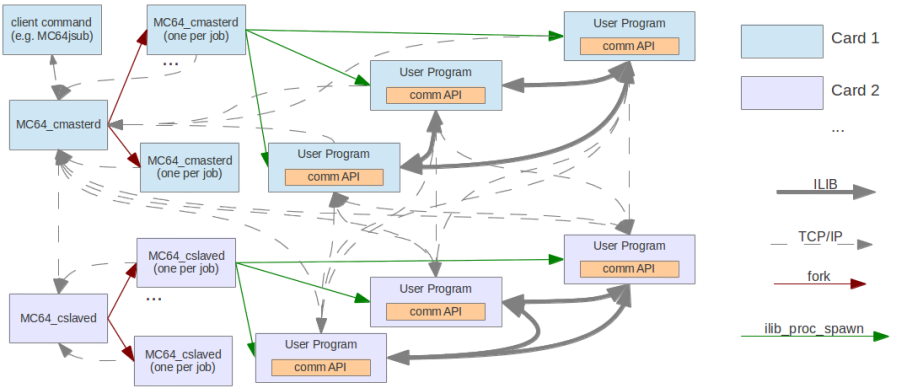


Fig. 2. Communication strategies used in the MC64-Cluster. The main master command uses TCP/IP to connect to any other process. The user processes use API functions that, in turn, use either the Tilera's iLib or TCP/IP, depending on the partner placement in the same or another card, respectively. New processes are launched using the Unix's fork() or the Tilera's ilib_proc_spawn() functions.

4 Preliminary Results and Conclusions

We have tested the MC64-Cluster design with a set of programs involving communications between the tiles of the same and different cards, pairwise and broadcast, connection-oriented and message-passing based on both short-length messages and larger ones, involving just a few tiles, as well as with all available ones. The Table 2 shows the mean execution times of a hundred simple message-passing operations with 48 tiles involved, where the first 24 ones send messages to the other 24. As can be observed, remarkable difference in performance between local and remote communications was found.

Table 2. Message-passing operation execution time (milliseconds) under different situations. Every communication is local, mixed or remote, for columns 48/0, 24/24 and others, respectively.

Length (bytes)	Local/remote			
	48/0	**40/8**	**32/16**	**24/24**
10	0.795	56.307	162.672	366.131
10,000	4.509	96.485	303.085	680.254

Our currently developed bioinformatics algorithms obtain a significant benefit from this cluster, due to the memory hierarchy of the Tile64 processor. Indeed, additional tests on real workflows have demonstrated a very low impact on performance loss in a parallel share-memory environment. As a preliminary test, we have used the MC64-NW/SW (bioinformatics parallel algorithm for pairwise sequence alignment) [7] to predict its execution time in the cluster. To align two sequences of 100 kilobases (kb), the original version of this algorithm was executed in 18.87 seconds using 59 tiles in a non-clustered environment. Then, 77.5 milliseconds are spent in communications and data transfer from shared memory to local memory, in order to process 17,956 chunks of data. The communications were extremely fast, with each message containing a single pointer to shared-memory. However, processing the data in the local memory is advisable, being much faster than working with shared-memory (there is no need of data-coherence tasks). Thus, the time taken by a memory transfer from the shared-memory to the local one must be considered when calculating the time required for preparing data to be processed. Fortunately, this transfer step is not needed in a cluster environment with a message-passing API, since the data is sent inside the message and, thus, directly copied into the receiver's local memory. In a priori evaluations, a clustered version of this algorithm using the same execution conditions, with two TilExpress-20G cards with 108 tiles involved (54 in each card) would spend an average time of 272.05 ms in communications per tile, being just twice the time taken by the non-clustered version. The clustered version of the MC64-NW/SW takes 10.54 s in aligning the same sequences under the same circumstances (i.e., an almost-linear scalability).

In conclusion, we have reached our initial design goal: to extend the capabilities of the Tile64 architecture beyond its local limit of 64 tiles. In a recent work [8], we have analyzed the different performance that can be obtained when porting bioinformatics

algorithms to the Tile64 architecture. We are now focusing on the enhanced implementation of these algorithms on this new cluster, with the aim to extend the prior capabilities of the Tile64 microprocessors. Our work in progress also includes the development of a web portal to allow life scientists to remotely use such system.

Acknowledgments. We are grateful to Tilera for providing hardware and software tools <http://www.tilera.com>. This work was supported by "Ministerio de Economía y Competitividad" [MINECO grants AGL2010-17316 and BIO2011-15237]; "Consejería de Agricultura y Pesca" of "Junta de Andalucía" [041/C/2007, 75/C/2009 & 56/C/2010]; "Grupo PAI" [AGR-248]; and "Universidad de Córdoba" ["Ayuda a Grupos"], Spain.

References

1. LinuxHPC.org. Cluster Encyclopedia (2010), `http://www.clusterbuilder.org/encyclopedia/alphabetized.php` (cited November 4, 2012)
2. Adaptative-Computing. TORQUE Resource Manager (2012), `http://www.adaptivecomputing.com/products/open-source/torque` (cited November 4, 2012)
3. Cluster-Resources. Maui Cluster Scheduler (2012), `http://www.clusterresources.com/pages/products/maui-cluster-scheduler.php` (cited November 4, 2012)
4. Gropp, W., Lusk, E., Skjellum, A.: Using MPI: Portable Parallel Programming with the Message Passing Interface. MIT Press (1994)
5. Gálvez, S., et al.: Next-Generation Bioinformatics: Using Many-Core Processor Architecture to Develop a Web Service for Sequence Alignment. Bioinformatics 26(5), 683–686 (2010)
6. Esteban, F.J., Díaz, D., Hernández, P., Caballero, J.A., Dorado, G., Gálvez, S.: MC64: A Web Platform to Test Bioinformatics Algorithms in a Many-Core Architecture. In: Rocha, M.P., Rodríguez, J.M.C., Fdez-Riverola, F., Valencia, A., et al. (eds.) PACBB 2011. AISC, vol. 93, pp. 9–16. Springer, Heidelberg (2011)
7. Diaz, D., et al.: Parallelizing and optimizing a bioinformatics pairwise sequence alignment algorithm for many-core architecture. Parallel Computing 37(4-5), 244–259 (2011)
8. Esteban, F.J., et al.: Direct approaches to exploit many-core architecture in bioinformatics. Future Generation Computer System 29(1), 15–26 (2013)
9. Esteban, F.J., Díaz, D., Hernández, P., Caballero, J.A., Dorado, G., Gálvez, S.: Many-Core Processor Bioinformatics and Next-Generation Sequencing. In: Liñán Reyes, M., Flores Arias, J.M., González de la Rosa, J.J., Langer, J., Bellido Outeiriño, F.J., Moreno-Munñoz, A., et al. (eds.) IT Revolutions. LNICST, vol. 82, pp. 172–188. Springer, Heidelberg (2012)
10. Sunderam, V.S.: PVM: a framework for parallel distributed computing. Concurrency: Pract. Exper. 2(4), 315–339 (1990)
11. IEEE Standard for IT-Telecom. and Information Exchange Between Systems-LAN and MAN-Specific Requirements. Part 3: CSMA With Method and Physical Layer Specifications. Amendment: Physical Layer and Management Parameters for 10 Gb/S Operation, Type 10 GBASE-CX4. IEEE Std 802.3ak-2004 (Amendment to IEEE Std 802.3-2002 as amended by IEEE Stds 802.3ae-2002, 802.3af-2003 and 802.3aj-2003), p. 0_1 (2004)

Self-Portrait Images for Mobile Application

Teh Phoey Lee[1], Nael Kabbany[2], and Chan Kei Jun[2]

[1] Department of Information System,
Faculty of Science and Technology, Sunway Univeristy,
46150, Selangor Darul Ehsan, Malaysia
[2] School of Information Technology,
Faculty of Business and Infromation Science,
56000, UCSI University, Cheras, Malaysia
phoeyleet@sunway.edu.my, naelk@ucsi.edu.my,
keijun@hotmail.com

Abstract. Self-portrait using mobile phones has been a popular trend among backpackers, teenagers and travellers. The usage of mobile devices had been escalating over the years, as it is more convenient and portable. However, there were constraints in terms of camera view and angle the mobile devices can capture. Based on my preliminary research, the major problems faced by most backpackers concerns (i) hand distance, (ii) angle and view captured, and (iii) no alternative mechanism or functions in mobile devices that can capture a larger panoramic view in self-portrait. This proposal intends to study the current strategies used by travelers in taking self-portrait by using mobile devices and investigate future mobile image application. The subjects in this study include travellers (backpackers, people travelling alone) and social network users as viewers. The evaluation is to find out if a model of "destination image measurement" could be an attribute towards enhancing mobile application in the future. The concept of destination image measurement has four-distinct components; functional characteristics, psychological characteristics, holistic and attributes. The framework is to establish a strategy to improve self-portrait capturing using mobile devices and the evaluation outcome will be a measurement technique in capturing a more effective self-portrait.

Keywords: Self-portrait, Mobile Application, Image Measurement, Mobile System.

1 Introduction

Backpackers often labeled as a group of tourists who undertakes extended and multi-purpose trips that combines work, study and leisure into a travel experience [1]. They travel around the world independently with a low budget and spend minimum cost if possible. When they travel, it is important that they capture all the places and events of their journey and translating the best moments into lifetime experiences. Although many travelers have knowledge about basic photography, there was limited research on areas to improve self-portrait photography using mobile devices. According to

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 827–838.
DOI: 10.1007/978-3-642-36981-0_77

Pearce [2], the camera was one of the most popular technology device used by Australian backpackers. In addition, Jenkins [1] research on 30 backpackers provided evidence indicating nature and sceneries are the most favored type of photography captured during their travels. They also share the pictures of their own travel with friends and relatives on Facebook, etc.

Mobile devices are incorporated with functions of camera for sometimes. Photos captures are being done much simpler with lesser effort. It appears that with a camera always at hand, as well as the technology of easy viewing, uploading, and sharing, photo-taking becomes for many a frequent, even daily, activity [3].

The camera technology in mobile devices is improving. Camera phones represent the culmination of a trend in which the low incremental cost of digital imagery, both in terms of economics and effort Aoki [4]. Therefore, carrying a mobile device todays is the most sufficient to capture the best moment by travelers.

Being a backpacker, taking a self- portrait happens at all times. Self-image refers to capturing images of own self. In another words, a picture images the backpacker as the subject and a background at the back of the subject as scenery. Digital images combined with ICTs increase the choices available to users for, other things, our four social uses: memory construction, communication, self-expression, and self-representation House [5]. Self-portrait is one of the ways of backpacker's self-expression. Self-expression is about giving voice to one's unique view of the world [3].

Sending photos to distant others is a way of keeping up on one another's lives. Telling stories about photos helps nurture relationships as well [3]. Given the rise in social networking platforms, it is not a surprise to see backpackers uploading self-portrait images onto the sites and shares them with their friends and families or even the whole world.

The main purpose of this paper is firstly to gain an insight of the feasible strategies on self-imaging on mobile devices. Many available mobile applications are able to support panoramic photographs shooting with seamless connection. However, the image of a self-taken panoramic photo from a nodal point on curving frames has led to the challenges of connection on stitching line; that lead to the human face to be distorted. To proposed strategies; illustrate the outcome with photo editing technique using self-images and at last obtaining the evaluation from the photo viewers. Thus, section 2 discusses the related strategies; section 3 explains methodology, section 4 analyses with the feedback using structure measurement, lastly conclusion and end with future works.

2 Related Works

Jenkins [1] defined the term 'backpackers' as a group of tourists that undertakes extended, often multi-purpose trips combining work, study and leisure into a travel experience. He further added that a backpacker's motivation for travel involves meeting other travelers and local people, absorbing the culture of a destination, as well as seeing and experiencing different environments. Most businessmen on

working trip would have wanted to take an image of them when they travel alone. What they normally used would be their mobile device.

Pearce [2] pinned pointed that the technology that Australian backpackers used the most was camera. Jenkins [1] once interviewed 30 backpackers, they responded that most favored type of photograph was the natural scenery, and the most common reasons for taking photographs were to capture memories and to show people. It is because the photographic record provides a point of validation, proof of the visit to show people at home and to trigger personal memories. In short, most of the backpackers nowasdays does share their photos onto social network platform to triggers their personal whereabout with mobile platform.

Mobile phones are not only efficient communication tools, but also capable computational devices equipped with high-resolution digital cameras, high-quality color displays, and GPU hardware [6]. It appears that with a camera always at hand, the technology of easy viewing, uploading, and sharing of photos has becomes our frequent event and daily activities [3]. Camera phones represent the culmination of a trend, which is low incremental cost of digital imagery, both in terms of economics and effort [4]. Digital images as well combined with ICTs increase the choices available for our four social uses: memory construction, communication, self-expression, and self-representation [5]. Self-expression is about giving voice to one's unique view of the world [3].

2.1 Destination Images

Image is the mental structures (visual or not) that integrate the elements, the impressions and the values that people have about a specific place; and, it is based on a series of more or less material knowledge of those people, and on a series of perceptions of a more emotional and effective nature [7]. Echtner & Ritchie (2003) defined destination image as frequently described as simply "impressions of a place" or "perceptions of an area". Destination image consists of functional characteristics and psychological characteristics. Functional characteristic referred to the more tangible aspects of the destination, and psychological characteristics referred to the intangible aspects.

2.2 Video Panorama and Mosaic

Video panorama is one of the techniques that able to get most coverage area in an image for a scenery view. The term "panorama" typically refers to single-viewpoint panoramas, which is created by rotating a camera around its optical center [9]. In year 1999, Peleg & Ben-Ezra referred panorama as an image having a wide field of view, up to a full 360 degrees and a panorama can be viewed as a projection of the scene onto a cylinder or a sphere through this viewpoint. Luhmann [10] referred panorama in cartography, as the techniques used for centuries to map three-dimensional objects into a two- dimensional drawing. He expressed that it is part of natural human behavior to observe the environment not only through stereoscopic viewing but also all around view over the partial or total horizon.

The need to combine pictures into panoramic mosaics existed since the beginning of photography, as the camera's field of view is always smaller than the human field of view. A more common solution is photo-mosaicing: aligning, and pasting, frames in a video sequence, which enables a more complete view[11]. Therefore, by registering all the frames of a video into a panoramic mosaic (as opposed to just select frames of the video) it enables many new creative possibilities [12]. Another advantage is the accuracy of stitching video frames is higher than the still photos[13]. Xiong [6] claimed a panoramic image has a wide field of view, much wider than is available on normal cameras such as those in mobile phones.

Another term used is mosaicking, which is referring to the process of seamlessly stitching a set of overlapping images of a scene into a large and wide view image. When the input images are the frames of a video, the process is called video mosaicing Hoseini [14].

2.2.1 The Process of Panorama and Mosaic

According to Hoseini [14], mosaicing consist of direct methods and feature-based methods. Direct methods use information from all pixels and discover parameter set through an iterative process to minimize the sum of squared difference (SSD). Meanwhile, these feature-based methods assume that feature correspondence between image pairs are available and ultilizae these correspondences to find transforms which register the image pair.

Brown [15] uses object recognition techniques based on invariant local features to select matching images, and a probabilistic model for verification. They implemented Lowe's Scale Invariant Feature Transform (SIFT features), which are geometrically invariant under similarity transforms and invariant under affine changes in intensity. They introduced a probabilistic model for image match verification. This gives them a principled framework for rejecting noise images and recognizing multiple panoramas in an unordered image set.

During the year of 1996, Szeliski [16] developed techniques that are based on the ability to align different pieces of a scene (tiles) into a larger picture of the scene (mosaic) and then seamlessly blend the images together. Next, Steedly [17] presented an automatic and efficient method to register and stitch thousands of video frames into a large panoramic mosaic. Their method preserves the robustness and accuracy of image stitches that match all pairs of images while utilizing the ordering information provided by video. Their approach able to substantially reduce the time required to estimate the image orientations with minimal loss of accuracy.

Xiong and Pulli [6] stated a panorama construction process requires a lot of computation and memory and at the same time mobile phones only have limited resources. They knew that it is necessary to develop efficient stitching methods to fit mobile applications. Figure 1 shows the process of the fast panorama procedure. Lin [13] explains the process of video to photo as shown in Figure 2. His process started with motion analyisis and clustering, followed with intent determination, and with the capturing, exttrated key frames to generate the panorama. Mean while, Deng and Zhang's two pass scheme [18] showing the basic processing scheme on two phases:

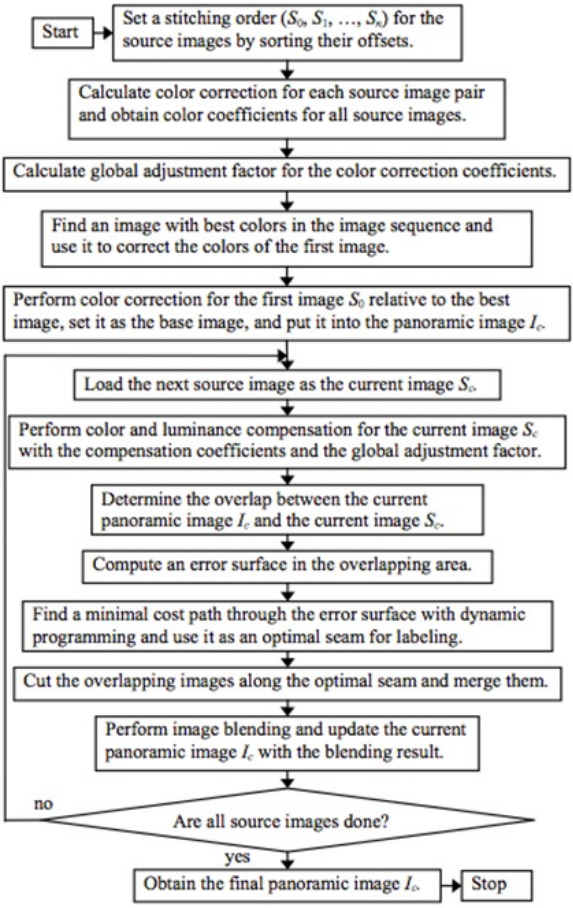


Fig. 1. The process of fast panorama stitching approach [6]

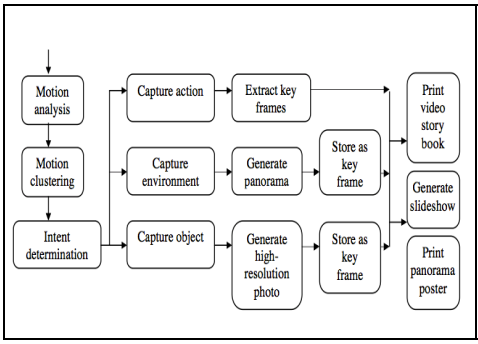


Fig. 2. Process for video to photo [13]

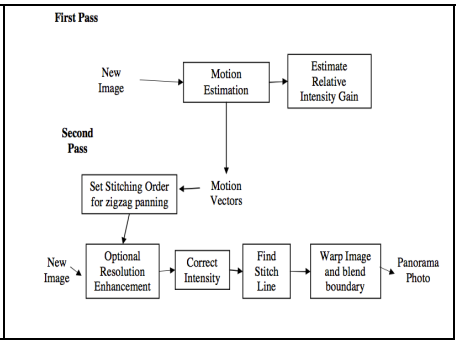


Fig. 3. Process of Two-pass schemes to generate a panorama photo [18]

the alignment phase and the stitching phase. During the alignment phase, motion estimation is performed on each video frame and subsets of frames with significant motion shifts in between are selected. During the stitching phase, these select frames are stitched together to form a panorama photo.

2.3 Image Composition

According to Pritch [19], image composition is common in digital image editing, whose objective is to combine images from different shots into a single output image that looks natural and realistic. They also stated that there are three approaches that are common for image composition which are matting, blending, and optimal cuts. Jia [20] looks at image composition as the process of creating a new image by pasting an object or a region from a source image onto a target image. In Rankov [21] research, they claimed that image stitching able to combine a number of images taken at high resolution into a composite image. Therefore, they listed the two main stages in processing these images which included image stitching and image blending.

2.3.1 Image Matting

Pritch [19] said matting is an attempt to make an accurate segmentation of an object, allowing it to place it in a new image. Meanwhile Singaraju and Vidal [22] refers image matting to the problem of assigning to each pixel in an image, a probabilistic measure of whether it belongs to a desired object or not. Year 2006, Jia [23] claim that image matting is a common way to extract an object from a source image, which is to be pasted onto a target image naturally using an alpha channel.

Chuang [24] said in digital matting, a foreground element is extracted from a background image by estimating a color and opacity for the foreground element at each pixel. They also said matting is used in order to composite the foreground element into a new scene. Chuang [24] also have developed a Bayesian approach to solve several image matting problems: constant-color matting, difference matting, and natural image matting.

2.3.2 Image Stitching

Image stitching is a common practice in the generation of panoramic images and applications such as object insertion, super resolution and texture synthesis. The aim of a stitching algorithm is to produce a visually plausible mosaic with two desirable properties: First, the mosaic should be as similar as possible to the input images, both geometrically and photo metrically. Second, the seam between the stitched images should be invisible [25].

Jia & Tang [26] claimed that in order to obtain a satisfactory results in image stitching, a natural transition from one image to another is required, where both structure and intensity should be aligned or matched within, or possibly beyond , the overlapped area. In their experiment, there are 3 properties which are in general agreement with their visual perception which are, structure preservation, intensity alignment and image context consideration. Rankov et al. [21] perform their image processing stages which are image stitching by sliding the new image over the composite image and finding the best cross-correlation point.

2.3.3 Image Blending

Image blending is part of image composition process. Pritch *et. al.* [19] refers image blending as building a new image from patches taken from the input images, and the seams between these patches are eliminated by the blending. Rankov et al. [21] proves that application of Image blending is necessary to remove the edges and make one compact image. They also said that the process of image blending is restricted to zone of overlap which is determined during the stitching process. They apply the blending process in the second stage of image processing is done by separating color planes, applying blending algorithm for each color band and recomposing planes together to get full color image at the output. They applied the gradient blending method in their research that gives advantage to them in improving visual quality of the composite image and making the edges invisible.

2.4 Image Qualities and Measurement

The measurement of image quality is crucial to many image and video processing systems Sheikh [27]. Due to inherent of physical limitations and economic reasons, the quality of images and videos could visibly degrade right from the point when they are captured to the point and viewed by a human observer. Wang [28] stated that the objective of image quality assessment aims to design quality measures that can automatically predict the perceived image quality. According to them, quality measures play an important roles in a broad range such as in image acquisition, compression, communication, restoration, enhancement, analysis, display, printing and watermarking. Thus the framework of image measurement should considered the above.

Besides, digital image and video processing systems are generally involved with signals that are meant to convey reproductions of visual information for 'human consumption'. The obvious way of measuring quality is to solicit the opinion of human observers [27]. Echtner & Ritchie [29] did a review in their research on the two basic approaches in product image measurement; structured and unstructured. In a structured methodology, various common image attributes are specified and incorporated into a standardised instrument, usually in a set of semantic differential or Likert type scales. Thus, structured methodologies are attribute focused. In other words, the respondent think about product image in terms of the attributes specified by the scales. Scale items are not designed to measure the unique characteristics of the product. Rather, the respondent are requested to rate the product on more general, common traits. Figure 4 shows the components of the model.

The structured methodologies are easy to administer, simple to code and the results can be analysed using sophisticated statistical techniques, it uses standardised scales [29]. Structured methodologies facilitate the comparison of several products across each of the attributes as scale items. Meanwhile, unstructured methodologies uses a free form descriptions to measure image [29]. With the approach, the attributes of image are not specified at the onset of the research. The respondent is allowed to freely describe his/her impressions of a case. Next section explains how the data are collected.

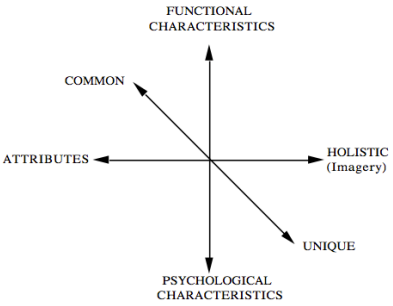


Fig. 4. Components of destination image [29]

3 Methodology

3.1 Data Collection

A preliminary study is conducted by distributing the illustrated comparison image to 200 respondent using the qualtrics web survey system. The resutls mainly collected from respondents who had experienced travelling and taking self-portrait. The reason is to understand the current self-portrait trend among the mobile user as well as the conparison of strategies used on the resulting destination images. The participants consisted of those usally access to social network platform as the main source for their information update and friend's information updates.

3.2 Questionaire Design

Based on the components of the destination image by Echtner & Ritchie (2003), the variables in the questionaire are designed into four dimensions as following;

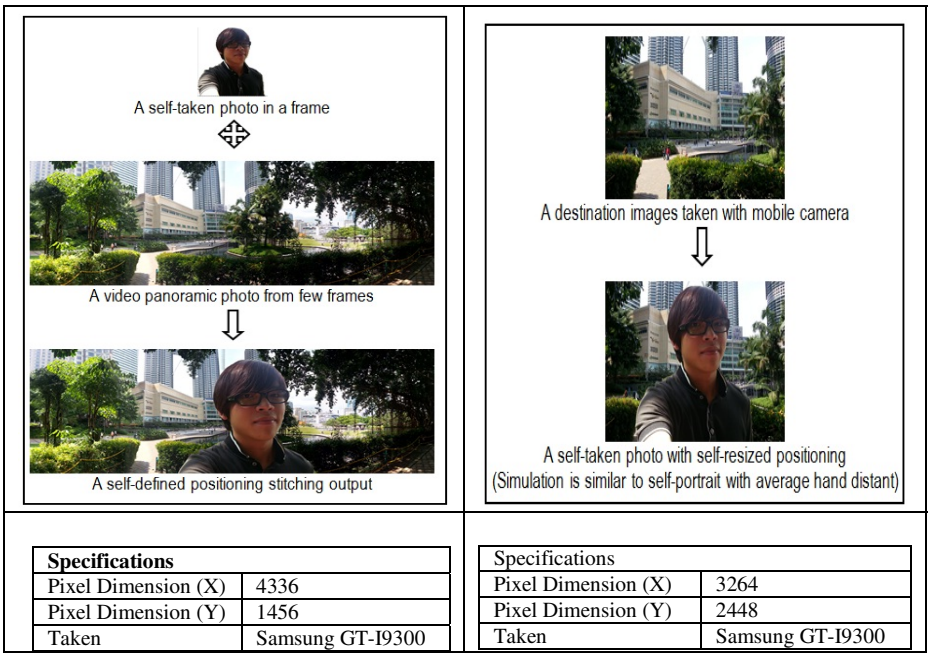
- Attributes vs Functional Characteristics.
- Attributes vs Phychologicl Characteristic.
- Holistic vs Functional Characteristic, and
- Holistic vs Psychological Characteristics.

The 5 likert scale-points are used to measure the respondents interpretation, by comparing the two of destination images as illustrated on the next section.

3.3 Conceptual Development

3.3.1 Panorama and Composition Image Illustration

Figure 5 has presented an illustration of panorama images. The application is suggested with a selection of sizing function for instant allocation of a self-portrait that allows seamlessly share function to social network. The reason of this illustration is to produce an illustration for the use of the exploratory study from the perspective of image viewer and traveler.



Specifications	
Pixel Dimension (X)	4336
Pixel Dimension (Y)	1456
Taken	Samsung GT-I9300

Specifications	
Pixel Dimension (X)	3264
Pixel Dimension (Y)	2448
Taken	Samsung GT-I9300

Fig. 5. Process of Composing Panorama Image Illustration and Image Composition Illustration

This image is used as the illustration for the preliminary study due to the current technology of panoramic photo that produced a distorted image. The suggestion for the physical development would contain a clear human face as in Figure 5 the "self-defined position stitching output". The physical development should come with the suggested process, including with share function to social network, instant positioning of the human face with an indicator, handy and small image stitching processing time, and a column for the backpacker to express their whereabouts in just one click or two clicks.

During the composing image composition illustration, one original image is taken, and the backpacker's face is cropped from another image according to the selection lines and placed into the destination image. This image composition would only represent the normal two frames self-portrait. Why do we need to have two frames instead of one self-portrait of image like we used to have taken in one shot? The reason behind is the benefit of instant position of human face that able to resize ac-hoc just after the photo taken in the proposed physical application, it allows user to show a bigger area of the background and reduce the "big-head" human face problem that blocking the background. Thus with the two images illustrated, a part of the preliminary study is presented in next section.

4 Results

Figure 6 shown an example of a distorted panoramic image of self-portrait using current available technology. Table 1 shown 89% (177 out of 200) respondents has

their self-portrait taken, although only 91 out of 200 (46%) of them are backpacker, it shown that they take self-portrait images not only when they travel, but also during their normal leisure hour.

Fig. 6. Example of Distorted Panorama Image when self-portrait

Table 1. Self-portrait while travel

		Do you usually take self-portrait while travelling?		Total
		Yes	No	
Are you a backpacker?	Yes	87	4	91
	No	90	19	109
Total		177	23	200

Table 2. Central tendencies measurement of personal views towards Panorama and composition images

Description	**Panorama**	**Composition**
I can understand the message send through the image.	3.71	3.06
I can see most of the background area clearly.	4.44	2.76
I can see the backpacker's face clearly.	4.48	4.57
I'm able to identify the location of the background.	4.34	3.16
The wider area of background makes me able to identify the background instantly.	4.23	3.13
The backpacker's face blocked the important elements of the image.	2.79	4.41

Table 2 concluded that 1) Panorama image sends a better message to image viewer. 2) Panorama image show a wider area of the background to viewer.3) only a small difference (0.09 of mean) on the clarification of backpacker's face and this can be improved by the function of self-image resizing on spot. 4) Respondent able to identify the destination of the image once the par anomic photo is shown.

5 Conclusion and Future Works

The results data collected from the questionnaire of the preliminary study are partly presented above. The results firstly concluded that most of the respondents are

looking positive towards the paronomic photo when comparing to normal image composition based on the survey outcome. Therefore, a future work on the solution of the stiching line problem on the self-portrait with paronomic images are important. The study of nodal point and the curving of paronomic photo should be discussed. The measurement technique of the destination images model is suggested to be included for the use of mobile application measurement, expecially in terms of the trade-off in the size on the mobile device to support the process for the illustration above.

References

[1] Jenkins, O.H.: Photography and travel brochures: The circle of representation. Tourism Geographies 5, 305–328 (2003)
[2] Pearce, P.L., Murphy, L., Brymer, E.: Evolution of the backpacker market and the potential for Autralian tourism, pp. 1–2. CRC for Sustainable Tourism Pty Ltd., Australia (2009)
[3] House, N.V., Davis, M., Ames, M., Finn, M., Viswanathan, V.: The Uses of Personal Networked Digital Imaging: An Empirical Study of Cameraphone Photos and Sharing. In: CHI EA 2005, pp. 1853–1856 (2005)
[4] Aoki, P.M., Szymanski, M.H., Woodruff, A.: Turning from Image Sharing to Experience Sharing. In: Workshop on Pervasive Image Capture and Sharing, 7th Int'l Conf. on Ubiquitous Computing, Ubicomp 2005, pp. 1–3 (2005)
[5] House, N.A.: Distant Closeness: Cameraphones and Public Image Sharing. In: CHI EA 2007, pp. 2717–2722 (2007)
[6] Xiong, Y., Pulli, K.: Fast Panorama Stitching for High-Quality Panoramic Images on Mobile Phones. IEEE Transaction on Consumer Electronics 56, 1–6 (2010)
[7] Marino, E.D.: The Strategic Dimension of Destination Image. An Analysis of the French Riveira Image From the Italian Tourist's' Perceptions. PhD. Tourism Management. University of Naples "Fredrico II" Faculty of Economics (2008)
[8] Echtner, C.M., Ritchie, J.B.: The Meaning and Measurement of Destination Image. The Journal of Tourism Studies 14, 37–48 (2003)
[9] Szeliski, R., Shum, H.-Y.: Creating full view panoramic image mosaics and environment maps. In: ACM Conference Proceedings, pp. 251–258 (1997)
[10] Luhmann, T.: A Histrorical Review on Panorama Photogrammetry. International Archives of Photogrammetry and Remote Sensing 34, 1–6 (2004)
[11] Saravanan, P., Narayanan, C.K., Prakash, P.V.S.S., Prabhakara Rao, G.V.: Techniques for Video Mosaicing 5, 286–289 (2005)
[12] Steedly, D., Pal, C., Szeliski, R.: Efficiently Registering Video into Panoramic Mosaics. In: Proceedings of the Tenth IEEE International Conference on Computer Vision, vol. 2, pp. 1300–1307 (2005)
[13] Lin, Q., Zhang, T., Chen, M., Deng, Y., Obrador, P.: Mining Home Video for Photos (2004)
[14] Hoseini, S.A., Jafari, S.: An automated method for mosaicing of video frames with projective constraint. International Journal of Science and Advanced Technology 1, 112–116 (2011)
[15] Brown, M., Lowe, D.G.: Recognising Panoramas. In: Proceedings of the Ninth IEEE International Conference on Computer Vision, ICCV 2003, vol. 2, pp. 1218–1227 (2003)

[16] Szeliski, R.: Video Mosaics for Virtual Environments. IEEE Computer Graphics and Application 16, 22–30 (1996)
[17] Steedly, D., Pal, C., Szeliski, R.: Efficiently Registering Video into Panoramic Mosaics. In: Proceedings of the Tenth IEEE International Conference on Computer Vision, ICCV 2005, vol. 2, pp. 1300–1307 (2005)
[18] Deng, Y., Zhang, T.: Generating Panorama Photos. In: Proc. of SPIE Internet Multimedia Management Systems IV, vol. 5242, pp. 270–279 (2003)
[19] Pritch, Y., Poleg, Y., Peleg, S.: Snap Image Composition. In: Gagalowicz, A., Philips, W. (eds.) MIRAGE 2011. LNCS, vol. 6930, pp. 181–191. Springer, Heidelberg (2011)
[20] Jia, J., Tang, C.-K.: Image Stitching Using Structure Deformation. IEEE Trans. Pattern Analysis And Machine Intelligence 30, 617–631 (2008)
[21] Rankov, V., Locke, R.J., Edens, R.J., Barber, P.R., Vojnovic, B.: An algorithm for image stitching and blending. In: Proceedings of SPIE, vol. 5701, pp. 190–199 (2005)
[22] Singaraju, D., Vidal, R.: Interactive Image Matting for Multiple Layers. In: CVPR, pp. 1–7 (2008)
[23] Jia, J., Sun, J., Tang, C.-K., Shum, H.-Y.: Drag-and-Drop Pasting. ACM Trans. Graph. 631–637 (2006)
[24] Chuang, Y.-Y., Curless, B., Salesin, D.H., Szeliski, R.: A Bayesian Approach to Digital Matting. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 243–248 (2001)
[25] Levin, A., Zomet, A., Peleg, S., Weiss, Y.: Seamless Image Stitching in the Gradient Domain. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 377–389. Springer, Heidelberg (2004)
[26] Jia, J., Tang, C.-K.: Image: Stitching Using Structure Deformation. IEEE Trans. Pattern Analysis and Machine Intelligence 30, 617–631 (2008)
[27] [27] Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Transactions on Image Processing 15(2), 430–444 (2006)
[28] Wang, Z., Bovik, A.C., Sheikh, H.R.: Image Quality Assessment: From Error Visibility to structural Similarity 13, 1–13 (2004)
[29] Echtner, C.M., Ritchie, J.B.: The Meaning and Measurement of Destination Image. The Journal of Tourism Studies 14, 37–48 (2003)

Lessons Learned from Creating a General Purpose Tool for Experience Sampling Methods

André Coelho and Rui José

Centro Algoritmi, University of Minho, Guimarães, Portugal
{acoelho,rui}@dsi.uminho.pt

Abstract. Experience sampling methods (ESM) are a commonly used technique for capturing information from real or potential technology usage. The ubiquity of mobile phones has created a particularly appealing opportunity for all sorts of ESM studies, but there are complex technical challenges involved. In this work, we have sought to understand those challenges and the overall viability of a general purpose ESM application. The contribution of this work is the identification of the key challenges and approaches that may be more relevant in creating this type of general purpose study support. We have found that the main challenges are associated with data synchronisation, proper notification management, flexible questionnaire design, generalisation of study workflow processes, and efficient positioning. This contribution may inform the design of other generic tools for ESM-based studies.

Keywords: Experience sampling, User studies, User study tools.

1 Introduction

Experience sampling methods (ESM) [1], [2] are commonly used as a technique for capturing information from real or potential technology usage within the normal context of everyday life. Many opportunities and limitations associated with a new technology may only surface in the micro-contexts of everyday life and could easily go unnoticed in studies that do not consider this in situ or ecological validity. ESM consists in asking users, periodically or in specific circumstances, to register information that is relevant to an on-going study, e.g. stating the user's activity or feelings at that particular moment or taking a picture of a specified item in the local environment. This ability to collect information about what is happening at the moment when it is happening is a key distinguishing property of this method. For example, interviews are only able to gather data about events that have already passed, when people are no longer immersed in the situation itself and have probably forgotten much of the relevant information. Since ESM studies can be carried out without the continuous presence of the researcher they allow participants to maintain a normal daily practice without the pressure of being in a study, and they are also easier to scale to more participants or longer periods of time than what would have been possible, for instance, by using observation techniques. The ubiquity of mobile phones in everyday life [3], [4] has created a particularly appealing opportunity for all

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 839–848.
DOI: 10.1007/978-3-642-36981-0_78

sorts of ESM studies. Mobile devices have increasingly extensive data capture and communication features and they are inherently personal devices that we carry around with us every day, making them the obvious tool for ESM studies.

Despite this obvious potential, there are complex technical challenges involved in designing ESM mobile apps that are effective in their ability to collect relevant data while minimising the impact for the study participants. Moreover, this and the lack of common concepts for dealing with the specificities that user studies always imply, have both led to a situation where researchers wishing to undertake this type of studies need to create their own software to run their studies. Development of these custom-built tools significantly increases the costs and effort needed to run ESM studies and necessarily leads to simplistic designs with inevitable consequences on the data collection process. A flexible general purpose tool for ESM-based studies could considerably increase the range and number of such studies by allowing researchers to conduct them much more effectively. With such a tool, researchers could focus on their study goals rather than on their application for gathering data, considerably increasing the scope and potential reach of their research.

The work described in this paper has sought to create one such general tool for ESM-based studies with the ultimate goal of understanding the challenges involved and the overall viability of the idea. Based on an extensive study of the literature on ESM itself or describing ESM-based studies, we established a set of concepts that were common across ESM studies in general and we consolidated them in a system design that could support all those studies. The contribution of this work is the identification of the key challenges and approaches that may be more relevant in creating this type of general purpose study support. We have found that the main challenges are associated with data synchronisation, proper notification management, flexible questionnaire design, generalisation of study workflow processes, and efficient positioning. This contribution may inform the design of other generic tools that may improve support to ESM-based studies.

2 Related Work

ESM seems to be particularly well suited for studies in ubicomp scenarios, considering its impromptu and in situ nature, which allows it to capture not only the relevant information regarding occurrences but also, their context at the time. Vastenburg, M.H. and Herrera, N.R. [5] state that: “researchers cannot know all questions nor the detailed product use and context of use in advance”, therefore, they propose that ESM needs to be adaptive. This adaptability aims at enhancing the ESM technique (Adaptive ESM or A-ESM) so that it can provide data to researchers faster, with better quality, and also allow them to change the study according to their current needs and preliminary findings.

ESM, like many others, is an example of a technique from another field (psychology) that was brought into the field of ubiquitous computing, and that usually results in the technique not being fully adapted to the purpose it was brought in to serve. As a consequence, even though ESM preserves the ecological validity of its measurements, it produces self-reported data which is susceptible to cognitive bias. This led Cherubini, M. and Oliver, N. [6] to propose a refined ESM (rESM). This

technique consists of two things: 1) automating mobile device data collection through use of sensory capabilities and 2) triggering data collection on objective user-generated events that are non-self-reported. This is an important consideration to inform the design of an ESM application as it tells us that any and all data that is relevant to a study, and whose collection can be performed automatically, should be collected that way. If the user's intervention is not required and could possibly taint data due to cognitive bias then the user should be part of the process.

Another case of a technique designed to adapt ESM to ubiquitous computing settings is the work by Iachello, G. et al. on paratype [7]. This is an inquiry technique based on experience prototyping and event-contingent experience sampling, designed to survey users in real-life situations, about ubiquitous computing technologies. It is a simulation of a technology or interaction model that is evaluated alongside a real-world experience. The authors claim that paratypes are particularly useful when it comes to gathering early feedback on mobile and ubiquitous technologies, especially applications that collect information at unexpected times, that provide information when needed, or where interaction is embedded in unplanned social action or everyday routine. However, use of this technique requires overcoming certain obstacles related to the disruption in the flow of social interaction, sampling accuracy, and procedural requirements.

ESP/iESP [8] is a tool composed of a Windows or Linux, browser-based, desktop application and a mobile Palm Pilot client. The desktop application is used as a means for creating and managing the logic of the study, its content, and the timing at which the questionnaires are triggered. This tool's most prominent deterring factor is its mobile client which is tied to Palm Pilot, an out-dated device, with no connectivity and location or acceleration sensors.

CAES [9] (Context-Aware Experience Sampling) was a tool developed in an MIT project that was designed to enable questionnaires to be triggered by contextual clues, derived from the computation of data captured by the device. The tool ran in Microsoft PocketPC PDAs and was developed in a modular fashion, so as to allow it to capture data from context-sensing sensors. This would allow researchers to use context-sensitive sampling to trigger specific questions in relation to specific user actions. The proponents of this tool aimed to improve ESM by introducing contextual-awareness and thus, minimize the interruption annoyance of the ESM technique by sampling only during, or just after, an activity of interest. This project joined efforts with the team from MyExperience.

The MyExperience tool [10] is an open-source one that runs on devices with Windows Mobile 5.0. The MyExperience client can collect sensor data, allow for time-based (random or scheduled) or sensor data-based questionnaire triggering and also, provide phone usage data through logs. Researchers making use of this tool require some degree of technological knowledge as the creation of studies is done by editing XML files. However, that is not the only limitation as MyExperience's default approach translates into a black box for the researcher, this is because all the data is stored on the devices themselves. This means that all the data must be aggregated at a later stage and also, that there is no form of telling whether or not the users are engaged in the experiment, while it is on-going. This can be overcome by configuring a server to post data to, the inconvenient aspect being that it is not an out-of-the-box experience.

Fischer, J.E. [11] presents a review of ESM tools, ESP/iESP, CAES and MyExperience. The article exposes several shortcomings of each tool and makes recommendations on what would be a better approach to a general purpose ESM tool.

In MobXamp [3] and Track your happiness [4] we find mobile experience sampling tools that take advantage of a popular mobile OS (iPhone), the former is designed to evaluate contextual economic decision making and the latter is designed to track a user's happiness and, what factors in to it. As such, both of these tools were custom-built for a specific case study whereas we intend to build a more general purpose tool.

3 The TESS Platform for ESM-Based Studies

As part of the research methodology for this study, we have developed TESS (Tool for Experience Sampling Studies), our own instantiation of a generic tool for ESM-based studies. The initial system requirements have been elicited from the analysis of related work, more specifically the references included in the related work section. Each of those papers was analysed in search for references to desirable system features or references made by researchers about the necessary adaptation of the process to make it suitable for particular types of study. This initial set of requirements was then extended through the development of the platform and its use to support real user studies. This iterative design of the platform has been a key element of this research, allowing us to assess the relevance and impact of the potential requirements, and also to identify additional challenges resulting from technological limitations

TESS is a platform that can support the realisation of generic studies based on experience sampling methods. The system (Fig. 1) is based on a web server where multiple studies can be created and deployed, and on a mobile application that is used by participants to support the actual data collection.

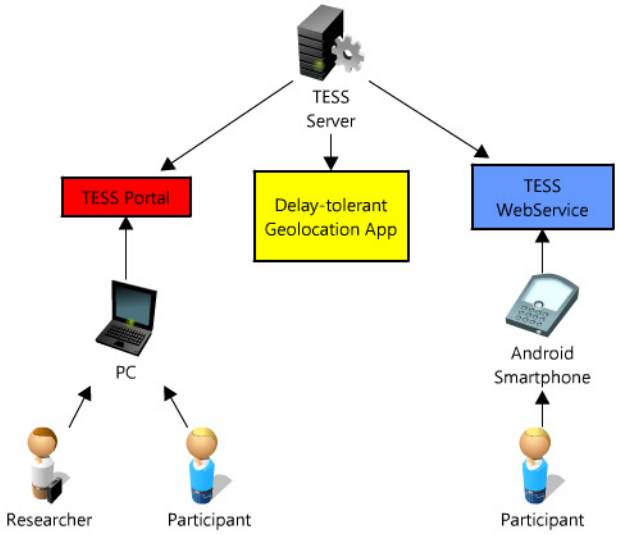


Fig. 1. Components running on the server and how platform users interact with them

3.1 Web Server

The web server (Fig. 1) acts as a portal for researchers and participants alike. The main features supported by the server are study management, data storage and user enrolment. Participants use the web site for managing their account, but they will primarily use the mobile application to answer questionnaires. Researchers use the web front-end to manage the execution of their studies. They will privilege flexibility, ease of managing studies and accessing study data for analysis. The core functionality from the perspective of researchers is to support the creation and management of user studies. From the initial set of requirements, we have clearly identified the following key properties:

• Simple tools for creating and managing studies;
• Flexible configuration of the studies, including dynamic changes to parameters;
• Integrated participant enrolment;
• Possibility to export data.

The server works in tandem with the mobile application, also managing the process of sending study instructions to mobile devices and handling the respective data collection as part of a specific study. The web server authenticates clients, before they download study specification files or before they upload study data. A system service running in the background periodically (hourly) scans study directories in search for new study data files that remain to be processed. A Perl script is then run to parse and treat all the data, producing new files with questionnaire data and location coordinates.

3.2 Mobile Application

The mobile application is shaped by the need to seamlessly integrate into the daily life of study participants. This means that participants will value the following characteristics of a tool:

- Minimize disruption when triggering user questionnaires;
- Support automated data collection;
- Avoid significant impact on the battery of the device;
- Support off-line operation to avoid the need for permanent connectivity;
- Configurable time period for alerts.

Our mobile application was developed as a native Android app that we made available through the Google Play Store. Once started, the application checks Android's account manager framework for credentials. If these are not present, the user will be prompted for the credentials he was assigned upon enrolling in a study. Participants can set their preferences regarding the type of data connection used to synchronise data. The application obtains from the server, the XML specification of the study in which the owner of that device is enrolled. This includes the questions to be asked, the number of notifications to issue throughout the study, if these are

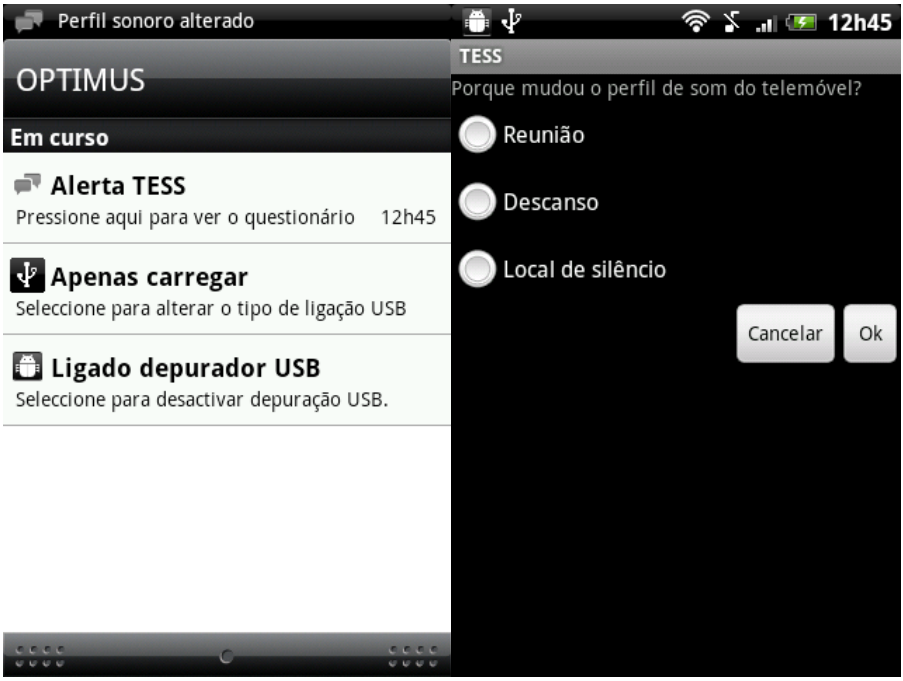


Fig. 2. Notification in Android's notification pane (*left*) and the experience sampling questionnaire resulting from pressing that notification (*right*)

random or periodic and what triggers them. Users are asked to fill-out questionnaires through system notifications in the notification centre (Fig. 2 *left*). Tapping the aforementioned notification will cause the application to access the relevant questionnaire's information (Fig. 2 *right*) which is then used to create a dynamic layout that will adapt the view to the type of question that it is presenting to the user, in this case multiple-answer type questions with variable number of choice-answers. All questionnaires are annotated with network landscape data by the application, so as to allow for posterior location data estimation (delay-tolerant positioning), when the data arrives at the server.

4 Evaluation

We took an iterative approach to the development of the platform, in which successive version were tested with increasingly complex studies. In this paper, we report on the last evaluation that we conducted to assess the value of the platform as a general purpose tool for experience sampling studies. We used the platform to run a study on ringer mode behaviours, i.e. what type of practices do people adopt in regard to changing the ringer mode of their mobile devices and how that is affected by location. The study involved reacting to any changes in mobile phone mode with a

short questionnaire on the reasons for that change. The tool would also gather data about the ringer mode selected by the user, the time of the occurrence, and the respective location. To minimise disruptions and avoid an excessive association between ringer mode changes and this study, the questionnaires were only triggered at each 5th ringer mode change or if more than 48h had occurred since the last change. Participants were members of the local academic community at our University and were rewarded with a €10 voucher. We will now present and discuss the key lessons from our evaluation work.

4.1 Notification Management

Notification is a central feature in the application. Multiple events can trigger notifications and they can be distributed in very diverse ways across time. This poses technical challenges in managing alarms, which trigger notifications and therefore questionnaires. Mobile devices can be rebooted and therefore any alarms that where armed are destroyed, which means that after the device completes a reboot, we need to rearm them. As such the application was coded to capture the Boot Completed Intent and when it does so, it will access the database via the Study Data Content Provider and rearm any alarms that have not already been triggered. Aside from technical challenges, notification management also poses issues when one considers the requirement of minimizing user interruption. As Fogarty et al. [12] shows us, especially in the work environment people consider themselves to be in a highly non-interruptible state most of the time, which translates into the necessity of devising mechanisms to evaluate a person's interruptibility. Additionally, their work shows us several models based on simple sensors that can be used to gauge someone's interruptibility, some of those sensors can nowadays be found or developed using mobile phones. Our approach in this respect was to not override the ringer mode, meaning that audible or vibratory notifications will be issued according to system settings and besides that, we do not force users to address notifications. If left unattended a notification will disappear when 5 minutes have elapsed since its time of origin, this is also important in keeping faithful to the principle of ESM. Having a user answer a question in relation to an event a long period of time after it has happened would subvert the principle of ESM.

4.2 Data Synchronisation

One other important aspect is that of ensuring that data synchronisation is performed in optimal fashion, so as to ensure consistency of data and to guarantee optimal use of energetic and monetary resources, which is one of the requirements we identified. The synchronisation of data is carried out by the system's synchronisation framework when an active connection is detected and faults in synchronisation are handled appropriately by the system (exponential back-off and retry). This means that synchronisation is handled gracefully without wasting valuable resources and significantly impacting user experience in a negative way that would hinder user adoption, therefore turning researchers away from the platform.

4.3 Flexible User Design

The specificities of our work and of the mobile OS we chose, posed a challenge for us in presenting questions dynamically. The content of Android layouts is usually static and inflated from XML files that are bundled with the application. We needed the tool to be able to inflate layouts with questions that where extracted from the study specification file and stored locally in a database. To achieve this we developed an Activity whose layout is programmatically inflated according to the question type and in the case of multiple choice answers, adjusts to the number of selectable choices.

4.4 Generalisation of Study Workflow Processes

An important realization of our work is that there are processes in the workflow of the platform that are hard to generalize and one such process is that of obtaining demographic data for studies. In our implementation we coupled the process of collecting this data with the process of subscribing to a study. Therefore, this meant that the demographic data we collect is of very general nature (gender, occupation, age, degree) and may not be sufficient for the needs of researchers. This is something that needs to be address in the sense of adding flexibility, in order to cover the needs of researchers. Perhaps this is something that could be done on the mobile client, after downloading the specification file for the study, this file could contain information on what demographic data researchers want.

4.5 Delay-Tolerant Location

Location data is essential for annotating events. However, if positioning implies significant energy, communication or privacy costs for users, it will become a severe obstacle to large scale use and volunteer recruitment. Therefore, in our location logging processes we have included two important design principles: the first is avoiding the use of GPS. We would always need an alternative solution because we need to consider indoor locations, but the key issue is that the continuous use of the GPS would necessarily have a very high cost in terms of power consumption; the second implication is to avoid depending on connectivity. In part this is also important to save energy, but since many people will not have a flat-rate data plan, there is also the issue that they will not accept the potential costs associated with data communications. The independence from connectivity would also allow us to perform the positioning without having to wait for the availability of a network connection.

To comply with these principles, we introduce the concept of delay-tolerant positioning. In most location-based services, location is normally part of an interactive feature and thus needs to be immediately available. On the contrary, in location-based logging, location information is needed to annotate an event. Therefore it is possible to just store the information needed to determine location, and leave the actual location calculation to some later point in time. Our data collection application, stores on the device the radio and Wi-Fi data generated that is used by the location API to determine location and when a connection becomes available, a batch of GSM

and Wi-Fi information is sent to a server that will then use that information to calculate the positions. This approach does not make any use of the GPS and works very well with only occasional connectivity. For location-based logging applications, this means that frequent positioning records can be generated without forcing the device owner to incur in significant power or network costs. We have studied the accuracy of this delay-tolerant technique in comparison with the real-time positioning and found them to be very similar [13].

5 Conclusions

We have assessed the viability of creating generic platforms for supporting experience sampling studies. While not reaching the point where we have a running system ready to be used by others, we have identified a number of important guidelines that may inform the design of such platforms as well as some technical challenges and trade-offs. In particular, the use of delay-tolerant positioning has revealed to be very important in addressing the specific positioning requirements of this type of studies. In the mobile application, there seems to be a major trade-off between the flexibility offered by a web-based approach in regards to interface design, and the additional control offered by a native application in regards to supporting sophisticated notifications and the specific delay-tolerant processes that we devised. Overall, the balance seems to sway in favour of the native application approach, especially if the expectations in relation to the types of questionnaires that can be made are not excessive.

References

1. Wac, K., Ickin, S., Hong, J., Janowski, L., Fiedler, M., Dey, A.K.: Studying the experience of mobile applications used in different contexts of daily life. In: Proceedings of the First ACM SIGCOMM Workshop on Measurements up the Stack - W-MUST 2011, p. 7 (2011)
2. Fischer, J.E.: Studying and tackling temporal challenges in mobile HCI. In: Proceedings of the 28th of the International Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA 2010, p. 2927 (2010)
3. Welpe, I.M., Reiss, M., Seiffert, P., Flake, C., Lauterbach, R., Tumasjan, A., Strobel, M., Hassold, S., Freundorfer, K.: Mobile Experience Sampling (MobXamp). Technische Universität München (2010), http://www.strategie.wi.tum.de/index.php?id=79&L=1
4. Killingworth, M., Pitiyanuvath, V.: Track your happiness. Harvard University (2009), http://www.trackyourhappiness.org/
5. Vastenburg, M.H., Herrera, N.R.: Adaptive Experience Sampling: Addressing the Dynamic Nature of In-Situ User Studies. In: Augusto, J.C., Corchado, J.M., Novais, P., Analide, C. (eds.) ISAmI 2010. AISC, vol. 72, pp. 197–200. Springer, Heidelberg (2010)
6. Cherubini, M., Oliver, N.: A Refined Experience Sampling Method to Capture Mobile User Experience. In: CHI 2009 Workshop: Mobile User Experience Research: Challenges, Methods & Tools, p. 12 (2009)

7. Iachello, G., Truong, K.N., Abowd, G.D., Hayes, G.R., Stevens, M.: Prototyping and sampling experience to evaluate ubiquitous computing privacy in the real world. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI 2006, pp. 1009–1018 (2006)
8. Consolvo, S., Walker, M.: Using the experience sampling method to evaluate ubicomp applications. IEEE Pervasive Computing 2(2), 24–31 (2003)
9. Intille, S.S., Rondoni, J., Kukla, C., Ancona, I., Bao, L.: A context-aware experience sampling tool. In: CHI 2003 Extended Abstracts on Human Factors in Computing Systems - CHI 2003, pp. 972–973 (2003)
10. Froehlich, J., Chen, M.Y., Consolvo, S., Harrison, B., Landay, J.A.: MyExperience. In: Proceedings of the 5th International Conference on Mobile Systems, Applications and Services - MobiSys 2007, p. 57 (2007)
11. Fischer, J.E.: Experience-Sampling Tools: a Critical Review. Journal of Youth and Adolescence, 1–3 (2009)
12. Fogarty, J., Hudson, S.E., Atkeson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C., Yang, J.: Predicting human interruptibility with sensors. ACM Transactions on Computer-Human Interaction 12(1), 119–146 (2005)
13. Coelho, A., Meneses, F., José, R.: Delay-tolerant positioning for location-based logging with mobile devices. In: 6th International Conference on Ubiquitous Computing and Ambient Inteligence (2012)

ManPro: Framework for the Generation and Assessment of Documentation for Nuclear Facilities

Cristina Olaverri-Monreal*, Carsten Dlugosch, and Klaus Bengler

Institute of Ergonomics,
Department of Mechanical Engineering
Technische Universität München, Germany
Boltzmannstrasse 15, 85747 Garching, Germany
{olaverri,dlugosch,bengler}@lfe.mw.tum.de
http://www.ergonomie.tum.de

Abstract. Nuclear plant operators must act in accordance with a number of requirements which are specified in the technical documentation that describes handling, functionality and architecture of the systems. Due to these many specifications to be considered, it is crucial to create technical documentation in order to improve the operation efficiency of such facilities and prevent human errors whenever possible. In this work, we propose a procedure (ManPro) for the computer-based creation of instruction manuals for the operation of technical systems. The "ManPro" approach is semiautomatic, which underlines its reproducibility and efficiency. Feasibility of the approach and its effects on usability are assessed and outlined.

Keywords: Nuclear facilities, Technical documentation, Mark up languages, XSLT, XML, UML.

1 Introduction

Recent unfortunate events related to safety systems within nuclear facilities have boosted worldwide concern about the downsides of using nuclear energy. Some control systems installed within nuclear facilities have been developed and established many years ago, and therefore no longer comply with the latest technological standards. New components need to replace existing ones in order to maintain the desired safety standards. This also applies to the production of instruction manuals that have been, in some cases, developed over 40 years ago, and have since rarely been updated. Nuclear plant operators must act in accordance with a number of requirements which are specified in the technical documentation.

This documentation describes the specifications that a plant has to meet, as well as its handling instructions, functionality and architecture [1]. The series of standards related to technical product documentation and nuclear power plants have been published by the International Organization for Standardization (ISO) [2, 3].

* Corresponding author.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 849–859.
DOI: 10.1007/978-3-642-36981-0_79

These standards describe how to establish which information users need, how to determine the way in which that information should be presented to users, and how to prepare the information and make it available.

The constantly increasing efficiency of computer-based control systems offers a wide range of possibilities for detecting and controlling undesirable operating conditions; it is therefore only natural to take advantage of such new technologies for the benefit of nuclear facility safety. These computer-based control systems require interaction with the facility operator, through their Human Computer Interfaces (HCI). Moreover, the information that these systems provide needs to be depicted to users in the most efficient way to understand electronic messages [4] and facilitate interaction, particularly in situations that poses an immediate risk. Several works have mentioned the fact that the nuclear industry has taken a long time to incorporate computerized systems, even if they have been found to have substantial advantages over the technology that they replace [5]. Instruction manuals have a direct impact on the efficiency of processes within these facilities and the productivity and safety of their employees. Thus, it is important to acknowledge the dynamic nature of these plants handling and functionality instructions for a more flexible representation of documents. We propose a reproducible and efficient procedure for the computer-based creation of instruction manuals to operate nuclear plants systems in this work. To ensure that the manual is searchable, a User Interface (UI) is generated from a modeling specification in the Unified Modeling Language (UML), which defines the components of a system. The information entered by the user is then stored for the further presentation of the document in different formats.

The remainder of this paper is organized as follows. The next section revises related work in the areas of automation and documentation of technical systems. Section 3 presents a detailed description of the methodology used to implement the proposed approach, and section 4 reports on the evaluation of the system. Finally, Section 5 concludes the paper.

2 Related Work

2.1 Automatic Generation of User Interfaces

To automatically generate user interfaces, the information content design has to enable the user to successfully perform the tasks keeping the information reduced to the minimum necessary as suggested in the formal approach and methodology for analysis and generation of human-machine interfaces, with special emphasis on human-automation interaction in [6]. A formal documentation of the software requirements guarantees a precise and correct user interface. Therefore a common practice in the early stages of software development is to build a Unified Modeling Language (UML) system model with UI prototypes that can capture the different aspects of a system [7, 8]. The authors in [9] presented in this context current model-based approaches to user interface automatic generation such as the following:

- The XSI approach, that is based on the segmentation of a model into different sub-models and provides an Entities view, a Use Case view and a User Interface view [10].
- The OO-Method approach that aims at producing a formal specification of a software system in the object-oriented language OASIS through a graphical system model at a conceptual level. This conceptual model is then translated to an OASIS specification [11, 12].
- The ZOOM approach to enable model-driven development through the structure, behavior, and user-interface of an application, providing thus three related notations to describe each of those parts [13].

Our approach uses a formal system architecture model in UML form to represent a correct user interface that map the different system components on the user interface so that the user can then enter the information that will later appear in the user manual.

2.2 Automatic Documentation Generation

Design, documentation and evaluation of safety-critical applications used in nuclear power plants have been addressed in several works. For example, the authors in [14] proposed a software formal documentation approach through both requirements and software design based on the systematic comparison of program behavior with the engineering specifications of the computer system. In the same context, an authoring tool for automating the software generation process has been presented in [15]. The authors investigated how to provide on-line assistance for software professionals based on the information users require. Further, an XML-based approach for generating the architectural documentation of a software system from the implementation code was presented in [16]. The authors identified relevant concepts for the software documentation and extracted them through an analysis of the source code. Additionally, they organized the documentation in a hierarchical form, which they presented in a human-readable format.

In the same XML content, a software solution for a variable information structure display that was based on the combination of markup languages was proposed in [17]. The authors used XML for representing a document structure, which could subsequently and independently be transformed into new ones using XSLT Stylesheets. In [18], the authors also introduced a related approach to an iterative reconstruction process that allows users to create a source code model in a database, and extract the reconstructed architecture through Structured Query Language (SQL) queries. Most approaches to generate software documentation use UML diagrams that include the system design information. For example, an approach to access and review UML software engineering diagrams on the Web-based on markup technologies was the topic in [19]. Additionally, an approach was developed in [20] which ensured a stronger link of user documentation's generation with software's life cycle phase. The approach enabled the transfer of required information from the software's functional specification described in the Unified Modeling Language

(UML) into user documentation. Also several techniques have been proposed more recently which aim to integrate Web-based data into a suitable format for system designers and end users, constructing UML diagrams from XML data [21, 22]. In other works, tools have been proposed to automatically create graphical input dialogs from text specification entered by software developers [23]. These tools differ from our approach in that they automatically extract information from an existing system. Our approach bases on user specifications combining some technologies from these frameworks to generate specific documentation for nuclear facilities semi-automatically through a computer-aided process.

3 Methodology

The main goal of the "ManPro" procedure is to guarantee that information contained in the final user manual is searchable and understandable for the user, and that ambiguity does not exist. The tool adheres to internationalization principles, allowing for multilingual content. In "ManPro", a relational database contains all the tables and relationships needed for the reliable and correct storage of information previously entered by the user through a Web-based form, which has been dynamically created from a UML specification of the system architecture, behavior, structure and maintenance. The descriptive components of the instruction manuals are then stored in XML format for the further compilation of the final document. "ManPro" consists of the following components:

- An XMI file containing the structure for a modeling specification of a nuclear plant system in UML;
- A Web form that ensures accuracy in the content of the documentation through guided questions and predefined fields extracted from the UML system specification, preventing errors. Additionally, information is immediately available and can be accessed from any computer with Internet access. It also assists system developers with templates and descriptions of requirements for nuclear facilities during the documentation writing step;
- A relational database that enables simple data updates, validation and accuracy through error check. Additionally, the storage of data in a database allows the user to cross-check data against existing information and to map already available documents. Information can be easily accessed, visualized and manipulated through SQL. The information available in the database can then be extracted through Java and MySQL queries, creating information related to chapters, sections, graphics, paragraphs or subsections, also including different kind of lists. Document internationalization requirements are also specified in Java;
- An XML document, extracted from the database's information using JDBC and containing the database structure mentioned above that will be further transformed into the final instructions manual version in PDF;
- The final instructions manual in PDF achieved through the XSL Transformation language XSLT;

The tool implementation process is depicted in Fig. 1.

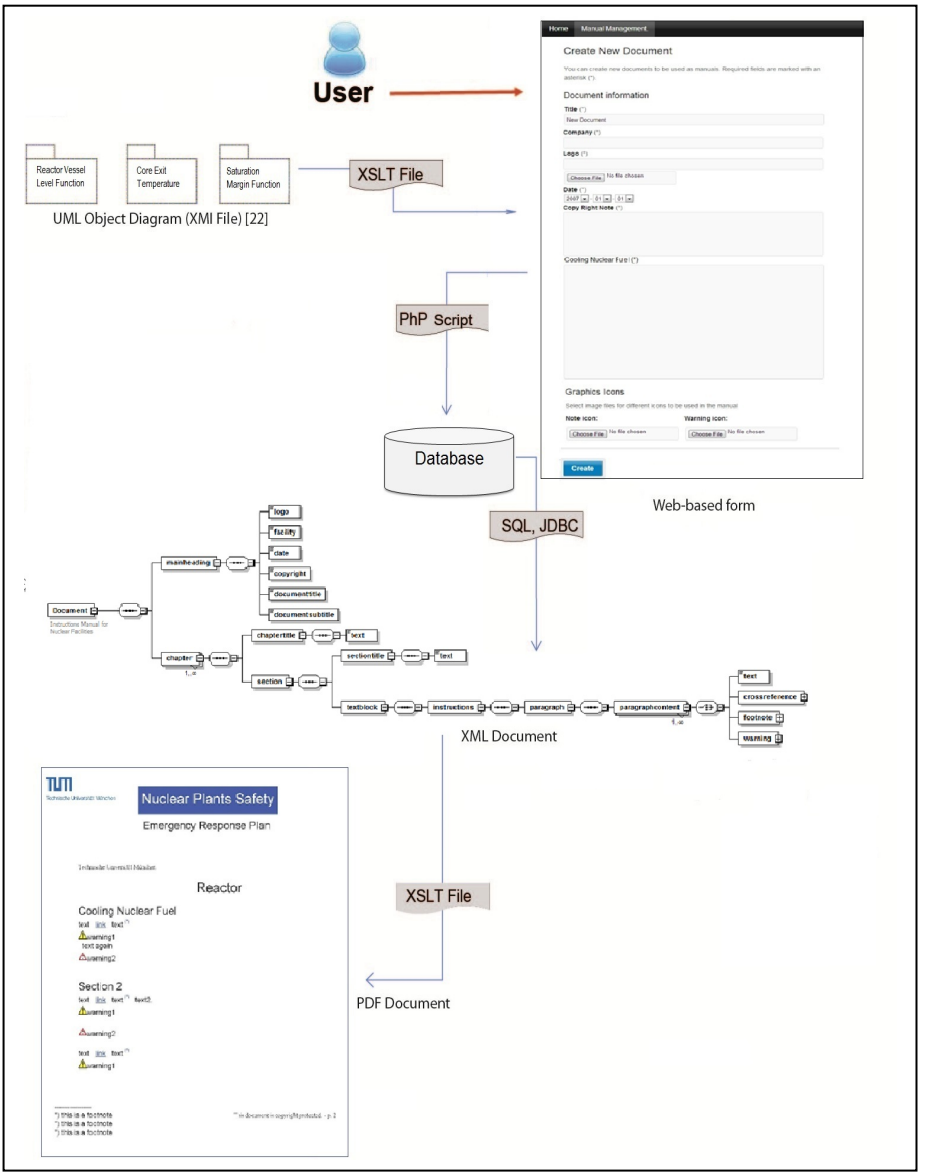


Fig. 1. "ManPro" information processing steps and output files

The figure illustrates the output files resulting from the database and XML structure after processing the information with the "ManPro" tool. The input UML depicts an example of three possible functions of a Core Cooling System [24]. To edit an instruction manual, the user needs only to accesses the appropriate Web forms and

can then create a document, add information or change contents for an existing document, and then save it to the database.

Our approach uses markup languages as the common feature in the development process, integrating XML-based technologies with the rest of the development framework, Java technology, databases, and the Web. We use a modular structure, with each module responsible for one part of the process. The framework begins with a UML object diagram specification of the components of a nuclear power system describing the system's behavior, structure and operating rules.

The input fields for the Web form, which enables text data entry and submission to the relational database is generated from the objects specified in the UML diagram. In the next step, an XML document is extracted for further transformation into a human-readable format.

3.1 Web Form Generation

The XML Metadata Interchange (XMI) format for UML model interchange was accessed through a XSLT file to generate the input fields of the Web form.

Since XMI is XML-based, the file contained the patterns to match the elements within the XMI document and the transformations to apply when a match was found.

To build the Web form for data entry, we used the HyperText Markup Language (HTML), to display information in a Web browser.

Additionally, the open source, server-side PHP scripting language was used to insert the form data into the database and to validate the user input on the server.

3.2 Database Platform

To develop the relational database platform, we used the XAMPP PHP-Apache-MySQL solution. The "document" table is connected to the rest of the tables containing document parts. This architecture allows for reuse of such data for multiple documents. Many-to-many relationships are saved in separate tables, which refer to primary keys with foreign keys of further tables to easily join information for a later data extraction based on queries.

Fig. 2 illustrates the architecture of a sample database section. Table 1 describes the tables' content.

3.3 Database XML Extraction

To be able to access database information in an XML format, we have built our approach upon the JDBC data access technology.

Through a mapping process, we first selected the information that we wanted to retrieve leaning on an intermediate xml mapping file used by java code to acknowledge the database structure [25]:

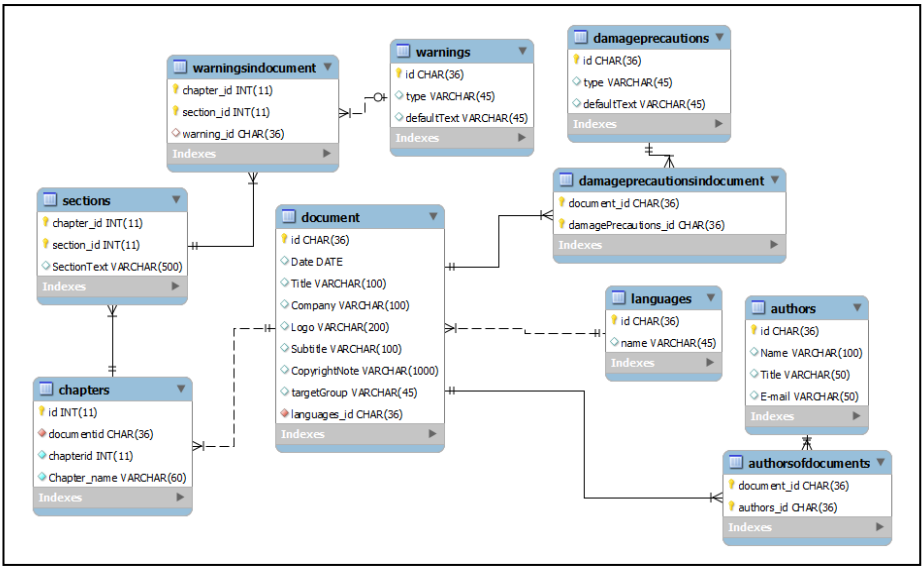


Fig. 2. Database Section Architecture

Table 1. Description of the database architecture

Table Name	Functionality
document	Main table containing the documents
authors	Table of all authors
authorsofdocument	Many-to-many relationship table between the “document” and “authors” tables
warnings	List of possible warnings
warningsindocument	Many-to-many relationship table between the “document” and “warnings” tables
damageprecautions	List of damage precautions or necessary precautions to prevent damage
damageprecautionsindocument	Many-to-many relationship table between the “document” and “damageprecautions” tables
languages	List of languages
chapters	List of chapters
sections	List of sections

1. An intermediate mapping XML file retrieved the necessary information from different database tables through Standard Query Language (SQL) queries.
2. In the same file, we specified the overall structure of the new XML document, describing the tree elements and their attributes, namely the root and the rest of elements that represented the database rows.
3. Additionally, we completed the structure information with the names and contents of data elements, creating new elements and attributes.

An XSL Stylesheet contained the rules for depicting information included in the XML with the appropriate layout. We exploited the resources of XSLT 2.0 for regular expression matching, to describe the text strings patterns we wanted to manipulate. To create the new XML document, the extracted data was made available by parsing the file created above. We then recovered the data element and the SQL statement, determining first the root element and then obtaining data nodes. The data was then stored in the root of the document tree, in order to gain primary access to the document's data, and to be able to create element nodes, text nodes, comments, processing instructions, etc. through the methods contained in the Document Object [26]. A section of the mapping file code is illustrated below.

Section of the intermediate mapping XML file to retrieve selected information from a database

```
<?xml version="1.0"?>
<mapping name = "map">
   <data sql="select * from document" />
   <root name="documentInfo" rowName="document">
      <element name="Title">
         <content>Title</content>
      </element>
      <element name="Date">
          <content>Date</content>
      </element>
      <element name="Company">
          <content>Company</content>
      </element>
      <element name="Logo">
          <content>Logo</content>
      </element>
      <element name="subtitle">
          <content>Subtitle</content>
      </element>
   </root>
</mapping>
```

3.4 Instructions Manual Creation in PDF Format

Using XSLT, we performed the final XML document transformation into a PDF. For this step, Apache FOP Java-based open source application was used to generate PDF

documents from FO files using XSL. We accessed the tree nodes of the basic XML document and selectively copied the content of the XML document into a new XML document, which represented the final structure. Pattern matching was used to identify the variable options for the copying process.

4 Final Document Evaluation

The manuals that were semi-automatically generated by "ManPro" were intended to serve as a reference for the evaluation of other existing manuals, as they enabled the verification of Quality Assurance (QA) requirements. Thus, we tested the approach to produce technical documentation through an evaluation process that was performed in terms of a level of detail analysis.

The printable PDF file produced in the process was tested by persons familiar with the technology described in the manual in early stage tests to improve the tool through a formative evaluation early in the design process. Relying on the guidelines to evaluate information contained in instructions manuals [27], we made sure that the document followed the under mentioned specifications:

- The instructions described all the product characteristics in a step-by-step procedure;
- The manual included a quick start guide;
- It also includes a list of the functions;
- The manual included line numbers to help with cross references;
- Instructions were presented in the form of step-by-step procedures;
- The information was written in a consistent way;
- Sections were ordered by frequency of use;

After this, we asked five additional persons to analyze the final document and look for potential inconsistencies. We re-designed the first prototype based on these evaluations results. The Graphical User Interface created to enter the manual information through the web form was additionally evaluated to determine the usability of the software. The five users considered the platform to be friendly and easy to use.

5 Conclusion and Future Work

In this paper, we have presented "ManPro", a novel framework for the computer based creation of instruction manuals for the operation of technical systems in nuclear power facilities.

Our approach is language and platform independent, ensures accuracy of documentation content through predefined fields for data entry, preventing errors, and the information is accessible to multiple users. Additionally, the tool allows for multilingual content. Our approach has been proven to be efficient and user-friendly, as it guarantees that the information contained in the final user manual is searchable

and understandable for the user. Therefore, the results presented in this paper establish the base for our further research in the field. We will refine UML diagrams in future work adding more extensive information to the instruction manuals and will assess the content design, organization and operating instructions through a comprehensive database-system for the evaluation of human factors and ergonomics.

Acknowledgments. This work was supported by the VeNuS 2 Project - Approach to the Efficient Assessment of Safety and Usability of Computer-Based Control Systems - Number 1501282/2011, funded by national funds through the Federal Ministry of Economics and Technology, Germany, BMWi. It contributes to the research on reactor safety promoted by the Federal Ministry of Economics and Technology, Germany, BMWi.

References

1. Pyzdek, T., Keller, P.: Quality engineering handbook, vol. 60. CRC (2003)
2. ISO: 01.110: Technical product documentation, international organization standardization (July 2012), http://www.iso.org/iso/catalogue_ics_browse?ICS1=01&ICS2=110&
3. ISO/IEC: Isoiec 26514-2008 systems and software engineering. Requirements for designers and developers of user documentation (July 2012), http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=43073
4. Carver, D.: Designing the user interface, strategies for effective human-computer interaction. Journal of the American Society for Information Science 39(1), 22–22 (1988)
5. Parnas, D., Asmis, G., Madey, J.: Assessment of safety-critical software in nuclear power plants. Nuclear Safety 32(2), 189–198 (1991)
6. Heymann, M., Degani, A.: Formal analysis and automatic generation of user interfaces: Approach, methodology, and an algorithm. Human Factors: The Journal of the Human Factors and Ergonomics Society 49(2), 311–330 (2007)
7. Jacobson, I., Booch, G., Rumbaugh, J.: The Unified Software Development Process. Addison-Wesley (1999)
8. Pressman, R.S.: Software Engineering – A practitioner's approach, 6th edn. Mc Graw Hill (2005)
9. da Cruz, A.M.R., Faria, J.P.: Automatic Generation of User Interface Models and Prototypes from Domain and Use Case Models
10. Silva, A.R., Saraiva, J., Silva, R., Martins, C.: XIS - UML profile for extreme modeling interactive systems. In: Proceedings of the 4th International Workshop on Model-based Methodologies for Pervasive and Embedded Software (MOMPES 2007). IEEE (March 2007)
11. Pastor, O., Molina, J.: Model-driven Architecture in Practice – A software production environment based on Conceptual Modeling. Springer (2007)
12. Molina, P.J., Hernández, J.: Just-UI: Using patterns as concepts for IU specification and code generation. In: Perspectives on HCI Patterns: Concepts and Tools (CHI 2003 Workshop) (2003)

13. Jia, X., Steele, A., Liu, H., Qin, L., Jones, C.: Using ZOOM approach to support MDD. In: Proceedings of the 2005 International Conference on Software Engineering Research and Practice (SERP 2005), Las Vegas, USA (2005)
14. Courtois, P.J., Parnas, D.L.: Documentation for safety critical software. Book Documentation for safety critical software, pp. 315–323. IEEE, City (1993)
15. Johnson, W., et al.: Dynamic (re) generation of software documentation. In: Proceedings of the Fourth Systems Reengineering Technology Workshop (1994)
16. Riva, C., Yang, Y.: Generation of architectural documentation using xml. In: Proceedings of Ninth Working Conference on Reverse Engineering, pp. 161–169. IEEE (2002)
17. Olaverri-Monreal, C., Bengler, K., Breisinger, M., Draxler, C.: Markup languages and menu structure transformation during the internationalisation process of driver information systems. Localisation Focus, 4 (2010)
18. Kazman, R., O'Brien, L., Verhoef, C.: Architecture reconstruction guidelines (2001)
19. Nentwich, C., Emmerich, W., Finkelstein, A., Zisman, A.: Box: Browsing objects in xml. Software-Practice and Experience 30(15), 1661–1676 (2000)
20. Zöllner, R., Patakui, K., Rosenthal, K.P.: Generation of the user documentation by means of software's specification in uml. In: Proceedings of the International Symposium Facteurs Humains et Conception des Systèmes de Travail. Optimiser les Performances de lEnterprise. Nizza. Comit AISS Recherche (2006)
21. Yu, A., Steele, R.: An overview of research on reverse engineering xml schemas into uml diagrams. In: Third International Conference on Information Technology and Applications, ICITA 2005, vol. 2, pp. 772–777. IEEE (2005)
22. Jensen, M., Møller, T., Pedersen, T.: Converting xml dtds to uml diagrams for conceptual data integration. Data & Knowledge Engineering 44(3), 323–346 (2003)
23. Zanden, B., Myers, B.: Automatic, look-and-feel independent dialog creation for graphical user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People, pp. 27–34. ACM (1990)
24. Koo, S., Son, H., Seong, P.: A method of formal requirement analysis for npp i&c systems based on uml modeling with software cost reduction. Journal of Systems and Software 67(3), 213–224 (2003)
25. Chase, N.: Using jdbc to extract data into xml (November 2012), http://www.ibm.com/developerworks/xml/tutorials/x-extract/
26. XMLDOM: Xml dom the document object, 57–66 (November 2012), http://www.w3schools.com/dom/dom_document.asp
27. Inaba, K., Parsons, S.O., Smillie, R.J.: Guidelines for developing instructions. CRC (2004)

Text Mining Indicators of Affect and Interaction: A Case Study of Students' Postings in a Blended-Learning Course of English for Specific Purposes

Helvia P.P. Bastos[1], Magda Bercht[2], Leandro K. Wives[2], Júlia Kambara-Silva[2], and Yasmmin Martins[1]

[1] Instituto Federal Fluminense, RJ, Brazil
helviabastos@yahoo.com.br, nim_asay@hotmail.com
[2] Instituto de Informática, UFRGS, RS, Brazil
{bercht,wives,jkksilva}@inf.ufrgs.br

Abstract. This is a case study of text cues of affect and interaction in students' postings in a virtual learning environment designed to support and expand face-to-face classes of English for Specific Purposes at university level. These categories of text analysis are indicators of the learner's social presence in online activities. Methodology in this study included (i) mining text indicators of affect and interaction; (ii) applying a questionnaire to investigate students' perception of the environment; (iii) verifying how these results reflect the students' level of satisfaction with the blended-learning approach.

Keywords: Discussion Boards, Presence Plus Model, Text Mining, Blended-learning, English for Specific Purposes.

1 Introduction

This paper presents a case study on messages posted in a virtual learning environment (VLE) devised to enhance conventional classes of English for Specific Purposes. The study aimed at identifying text cues of interaction and affect and process them in a text mining software in order to verify the students' level of satisfaction and acceptance of using a hybrid approach to language learning – aspects related to the notion of social presence in distance learning environments.

Digital technologies (DTs) have provided a range of educational opportunities and resources both in distance learning and as support to traditional classes. One of the many possibilities for educational actions enabled by DTs is blended-learning, that is, face-to-face classes expanded into VLEs. In this scenario, teachers and learners no longer need to be contrived by limited traditional class time to explore, discuss, and expand course contents.

Discussion boards (also known as message boards, discussion forums, online forums) are important tools that allow for interaction in VLEs. According to Sahu (2008), they play a relevant role in e-learning, despite the continuous development of technological innovations. In these environments, course participants can, among

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 861–872.
DOI: 10.1007/978-3-642-36981-0_80

other possibilities, reflect on course materials, insert new topics, and socialize in a different manner. Since asynchronous communication may contain multilayered dialogues, as well as significant text cues of the learner's engagement in the group and/or course, students' posts make up an interesting corpus for quantitative and/or qualitative analysis of various features, including satisfaction with the pedagogical methodology, an aspect of the social presence of learners as individuals or as group.

Social presence (SP) is defined by Gunawardena and Zittle (1997) as the degree to which individuals are perceived as 'real' in mediated communication". According to Garrison et al. (2001), SP is the ability of people to project themselves both socially and emotionally in online encounters. This sensation of being present, despite the physical absence, is fostered by text-based Computer Mediated Communication (CMC) in tools such as e-mail, forums and chats, and usually enhanced by non-verbal elements (graphic resources).

Following a review of related studies in Section 2, the article gives an overview of blended-learning and English for Specific Purposes in Section 3. The methodology used in the study and analysis of the results are presented in Section 4 and 5.

2 Related Studies

Identification and assessment of students' interactions in online forums has been object of extensive research in using different approaches and methods. This section list some research and efforts to establish criteria (*rubrics*) to help assessments of postings in discussion boards. For Jackson (2010), rubrics have the three following purposes: (i) "as a tool to judge the value of individual students' contributions to discussion"; (ii) "as a framework to evaluate the overall effectiveness of discussions, such as levels of cognitive engagement achieved and balance of contributions across these levels"; (iii) as a means to make explicit and transparent to students the expectations of their engagement in discussions, and thereby shape that engagement.

The flow of discussions was examined in Chen and Chiu's study (2008). The authors analyzed the effect of previous messages on subsequent postings by following five "dimensions": *evaluations* (agreement, disagreement, or unresponsive actions); *knowledge content* (contribution, repetition, or null content); *social cues* (positive, negative, or none); *personal information* (number of visits); and *elicitation* (eliciting response or not). Data obtained in the study show that disagreements or contributions in previous postings produced more disagreements and social cues in following messages.

Rourke and collaborators (2001) propose discursive indicators based on criteria previously given by Garrison et al. (2001) to calculate the learner's density of SP (low, medium, and high). Rourke et al. (ibid) coded and studied transcripts of postings in two distance learning courses, and submitted them to quantitative analysis in the Atlasti software[1]. The authors suggest three main categories and twelve indicators to guide content analysis of forum postings: (i) *affect* (*expression of emotion, use of humor,* and *self-disclosure*); (ii) interaction (*continuing a thread*, *quoting from others'*

[1] http://www.atlasti.com/index.html

messages, *asking questions*, *complimenting*, *appreciation,* and *agreement*); (iii) cohesion (*inclusive pronouns*, *phatic expressions*, and *salutations*).

Expanding Rourke et al.'s three categories, Bastos (2012) presented a model for identifying SP marks in educational forums and chats. The Presence Plus Model (PPlus) contains a fourth category – Force, a rubric inspired by Martin and White's Appraisal System of Language Evaluation (2009). To carry out her study, Bastos (2012) crossed three different techniques to analyze 2,049 posts: (i) manual content analysis (mapping and categorizing SP indicators); (ii) mining and calculation of SP density; (iii) survey questionnaire answered by teachers and students in a distance learning course.

As made clear in the Introduction, this work is concerned with affect and interaction as manifestations of the students' engagement and satisfaction in the learning experience described in this article. By *affect* we mean *affective states* (Scherer, 2005), while *interaction* is used to refer to certain communicative strategies people use to achieve purposes such as *congratulating, thanking, apologizing, offering help* etc. (See item 4.1 below). The framework selected to guide this study is the PPlus set of indicators. Its automatic analysis was made by Kambara-Silva's software *Presente!* (2011) – a program developed to process postings based on the PPlus Model. Other theoretical and research resources used in the study are described in the coming sections.

3 Context of the Investigation

This section presents some notions on blended-learning, as well as the language learning approach used with participants of this case study.

3.1 Blended Learning

According to Granham et al. (2005), "blended-learning" (BL) combines face-to-face (F2F) instruction with computer-mediated (CM) instruction". Granham et al. (2005) identify four levels in which BL can occur: activity, course, program, and institution. In the course observed in this study, use of BL is mostly related to activity and course levels, since application of BL in program and institutional levels rely on other school technical and pedagogical staff.

Developing a blended-learning course presents some challenges for teachers and students as well. Graham (2006) draws attention to these: (i) teacher training; (ii) customized and culturally relevant materials; (iii) technological-skilled students; (iv) students' difficulties in accessing technology due to social-economic reasons.

For a comprehensive view of blended-learning as a methodology in the teaching and learning of English for Specific Purposes (ESP), two studies are worth mentioning: (i) Álvarez and González (2007) report on how to integrate BL with ESP in the Moodle environment; and (ii) Daragmeh's analysis of the pros and cons of this hybrid methodology (2010).

3.2 Moodle-A&UD

This case study was carried out from August through November 2012, with 17 (seventeen) undergraduate students of Architecture and Urban Development at the Instituto Federal Fluminense (IFF-Campos, RJ, Brazil). As the subject "English for Specific Purposes" (ESP) is not part of the course curriculum, it was offered as a semester elective discipline (40h). The course program focused on the review of relevant syntactic structures, and presentation of specific vocabulary via texts covering transversal topics in Architecture and Urban Development (A&UD).

In this particular situation, the ESP approach was selected as it caters for the learners' needs to understand texts of technical content. Accordingly to ESP principles (Hutchinson and Waters, 1987), materials and online activities were designed to develop competences and skills that students may need, to accomplish academic and/or professional tasks. Thus, the language program integrates linguistic skills with themes studied in other course programs. This means that the language is not studied *per se*, but as a means to read about relevant topics in the students' overall training.

The Moodle setting[2] developed for the A&UD group of students, named "Moodle-A&UD" for this study, was designed to meet the needs and interests of the students, allowing for more time to solve questions and do collaborative work. Therefore, Moodle-A&UD was not only a repository of materials studied in class, but also as a space for interaction and socialization. Figure 1 shows the top of the course homepage. Distribution of the basic elements of the page follows the conventional layout available in the system.

Fig. 1. Homepage of Moodle-A&UD

Table 1 presents the list of materials in Moodle-A&UD. Due to the fact that the subject was given both in conventional classes and expanded into a VLE, part of the tools and functionalities offered by the system were not used – chat and video conference, for instance. The objective of inserting materials covered in class was to give students more flexibility in reviewing and reflecting on the various units of the program. The presentations in slides offered opportunity for reviewing relevant grammar structures according to individual needs or interest.

[2] http://www.ptce2.iff.edu.br/moodle/course/view.php?id=78

Table 1. Materials and resources available in Moodle-A&UD

Purpose	Materials and Resources
Support and Expansion	- program and objectives - performance evaluation criteria - presentations seen in class - presentations for grammar review - links to supplementary technical texts - links to English learning sites - guidelines for group projects
Interaction and Collaboration	- e-mail - bulletin board (suggestions, ideas) - discussion board

Interaction was mostly done via the discussion board. Topics were proposed either by the teacher or by the students themselves. Examples of discussions proposed by the teacher are: *English as a necessary professional asset, English words or expressions in everyday Brazilian life.* This tool was also used for research and comments on these tasks*: idioms, evaluation of a blog or site for students of English as a foreign language, comparison of the students' course curricula with those in a foreign school of Architecture, selection and description of a sustainable architectural or urban development project.*

Students collaborated by inserting new topics related to their graduation course and professional practice, including news, articles, cartoons, as well as links to current innovative architectural or urban projects. Students were asked to post their collaborations in threaded format – a structure that allows for multiple comments, avoiding individual and unanswered postings.

4 Methodology

For this case study, three strategies were used: (i) selection of the corpus of analysis; (ii) text mining of students' postings according to the PPlus model, and (iii) an online survey questionnaire. According to Creswell (2003), the triangulation of different research strategies and sources may guarantee consistency of results.

The questionnaire was organized on GoogleDocs[3] and answered by 11 out of the 17 students enrolled in the course. Part 1 of the questionnaire focused on the learners' familiarity with the Moodle platform (new for 07 students – 64%). Part 2 of the survey presented items on Moodle-A&UD, and Part 3 concentrated on the student's interaction and satisfaction with the environment. Items were presented in the Likert[4] scale format, with 04 options of answers: *strongly agree, agree, disagree, strongly disagree*). Results of this survey are discussed in Subsection 5.1.

[3] `https://docs.google.com/spreadsheet/viewform?formkey=dEhaakdrMXdGQWVsQk9YRjZhcDdVWXc6MQ`

[4] Likert, R. "A technique for the measurement of attitudes". In *Archives of Psychology*. n. 140, 1932.

The study used content analysis to map and classify SP indicators (*units of* analysis). According to De Weaver et al. (2006), the term "unit of analysis" refers to the basic text unit to be classified; ranging from one or more words to graphic elements. For text mining, we used Kambara-Silva's software *Presente!*, developed to carry out lexicometric[5] analyses of SP text cues indicated in the PPlus table. To do this, words, phrases, and idiomatic expressions were inserted in the program as algorithms – one for each SP indicator (e.g. the algorithm "color" assures that instances in a color other than black will be identified). In the same fashion, other algorithms in the software are "regular expressions", "bold", "proper names", and so forth. The software is made up of three main tools, as follows: (i) a *converter* that converts HTML files retrieved from forums and chats XML files; (ii) a *category builder* which allows teachers to create and edit classes and subclasses of SP; (iii) an *analyzer*, which makes analysis of the files based on the category builder, and provides the SP degree of students. Thus the program processes terms and expressions based on the list of regular expressions identified in the manual analysis carried out by Bastos (2012) on two online courses.

4.1 Identification of Text Indicators of Interaction and Affect

Indicators found in the PPlus Model were used to identify, process and analyze text cues of interaction and affect. As stated in Section 2, this model derives from Rourke et al.'s (2001), but expands it with more subcategories and indicators.

Used as a synonym for *affective processes*, the term *Affect* (AF) encompasses these notions pointed by Scherer (2005): preferences, attitudes, mood, affective dispositions, and interpersonal stances. Our template has the subcategories described in Table 2 to express *Affect* (AF).

Table 2. Subclasses and indicators of Affect (AF) – *PPlus* (Bastos, 2012)

Subclasses	Indicators	Examples
AF1. Emotion	AF1a. Interjections	*Argh; Oh; Eek*
	AF1b. Onomatopoeia	*Kkkk*
	AF1c. Emoticons and gifs	*;) / =P /* 
	AF1d.Repetitious punctuation	*!!!! / ??? / !?!?*
	AF1e. Repeated letters	*Gooood*
	AF1f. Enhancement	*Capitalization*
		Color
		Bold
		Italics
		Quot. Marks
AF2. Humor	AF2a. Sense of humor	*What a moron*
AF3. Disclosure	AF3a.Vulnerability	*Can´t do it!!!*
	AF3b. Personal life	*Had to travel…*

[5] Measurement and analysis of the frequency of words or segments in a text (Cousteaux, 2009).

For *Interaction* (IN), Table 3, the model contains subclasses related to strategies that, according to Rourke et al. (2001), "build and sustain relationships, express a willingness to maintain and prolong contact, and tacitly indicate interpersonal support, encouragement and acceptance of the initiator". For this reason, indicators like agreeing, complimenting, thanking, offering help, and requests were included. The "reply" resource is also a clear indicator of interaction in online forums.

Table 3. Subclasses and indicators of Interaction (IN) – *PPlus* (Bastos, 2012)

Subclasses	Indicators	Examples
IN1. Sustaining conversation	IN1a. Use of the reply function	*Re:*
	IN1b. Discourse markers	*You know,…*
IN2. Reference to other posts	IN2a. Paraphrasing	*As Mario said*
	IN2b. Quoting	*"Blogs are…"*
	IN2c. Complimenting	*Congratulations!*
	IN2d. Agreeing	*I agree with…*
	IN2e. Disagreeing	*I must disagree…*
	IN2f. Thanking	*Thanks for...*
IN3. Recognition	IN3b. Apologizing	*So sorry for…*
	IN3c. Offering help	*Need any help?*
	IN3d. Motivating	*Upi can do it!*
	IN3e. Sympathizing	*I understand…*
IN4. Questions	IN4a. Clarifications	*How can I …?*
	IN4b. Requests	*Could you…?*

5 Results and Discussion

To find out the number of SP indicators of Affect (AF) and Interaction (IN), we mined 169 student posts in 29 topics, resulting in 203 instances of AF and 161 of IN. With 198 occurrences, subclass Emotion (AF1) exceeds by far Humor (no results) and Disclosure (05 occurrences). The reason for zero results in Humor can be explained by the fact that it is a language expression that can be apprehended by a reader, but

Table 4. Distribution of Indicators of Emotion (AF1)

Indicators of AF		Occurrences
AF1a. Interjections		11
AF1b.Onomatopoeia		14
AF1c.Emoticons and gifs		15
AF1d.Repetitious punctuation		22
AF1e.Repeated letters		23
AF1f. Enhancement	Capitalization	19
	Color	19
	Bold	26
	Italics	02
	Quot. Marks	20

difficult to be identified by a text-mining program. Since the automatic analysis performed by the software *Presente!* is based on regular expressions, manifestations of Humor, though a relevant feature of Affect, were not processed for this study.

As shown in Table 4, all indicators of AF1 are evenly distributed throughout the corpus of analysis. The large use of AF1 cues can be attributed to the students' familiarity with one another and with online communication. Being free to express themselves in either colloquial Portuguese (their mother tongue) or English, their texts present many instances of oral and non-verbal language.

The following post illustrates an intense use of AF1 units in one single message. The student used the "kkkkk" characters to simulate laughter (onomatopoeia), quotation marks to enhance a metaphor, and repetitious punctuation to signal her satisfaction.

	Re: ENGLISH LESSONS ON YOU TUBE by xxxxxxxxxxx - Friday, 22 June 2012, 09:38 PM
	kkkkkkkkkk Muito figura o "docinho de pêssego". Funcionou!! Show parent | Reply

Fig. 2. Example of Emotion indicators in a post

For Interaction (IN), results from the mining experiment show an expressive difference between the "Reply" feature (*continuing a thread* - IN1) in the discussion board (total of 123 occurrences), and the other IN indicators.

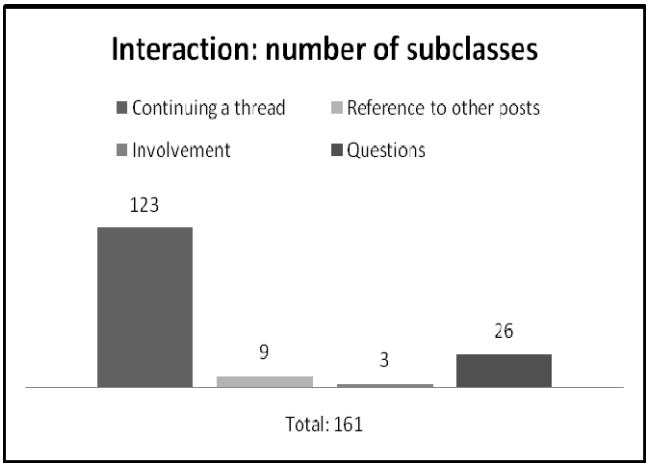


Fig. 3. Number of occurrences of Interaction

The fact that this feature had the highest score indicates that the thread of discussion was achieved, since unanswered or no comments to posts were major concerns of the teacher. Subclass "Questions" (IN4), with 26 instances, includes those via which the student asks for clarification (IN4a) or makes a simple request (IN4b). Other indicators of involvement such as *thanking, apologizing, agreeing,* and *disagreeing* were sparsely used. Such lower results can be justified by the fact that, meeting one another face-to-face on a regular basis, students felt no need to fulfill those functions.

5.1 Survey Questionnaire

The online survey aimed at verifying students' perception of the various resources offered in the environment, and of their participation or attitude toward the environment.

Table 5. Percentages (%) of Students' Perception of Features offered in Moodle-A&UD

Features	Strongly Agree	Agree	Disagree	Strongly Disagree
1.Fostered learning autonomy	36	55	9	0
2.Enabled knowledge sharing	73	27	0	0
3.Expanded contents and discussion	64	36	0	0
4.Offered materials and links of student interest	45	55	0	0
5.Provided grammar support materials	55	45	0	0
6.Promoted student-student and student-teacher interaction	91	09	0	0
7.Established adequate task deadlines	64	36	0	0
8.Added pedagogical value to the course	82	09	9	0

Results in Table 5 show that learners acknowledged the several features offered by Moodle-A&UD, and were satisfied with them. Options "strongly agree" and "agree" received high scores in all items. Interaction in the environment was pointed out by 91% of the subjects, and 82% of them strongly agreed that the hybrid use of a VLE to support and enhance face-to-face classes added pedagogical value to the course.

Table 6. Percentages (%) of Participation and Acceptance of Moodle-A&UD

Features	Strongly Agree	Agree	Disagree	Strongly Disagree
1.You logged in at least once a week	36	27	27	09
2.You completed tasks assigned by the teacher	55	27	18	0
3.You made at least one post in the forums	82	09	09	0
4.You had difficulty in using the environment	18	27	18	36
5.Your difficulty was due to lack of experience with VLEs	0	18	55	27
6.Using the VLE should not be compulsory	27	36	0	36
7.You would have logged in the VLE even if it were not compulsory	36	55	0	9

The questionnaire also investigated issues related to individual behavior in the environment – participation, ease or difficulty in using the platform, and compulsory registration.

Data in Table 6 show greater distribution of agreement and disagreement. Participation in the discussion board (item 3) received the highest score for option "strongly agree". The activity report provided by the Moodle platform shows that all students made more than one post per week – an evidence of their participation and interaction with peers or teacher. Results for item 4 indicate that 45% of the students had some level of difficulty in using tools and resources in the environment. However, these difficulties were not reported in the forum or to the teacher in person.

Participation and tasks in Moodle-A&UD amounted for 40% of the student's final grade. However, results in item 6 show that 63% of the subjects declared that it should be an elective academic activity. Accordingly, 91% of the subjects stated they would have used the environment even if it were not a compulsory component of the course program. This may indicate that learners were willing to take advantage of the benefits of having a VLE to expand their learning experience. These comments posted by participants summarize satisfaction with teacher's methodological strategy[6]:

> *I think your concern in offering this VLE is of great value. It's absolutely pertinent for the course, since language has to do with interaction. Use of the forums and extra material, all in one place, as well the guidelines for projects, facilitate our academic life, making it an efficient and pleasing way of studying.*

> *At first, I thought we wouldn't use it very often, but now I see that the class is participating actively. I'm enjoying this experience a lot, and I think that more teachers should use this platform too. This site enables student-teacher integration outside the classroom; it organizes our academic contents, which is great for us students. Evaluation of our participation in the environment is important as it stimulates us to be always interacting in it.*

6 Conclusion

This study focused on identifying text indicators of affect and interaction in students' postings in a VLE specifically designed for an elective ESP course. The strategy of using an online environment to support and expand conventional face-to-face classes was assessed by a survey questionnaire and by mining units of analysis found in the PPlus Model.

Results obtained by text-mining messages with the software *Presente!* show predominance of affective aspects over interactive ones. Data obtained in the survey questionnaire add to the conclusion that use of a blended-learning (BL) approach was an adequate pedagogical decision. Results concerning affect and interaction indicate satisfactory density of student social presence in the VLE.

[6] Original comments in Portuguese translated by the authors.

Considering that current learners are used to dealing with digital technologies, use of a hybrid methodology may enhance learning as it incorporates the dynamics and resources of hypermedia into academic activities traditionally limited to physical encounters.

References

1. Álvarez, I.A., González, L.P.: Blended-Learning in ESP: an Insight in Moodle. Revista de Lenguas Aplicadas (2007), http://www.uax.es/publicaciones/archivos/LINIDI07_001.pdf
2. Bastos, H.P.P.: Presença Plus: Modelo de Identificação de Presença social em Ambientes Virtuais de Ensino e Aprendizagem. Dissertation. Programa de Pós-Graduação em Informática na Educação. Universidade Federal do Rio Grande do Sul. Porto Alegre, RS, Brazil (2012)
3. Chen, G., Chiu, M.M.: Online Discussion Processes: Effects of Earlier Messages' Evaluations, Knowledge Content, Social Cues and Personal Information on Later Messages. Computers and Education 50, 678–692 (2008)
4. Creswell, J.W.: Research Design: Qualitative, Quantitative, and Mixed Method Approaches. Sage Publications, Thousand Oaks (2003)
5. Daragmeh, A.M.: Teaching ESP in a Blended-Learning Setting. In: 2nd ICONTE, Antalya, Turkey, pp. 105–103 (2011)
6. Garrison, D.R., Anderson, T., Archer, W.: Critical Thinking, Cognitive Presence, and Computer Conferencing in Distance Education. American Journal of Distance Education 15, 7–23 (2001)
7. Graham, C.R.: Blended Learning Systems: Definition, Current Trends, and Future Directions. In: Bonk, C.J., Graham, C.R. (eds.) Handbook of blended learning: Global Perspectives, Local Designs, Pfeiffer Publishing, San Francisco (2006)
8. Graham, C.R., Allen, S., Ure, D.: Benefits and Challenges of Blended Learning Environments. In: Khosrow-Pour, M. (ed.) Encyclopedia of Information Science and Technology, pp. 253–259. Idea Group, Hershey (2005)
9. Gunawardena, C., Zittle, F.: Social Presence as a Predictor of Satisfaction within a Computer Mediated Conferencing Environment. American Journal of Distance Education 11, 8–26 (1997)
10. Hutchinson, T., Waters, A.: English for Specific Purposes: A Learning-Centered Approach. Cambridge University Press, Cambridge (1987)
11. Jackson, K.: What value Assessment Rubrics in Shaping Students' Engagement in Asynchronous Online Discussions? In: Steel, C.H., Keppel, M.J., Gerbic, P., Housego, S. (eds.) Curriculum, Technology & Transformation for an Unknown Future, pp. 454–458. ASCILITE, Sydney (2010)
12. Kambara-Silva, J.: Automatização do Processo de Aquisição de Presença Social em Fóruns e Chats. Term Paper. Instituto de Informática. Universidade Federal do Rio Grande do Sul. Porto Alegre, RS, Brazil (2011)
13. Martin, J.R., White, P.R.: The Language of Evaluation: Appraisal in English. Palgrave/Macmillan, London, UK (2005)
14. Rourke, L., Anderson, T., Garrison, D.A., Archer, W.: Assessing Social Presence in Asynchronous Text-Based Computer Conference. Journal of Distance Education 14, 51–70 (2001)

15. Sahu, C.: An Evaluation of Selected Pedagogical Attributes of Online Discussion Boards. In: Hello! Where Are You in the Landscape of Educational Technology?. ASCILITE, Melbourne (2008)
16. Scherer, K.R.: What are Emotions? And How Can They Be Measured? Social Science Information 44, 695–729 (2005)
17. Vian Jr., O.: O Sistema de Avaliatividade e os Recursos para Gradação em Língua Portuguesa. Documentação de Estudos em Linguística Teórica e Aplicada 25, 99–129 (2009)

GEMINI: A Generic Multi-Modal Natural Interface Framework for Videogames

Luís Filipe Teófilo[1,*], Pedro Alves Nogueira[1,*], and Pedro Brandão Silva[2]

FEUP – Faculty of Engineering, University of Porto – DEI, Portugal
[1] LIACC – Artificial Intelligence and Computer Science Lab., University of Porto, Portugal
[2] INESC PORTO – Instituto de Engenharia de Sistemas e Computadores, Porto, Portugal
{pedro.alves.nogueira,lteofilo,pedro.brandao.silva}@fe.up.pt

Abstract. In recent years videogame companies have recognized the role of player engagement as a major factor in user experience and enjoyment. This encouraged a greater investment in new types of game controllers such as the WiiMote™, Rock Band™ instruments and the Kinect™. However, the native software of these controllers was not originally designed to be used in other game applications. This work addresses this issue by building a middleware framework, which maps body poses or voice commands to actions in any game. This not only warrants a more natural and customized user-experience but it also defines an interoperable virtual controller. In this version of the framework, body poses and voice commands are respectively recognized through the Kinect's built-in cameras and microphones. The acquired data is then translated into the native interaction scheme in real time using a lightweight method based on spatial restrictions. The system is also prepared to use Nintendo's Wiimote™ as an auxiliary and unobtrusive gamepad for physically or verbally impractical commands. System validation was performed by analyzing the performance of certain tasks and examining user reports. Both confirmed this approach as a practical and alluring alternative to the game's native interaction scheme. In sum, this framework provides a game-controlling tool that is totally customizable and very flexible, thus expanding the market of game consumers.

Keywords: Multi-modal, natural interfaces, videogames.

1 Introduction

Videogames and multimedia applications have initially tried to convey increasingly immersive experiences through increased character and environment believability, having in recent years started to dedicate their attention to the interaction artefacts (e.g. the WiiMote™, Kinect™, Move™ and Guitar Hero's controller) [1]. Traditionally, the player is forced to press arbitrary button combinations, which correspond to mapped action in the game world. Often, these combinations are standard (e.g. using the WASD keys to move the game character) or rely on cultural conventions ('R' key for reload, 'F' key for flashlight). Controller-type artefacts, such as the Rock-Band™

* Corresponding authors.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 873–884.
DOI: 10.1007/978-3-642-36981-0_81

instruments, allow players to have a physical mean to interact with the game world. Natural interaction devices (e.g. Kinect™, voice recognition) allow players to control their avatars by acting as if they were actually performing the task within the game.

Despite natural interaction devices lacking physical means, they allow a much greater range of interaction methods and they also have the ability to integrate them thus proving themselves a more powerful tool for interaction research.

An awarded example of this physical medium/natural interaction fusion is the interactive virtual-reality environment Osmose, by Char Davies [19]. In her experiment, Davies merged a kinaesthetic interaction scheme with traditional virtual reality technologies (a head-mounted display and 3D surround sound) to provide the physical medium. In her own words: "*(Osmose) shuns conventional hand-based modes of user interaction, which tend to reduce the body to that of a disembodied eye and probing hand in favour of an embodying interface which tracks breath and shifting balance, grounding the immersive experience in that participant's own body*" [19]. Davies' results show that some test participants had strong emotional reactions to the whole experience, suggesting that applications reporting high immersion levels (e.g. videogames), coupled with suitable kinaesthetic interaction schemes can drastically increase the enjoyability and sense of emotional engagement of said application. The creation of applications resorting to natural (or kinaesthetic) interaction thus enables more engaging experiences that, in turn may capture the interest of a larger audience and facilitate player engagement along the 4 factors proposed by Lazzaro [2].

Regardless of the increased investment by game designers in crafting more engaging experiences, natural interaction is an area that has not yet been thoroughly researched and thus lacks a set of adequate development tools. This work aims at providing such a tool, usable by both the academic and industry fields. It does so by introducing a versatile framework that can be used to quickly develop and test natural interaction schemes (IS) for applications that were not initially designed for them.

The framework allows the user to define a custom IS via a supplied graphical user interface (GUI) and then use it to interact physically, through recorded poses and verbally, via speech recognition, with the intended application. Poses are automatically recorded and recognised through the user's skeleton, which is detected through a Kinect™ device. Words and sentences are recognized as voice commands by using the Kinect's built-in microphone array and Microsoft's Speech API™ in conjunction. Currently only English is supported. More languages will be added as foreign language libraries become more robust. A dedicated library manages the communication and event handling with the WiiMote™ and Nunckuk™.

2 Related Work

2.1 Natural Interaction Modes

As previously mentioned, traditional interaction models in videogames resort to button combinations, implemented through keyboard and mouse schemes, only recently shifting to dedicated and natural controllers. However, users are still limited to the designed (native) IS, not being able to redefine or change it altogether. The flexible

action and articulated skeleton toolkit (FAAST) is a middleware to simplify the integration of full-body controls with games and virtual reality (VR) applications [3]. To the best of our knowledge it is the only approach that tackles this issue, enabling a toolkit for natural interface implementation. FAAST is able to detect various preset poses and map each one to a single entry of (also preset) actions. Its main limitations are restricting the user to the available preset poses, limiting the action mapping to the preset keyboard and mouse keypress dictionary and not allowing the usage of other interaction devices, other than the Kinect™.

This lack of a standard framework for the deployment of natural interaction schemes leaves individuals researching them with two options. To resort to a Wizard of Oz approach, simulating a non-working prototype [14], which in many cases isn't possible (e.g. playing a game or most real-time activities), or to build his own custom solution from the ground up [15, 16]. The latter is often the only available approach, requiring a huge commitment in terms of time and effort, while also limiting this research field to people versed or with access to people versed in computer science. Additionally, it also stifles the growth rate of the field and its adoption by the public, contributing to its loss of popularity.

2.2 Movement Detection

Recent approaches in reliable movement detection have introduced marker-based systems [4], accelerometers [5, 18], physiological sensors [6] and carbon-based strain measurement [17]. While these systems are, in general, accurate they are costly due to the necessary dedicated hardware; and intrusive, by requiring the user to wear the sensors or markers. Some of them also do not measure all of the relevant motions (e.g. strain sensors often do not measure torsion) or provide enough accuracy (e.g. cell phone-grade accelerometers).

With the introduction of the Kinect™ movement detection has become cheap and unobtrusive, alas with some inaccuracy as some of our preliminary tests showed a Gaussian fluctuation of nearly 7 cm on the X and Y planes when the subject was idle. Nevertheless, its cheap price, open source SDK and wide availability induced its use in this study. Despite providing a spatial representation of the user's skeleton, the Kinect™ does not support custom pose or movement recognition. This is an issue that has been vastly studied by the scientific community [7, 8, 9, 10]. Being a complex problem most solutions do not work in real-time or have limited tracking capabilities, which motivated the development of the presented lightweight pose detection method.

2.3 Speech Recognition

Speech recognition (SR) is also a complex problem with a multitude of approaches [7, 11, 12, 13]. The main issue with SR is that it requires a database of recognized phonemes and words, which is difficult to create on the fly, as each instance also requires considerable feature extraction and training. Another pressing issue is that it is difficult to identify various sound sources robustly, as well as differentiate from actual sound sources (speakers) and noise. This issue has been tackled by Shih [13], but has

yet to be implemented in commercial software. While proving itself resistant to the first issue, the Kinect™ is extremely vulnerable to the two foremost ones at medium distances (~2 meters, the distance required for optimal movement recognition [20]).

Microsoft's Speech API (SAPI) is a widely used package with native support that already provides an extensive database for the English language and features various runtime optimizations, which motivated its choice as our speech recognition engine.

3 Conceptual Framework Description

The Generic Multi-Modal Natural Interface (GeMiNI) is a framework meant to support an easy introduction and configuration of any computer compatible peripheral device to work as a game input. It acts as an abstraction layer between device events and the game's default controls. This allows users to experience new interaction methods not originally supported or even been devised by the game's developers. As an example scenario: a first-person shooter game, designed for mouse and keyboard input, could be enhanced using voice commands to trigger actions such as issuing orders to squad members or body poses for crouching, walking or setting traps.

GeMiNI's architecture is conceptually composed of three layers, as depicted in Figure 1. First, the input is captured on the input layer, generating an event type. Different devices with different software drivers (for example, cameras) may output the same type of event (e.g. a captured video frame). The logic layer then translates these into commands that are recognized by the game, according to a user defined IS. More precisely, this is achieved by mapping each event to the game's original controls (e.g. keyboard shortcuts). Lastly, the application layer is responsible to assure that, while in-game, the game actions are invoked when the corresponding events are triggered.

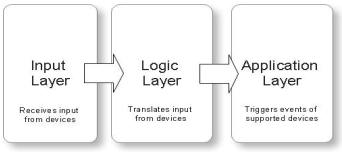


Fig. 1. GeMiNI conceptual Architecture

4 Implementation

The GeMiNI framework is not conceptually restricted to any specific input device. Still, addressing all possible interaction technologies has neither been considered feasible nor relevant at this point of our research. Instead, as a first step, three kinds of device technologies and corresponding events have been considered: Microsoft Kinect cameras, for detecting body poses; embedded microphones for capturing voice commands, and Nintendo Nunchuks for capturing auxiliary key inputs.

For this first approach, these technologies were deemed diverse enough to generate a wide variety of IS. Also, bearing in mind their popularity, maturity and affordability, they were considered the ones where gamers would be more familiarized with

and eager to experiment on. The next section addresses more technical details behind the integration of said devices with the proposed architecture's concept and operation.

4.1 Architectural Integration

The integration of the three interaction device technologies has been performed according to the conceptual architecture explained in the previous section (see Figure 2). Furthermore, to support an easy addition of new features, each layer is divided into independent modules. As such, an interfacing component was built for each device.

For the implemented devices, three components have been developed in the Input Layer (see Figure 2). The Skeleton Module uses the Microsoft Kinect SDK to process a human skeleton structure from the Kinect camera input, and feeds it to the Pose Recognizer module in the Logic Layer. The Speech Module uses the Microsoft SAPI to capture audio from the microphone, and feeds it to the Speech Recognizer in the Logic Layer. Nunchuk Module resorts to the open-source library WiimoteLib and feeds Nunchuk inputs to the general Input Manager in the Logic Layer.

The Logic Layer then processes the input events using the following components:

- Pose Recognizer: Uses our algorithm based on predefined spatial constraints to detect skeleton poses, and passes an identifier of the pose to the Input Manager;
- Speech Recognizer: Uses the Microsoft SAPI to recognize a designated vocabulary from the audio input, and sends the identified words to the input manager;
- Input Manager: Processes specific identifiers, such as corresponding to keystrokes (from the Nunchuk), to pose (from the pose recognizer) and pronounced words (from the speech recognizer). Then it translates these identifiers to specific game actions, according to a predefined mapping description.

Lastly, the Application layer is connected to the Logic layer and is composed of two distinct modules: The Configuration GUI, which allows an expeditious and intuitive configuration of all the necessary parameters of the architecture's components through a graphical user interface; and the External Application, which is executed simultaneously with the framework and reacts to the translated GeMiNI events.

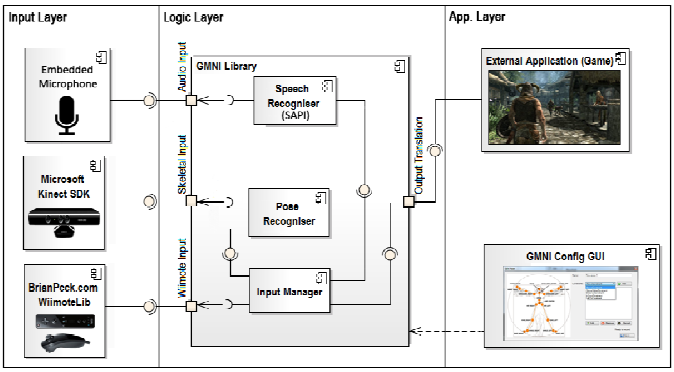


Fig. 2. GeMiNI Implemented Architecture

4.2 Architecture Components

In this sub-section we describe the various components that compose the proposed framework by outlining their features, performance ratings and possible limitations.

Pose Recognition: The skeleton data provided by the Kinect cameras contains the position of 20 distinct points (or joints) from the detected human body (see Figure 3).

Each point is imbued with a semantic identifier indicating the body part and a spatial reference in three-dimensional Cartesian space, relative to the cameras. This introduces the possibility to perform queries concerning the relative location of any body part towards another and its 'absolute' location with the Kinect as the origin. The used coordinate system is shown in Figure 4.

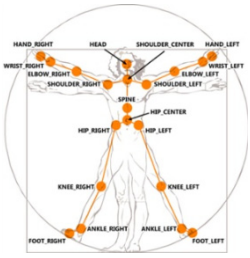


Fig. 3. Skeleton joint disposition[1]

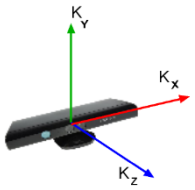


Fig. 4. Kinect Camera coordinate system[2]

Saving a pose by specifying precise absolute location of a joint is neither handy nor feasible, since that it greatly limits possibilities and oversees variations that occur naturally in body posing. Instead, in this approach, poses are specified through sets of spatial constraints, which are verified on body member dispositions. For example, to verify if someone is standing on one foot it is only necessary to consult the Y-coordinate values of the feet joints to check for any prominent differences. Another example: to check if the hands are touching, it is only necessary to calculate the distance between them. This approach allows for a greater flexibility on defining poses.

At this point, the following constraint types have been implemented:

- Distance: Imposes minimum and maximum Euclidean distance between joints;
- InFront: Defines whether a joint is in front of another, by comparing the values of the Z coordinates;
- LeftTo: Defines whether a joint is to the left of another, by comparing the values of the X coordinates;
- AboveOf: Checks if a joint is located above another, by comparing the values of the Y coordinates;
- AboveValue: Checks if the Y coordinate of a joint is located above a certain threshold.

[1] Found at: `http://embodied.waag.org/wp-content/uploads/2011/07/kinect_joints.png`

[2] Adapted from: `http://cvcinema.blogs.upv.es/files/2011/05/convert-kinect-to-standarized1.png`

In order to calibrate the pose detection, each constraint can be configured by fine-tuning the corresponding parameters. Still, single constraints may lead to unwanted body pose recognitions since they impose rather broad definitions. In order to reduce this ambiguity it is possible to resort to more than one constraint simultaneously, so as to define a pose. It is important to stress that this constraint-based approach has proven to be fast, so as to allow real-time detection, with an approximate 16-25ms delay, maintaining the game's frame rate.

Pose Recording: The correct choice and parameterization of the spatial constraints, to define a certain pose, can become a difficult endeavour. For this reason, an alternative to this manual trial-and-error approach has been created. GeMiNI provides completely automatic constraint and parameter definition though pose recording. The Kinect skeleton feed was used to analyse the body's main motion axes during a short (5 second) training phase and infer the relevant constraints and parameters. The method works by the set of joints located in the body's main motion axes (knees, feet, spine, neck, hands and head). Then, it performs peak removal by employing a median filter across a 1.5 second sliding window, with a 0.5 second overlap. This step is done to remove random fluctuations from the recorded signal, as we found the Kinect camera has ~7cm fluctuations on the X and Y planes. The method proceeds to analyse the relations between each possible (distinct) joint pair. For each of these joint pairs, we consider all the possible linear combinations of axes (i.e. X, Y, Z, XY, YZ, XZ and XYZ) and consider that a relevant movement was seen if the maximum observed difference between the joint pair in the recorded data is bigger than a set threshold (15cm in our experiments). Finally, the restriction set is defined as the relevant axis combinations in the analyzed joint pairs.

WiiMote Communication: The use of the WiimoteLib to access the Nunchuk inputs has greatly eased the key interpretation, releasing from the need of additional processing in a separate logic layer component. The choice of a Nunchunk driver implementation was motivated by its button diversity (it has both normal buttons and a D-Pad), popularity and compact form. This was considered a good practical alternative for 2D movement or camera control in 3D applications.

Speech Recognition: Speech recognition features were implemented, allowing any (pronounceable) word or sentence recognition. The system works at a maximum optimal distance of 2 meters and performs speech recognition with a 1 to 2 second delay. Issues found with SAPI included various user identification and noise cancellation, with some sounds from the environment sometimes being misinterpreted (false positives) as voice commands.

Input Management and Simulation: The defined poses, speech commands or external game device outputs can be mapped to a combination of both mouse and keyboard events. Regarding keyboard invocations, there are two event possibilities: Keyboard press, corresponding to a single key-stroke; Keyboard hold, equivalent to holding the key down for certain period, repeating single key-strokes with a certain, configurable frequency. Likewise, mice controls can be simulated in the following manners: Mouse Movement, simulating horizontal and vertical movements; Mouse button press and hold, following the keyboard example.

5 Tests and Validation

For the accuracy and usability tests presented in this section, 25 individuals with ages between 18 and 27 years, 76% male and 24% female with no known physical or mental limitations were recruited. Out of these 25 test subjects, 40% were casual gamers - i.e. reported playing less than 1 hour per day and occasionally whenever big titles are released, also having little to moderate familiarity with videogames in general. The remaining 60% were hard-core gamers – i.e. played an average of 3 or more hours per day and possessed advanced videogame and software applications familiarity.

5.1 Pose Detection and Inference Accuracy

All twenty-five test subjects were asked to perform twenty designated poses, which were recorded and automatically inferred by GeMiNI. Subjects were verbally instructed so as not to condition their interpretation of the required poses, thus enabling the most natural experience possible and also testing the method's robustness. They were then asked to re-enact each one of these poses ten times, to measure the detection accuracy. The poses' inference accuracy was also considered in these tests. The inference of poses is the process of automatically determining the constraints that characterize a given pose. We considered a pose learned if, after the inference process, it is detected with an accuracy of 80%, or more, on subsequent repetitions. This value was empirically defined as we found subjects reported frustration with the system bellow this accuracy threshold. Each pose was repeated 20 times by each participant. Overall results for the test population are depicted in Table 1.

Table 1. Average pose accuracy detection and inference (D – Detection, I – Inferring)

Pose	D	I	Pose	D	I
Step forwards/backwards	97%	95%	Crouch	76%	92%
Lean left/right	81%	90%	Flex	93%	100%
Left/right punch	99%	100%	Point a bow	94%	94%
Lift left/right leg (kick)	97%	100%	Hands behind shoulders	86%	87%
Jump	96%	96%	Outstretched arms	95%	96%
Raise left/right arm	100%	100%	Lean forwards/backwards	86%	87%
Grabbing motion	98%	96%			

5.2 Speech Recognition

For this test, fifty words were randomly selected from the game's item inventory list. These words were also used to generate fifty sentences composed of two or three of those words (as a voice command is usually comprised of one to three words). The set of sentences was then segmented into three complexity categories according to the number of syllabi of the sentences. Thus, these categories represent the simplest to most complex voice commands possible in our test scenario.

Each test subject then repeated a set of 15 random samples from each category 5 times, in order to test the SR accuracy. The tests were performed at a distance of about 2 meters - the optimal distance for body movement capture [20] - from the microphone in a quiet room. The accuracy for this task is depicted in Table 2.

Table 2. Speech recognition accuracy results

Syllabae count	Single words	Sentences
1-3 syllabae	78%	89%
3-5 syllabae	83%	93%
6-8 syllabae	N/A	94%
8-10 syllabae	N/A	93%

5.3 Usability Testing

To evaluate GeMiNI's GUI usability, each of the test subjects were asked to perform a series of tasks (Table 3) that represented each of the previously mentioned steps involved in defining a new IS. The mean values and standard deviations for each action's completion times were calculated, as well as the total number of errors performed by the subjects. These results are present in Table 3.

Table 3. Task completion times and errors

Task	Mean (sec)	Stand. Dev.	Total Errors
New pose (manual)	46.4	14.7	254
New pose (auto)	17.4	5.2	23
New voice command	8.6	3.5	9
Add simple action	13.7	4.8	29
Add complex action	32.6	13.3	99
Set Wiimote button	4.1	2.3	13

6 Case Study – Bethesda's Elder Scrolls V: Skyrim

Despite having developed ISs for various games (e.g. Devil May Cry 4, Super Mario, Legend of Zelda) for our case study we wanted to focus on a videogame with complex and diverse interaction mechanics. Thus, we sought a videogame that provided seamless and complex combat, social and user interface interaction, ultimately choosing the videogame: The Elder Scrolls V: Skyrim. Skyrim is an open-world action role playing game where the player must explore dangerous locations and interact with various factions to further the storyline. Given its genre, the game allows virtually unlimited freedom of movement and interaction through dozens of gameplay mechanics, thus proving an alluring test bed for our framework.

6.1 Interaction Scheme

The game's IS is mainly divided into three categories: exploration, social interaction and combat (exemplified in Figures 5 and 6). Table 4 shows the selected test actions and the relation between their native and newly defined IS.

Table 4. Comparison between native (Nat-IS) and new IS (New-IS)

Action	Nat-IS	New-IS
Move forward/backward	'W/S'	Right foot forward/backward more than 20cm
Strafe left/right	'A/D'	Lean left/right more than 20°
Orientation (look around)	Mouse cursor	Move Wii Nunchuk in desired direction
Invoke map	'M'	Outstretched arms
Initiate a conversation	'E"'	Wave/Say 'hello'
Quit a conversation	'Tab'	Say 'goodbye' or 'see you soon'
Buy/sell an item	'Enter'	Say 'buy/sell'
Equip weapon/spell	'1-8'	Say weapon/spell name
Use weapon / cast spell	Mouse click	Push equipped hand forward
Charge spell	Mouse hold	Raise corresponding arm
Raise shield	Right mouse click	Arm in front of chest with horizontal orientation
Charge at enemy	'Alt + W'	Right foot forwards more than 30 cm

Fig. 5. The "cast spell" and "raise shield" actions, respectively

Fig. 6. The "initiate a conversation" and "invoke map" actions, respectively

6.2 User Commentaries

A pilot study with sixteen test subjects with no previous knowledge of the video game's commands was carried out. Each subject played the game with the native and the natural interaction scheme. Afterwards, players were asked to answer a brief questionnaire, so as to gather data about their preferences. The questionnaire contemplated the three interaction categories. Each of the test subjects was asked to evaluate which

scheme they preferred and rate each IS. Results (Table 5) show that users found the natural scheme to provide a more enjoyable and intuitive user-experience.

Besides the aforementioned evaluation method, the subjects were also asked to assess the natural scheme qualitatively. The subjects highlighted that the system's response time (16-25 *ms*) was adequate. They also gave special emphasis to the possibility of customizing the interaction scheme. Finally, test subjects pointed out that due to the smaller amount of mix-ups in the natural interface, the learning curve becomes faster. This suggests that the games' first impression on users with a natural interface improves considerably thus creating a more addictive experience. Further studies are required to correctly assess the truthfulness of this statement and quantify how much faster the learning curve actually is.

Table 5. Interaction scheme user preferences. Overall, users preferred the natural one.

Game feature	Prefer Native	Prefer Natural	No Preference
Social Interaction	25%	50%	25%
Inventory Management	19%	37%	44%
Movement / Exploration	19%	56%	25%
Combat	13%	81%	6%

7 Conclusions and Future Work

In comparison with existing works, this approach is faster and provides more features, such as new pose definitions and their automatic recording, support for other devices, speech recognition and complex input mappings without the need for third-party software. However two technical limitations were found. Firstly, that the Kinect™ must be distant from the speakers so as to not interpret voices or sounds coming from the game as voice inputs. Secondly, some actions (e.g. shaking someone's hand, opening a door and casting a spell) require some form of context to be correctly identified, which is undoable without access to the game's engine. In other words, it only shows any limitations when the Kinect™ sensor was poorly placed or when used with complex applications that did not previously support natural interaction.

Retrospectively, the system has proved itself capable of delivering an accurate, versatile and satisfactory method for the implementation of multi-modal natural interfaces, as has been proved by our trials. It also succeeded in providing a pleasurable experience in one of the most complex action videogames currently available.

Future work should focus on performing a set of comprehensive immersion studies on a larger population so as to quantify how much physical involvement can benefit the overall experience. Head tracking could be used as an interaction mode or to gather involvement data. Finally, a streamlined version of the GUI could also be created so as to make the framework available to an even broader population.

References

1. Bianchi-Berthouze, et al.: Does body movement engage you more in digital game play? And Why? In: Proceedings of the International Conference of Affective Computing and Intelligent Interaction, pp. 102–113 (2007)
2. Lazzaro, N.: Why we play games: Four keys to more emotion without story. Technical report, XEO Design Inc. (2004)
3. Suma, E.A., et al.: FAAST: The Flexible Action and Articulated Skeleton Toolkit. In: IEEE Virtual Reality Conference, pp. 247–248 (2011)
4. Vera, L., et al.: Augmented Mirror: Interactive augmented reality system based on kinect. In: Interact 2011, vol. Part IV (2011)
5. Kratz, L., et al.: Wizards: 3D gesture recognition for game play input. Future Play (2007)
6. Zhang, X., et al.: Hand gesture recognition and virtual game control based on a 3D accelerometer and EMG sensors. In: International Conference on Intelligent user Interfaces (2009)
7. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech and Signal Processing 26(1), 43–49 (1978)
8. Kinect DTW (2011), `http://www.kinectdtw.codeplex.com`
9. Schwarz, L.A., et al.: Recognizing multiple human activities and tracking full-body pose in unconstrained environments. Pattern Recognition 45, 11–23 (2012)
10. Arantes, M., Gonzaga, A.: Human gait recognition using extraction and fusion of global motion features. Multimedia Tools and Applications 55, 655–675 (2011)
11. Saon, G., Soltau, H.: Boosting systems for large vocabulary continuous speech recognition. Speech Communication 54, 212–218 (2012)
12. Jelinek, F.: Statistical methods for speech recognition. MIT Press (1998) ISBN-13:978-0-262-1066-3
13. Shih, P.Y., Lin, P.C., Wang, J.F., Lin, Y.N.: Robust several-speaker speech recognition with highly dependable online speaker adaptation and identification. Journal of Network and Computer Applications 34, 1459–1467 (2011)
14. Bernsen, N.O., Dybkjær, L.: Exploring Natural Interaction in the Car. In: Proceedings of the CLASS Workshop on Natural Interactivity and Intelligent Interactive Information Representation, pp. 75–79 (2001)
15. Santos, E.S.: Interaction in Augmented Reality Environments Using Kinect. In: Proceedings of the XIII Symposium on Virtual Reality, pp. 112–121 (2011)
16. Linder, N.: LuminAR: a compact and kinetic projected augmented reality interface. M.Sc. thesis. MIT, Cambridge, Massachusetts, USA (2011)
17. Yamada, T., Hayamizu, Y., Yamamoto, Y., Yomogida, Y., Izadi-Najafabadi, A., Futaba, D.N., Hata, K.: A stretchable carbon nanotube strain sensor for human-motion detection. Nature Nanotechnology 6, 296–301 (2011)
18. Lockman, J., Fisher, R.S., Olson, D.M.: Detection of seizure-like movements using a wrist accelerometer. Epilepsy and Behavior 20(4), 638–641 (2011)
19. Text and videos available at the Immersence website, `http://www.immersence.com` (last visited on April 10, 2012)
20. Microsoft Corporation, Kinect sensor manual and warranty (2012), `http://download.microsoft.com/download/f/6/6/f6636beb-a352-48ee-86a3-abd9c0d4492a/kinectmanual.pdf` (last visited on April 9, 2012)

A Semi-automatic Negotiation Strategy for Multi-attribute and Multiple Participants

Rharon Maia, Carlos Dias, Marcus R. Laurentino, and Alisson Vasconcelos Brito

Universidade Federal da Paraiba, Campus Universitário, s/n,
João Pessoa, Paraíba, Brazil
{alissonbrito,rharon,hacks,marcus.rafael}@lavid.ufpb.br

Abstract. This work proposes the definition of a system to negotiate products in an e-commerce scenario. This negotiation system is defined as PLANE – Platform to Assist Negotiation – and it is carried in a semi-automatic way, using multi-attributes functions, based on attributes of the negotiated content. Also it presents an architecture to interconnect the participant through an inter-network in the television broadcasters context. Each participant of the inter-network applies policies for its own contents, and all of them must respect these policies. If a participant needs a content not covered by the policies, it is possible to start a negotiation process for this specific content.

Keywords: Negotiation, inter-network, policy, multi-attribute functions.

1 Introduction

Nowadays television broadcasters produce and keep a large amount of audiovisual content that is distributed among their commercial partners [1]. However the distribution and trading of this material involve complex contractual negotiations between television broadcasters, like the signing of contracts and establishment of rights and duties to be applied to the negotiated content [2]. Furthermore, the search and acquisition of content from other television broadcasters often became hard tasks, due to the lack of an efficient infrastructure that provides means for the broadcasters to connect themselves and negotiate their content.

In this context, the definition of a negotiation system that leverages the sharing of audiovisual content of television broadcasters with other partners is relevant, providing the means to negotiate this content, respecting the contractual policy established between two or more television broadcasters. More than that, the procedures of negotiation must be performed in a semi-automated way, in order to overcome the delay caused by negotiations carried by people.

This work aims to define of a negotiation system for content shared by television broadcasters. To support it, a logical architecture to interconnect the broadcasters was defined, creating an inter-network of broadcasters. Also it is possible to negotiate one or more audiovisual contents among two or more participants of the inter-network through the module called PLANE. This module considers attributes extracted from

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 885–892.
DOI: 10.1007/978-3-642-36981-0_82

the content negotiated, like price, view ratio and aspect ratio, to compute offers and counteroffers in a negotiation session between two or more participants of the content negotiation.

In Section 2 are presented some related work in audiovisual content sharing and negotiation using a semi-automatic process. In Section 3 is presented the architecture to support the concept of inter-network of television broadcasters and its services. In Section 4 the PLANE is shown, a mechanism for semi-automated negotiation of content in the inter-network. In Section 5 one scenario of negotiation with PLANE is present as well as its results. In Section 6 the conclusion is present altogether with ideas for future work.

2 Related Works

In [2], it was proposed the AXMEDIS framework to integration, production and distribution of digital content. The AXMEDIS creates a P2P network of digital content producers where content negotiation is possible through B2B or B2C. In the context of negotiation the approach presented in [3], which is an extended work of [4], presents improvements related to the semi-automatic way of negotiation, such as the number of attributes considered in negotiation and the utilization of a formal rule to generate an offer proposal. This proposal solves many of the problems faced by [4], but it is limited to the use of static attributes in the agreements, not being possible to consider other attributes. Another approach was presented in [4], adapting the concepts of [5] to the context of that work: the use of multi-value functions and weighted attributes, where the latter represents the degree of relevance given by a negotiator. Besides these works, the work of [6] treats the negotiation using one or more attributes, but like [4], actually, the negotiation is made using only one attribute, in this case, the price.

3 Broadcaster Inter-network

Although, it would be easily applied to any kind of product of an e-commerce context, as a case of study this work, focuses on audiovisual content negotiation and sharing through the definition of an inter-network, where other broadcasters select and publish their contents to be shared and negotiated with other participants connected to the inter-network.

In the architecture proposed here, the connection with other participants of the inter-network occurs through PTC (Point of Content Exchange). In a simple comparison, a PTC resembles a router, because it is located in the edge of the network, making the connection with other possible participants of the network. But the PTC is responsible for other functions, like the management of policies, shared content and the negotiation of some content published to the inter-network. Figure 1 presents the architecture of the PTC and its components.

The Peering System is responsible for peering system to manage the connections among the PTCs of the inter-network, including the discovery and establishment of

connections, and which of these connections must be used to forward the content. The Broker is responsible for intermediate the access to all components in a PTC. The Policy Manager defines and validates the established permissions on the shared contents by the inter-network participants. The policies are defined using XACML [8], a language for policy specification. The Content Provider Television is the component responsible for performing the integration of the Digital Asset Management (DAM) of each broadcaster, accessing the content base and creating a catalog of available content.

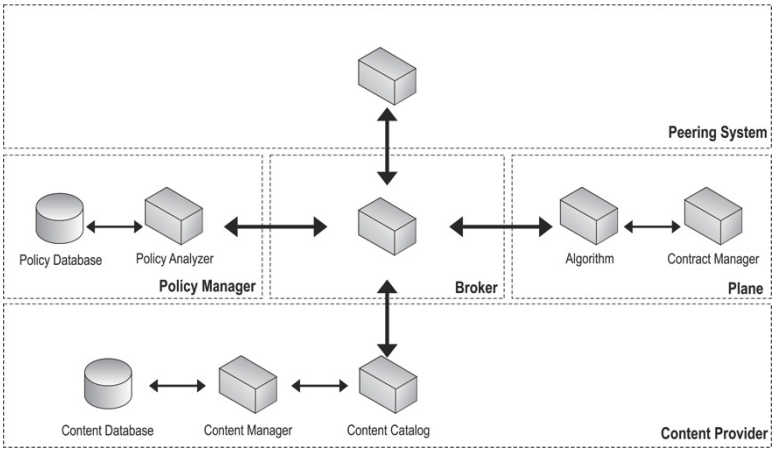


Fig. 1. Point of Content Exchange (PTC) architecture

There are several metadata standards to deal, like MPEG-21 [9], PBCore [10], TV-Anytime [11], making integration of DAM systems a complex task, because it is necessary to understand the semantics of the attributes defined in metadata, to make consistent relations between two different metadata standards [7].

The Negotiation System is the focus of this work. In this component occurs the effective negotiation among two or more participants of the inter-network. The negotiation module is called PLANE (PLatform to Assist NEgotiation), which is composed by a Contract Manager and an algorithm for negotiation. The algorithm analyses and creates new offers, does counter-offers and leads to a negotiation of the contents. In the next section the PLANE is discussed and its operation on the inter-network is detailed.

4 PLANE

For negotiation effectively occurs, it was developed an algorithm that is responsible for generating a set of offerings and counter-offerings, where its main goal is an agreement that brings gains for all parties involved in the negotiation. To understand the operation of the algorithm, prior knowledge of some concepts is necessary: what is an attribute, multi-attribute functions and the weight given to an attribute.

According to [12], attributes can be defined as the characteristics of a product. They can be concrete, observable or measurable of relevant importance. Another more general meaning is given by [13], which defines the attribute as a characteristic qualitative or quantitative of an observed member, in other words, each property that defines an object or entity. The algorithm presented here will focus only on quantitative characteristics.

The negotiation decision can be weighted by a single attribute, but situations like this are not as common. The most frequent problems require the measurement of more than one attribute [14].

The weight corresponds to the degree of importance that a negotiator (buyer or seller) defines to the attribute. It is a value in the range from 0 (least important) to 1 (most important). However it could be any other continuous range (with values belonging to the set of real numbers) that can be transformed into a percentage scale.

4.1 Functions

According to [4], the usage of Value Functions and Utility Functions is necessary for a negotiation tool to formulate possible decision options, where both specify a unique structure of preference. The Value Functions are a particular case of a Utility Function and is embedded in an environment of certainty; a Utility Function can also handle systems involving an environment of luck. In this paper, it was considered only Value Functions. The negotiation process also considers more than one attribute, so here functions with multi-attribute values are used. With this type of functions, it is possible to manipulate the impact of changing attributes throughout the process, making the negotiation flow faster and more efficient.

According to this criterion, three functions were used: a linear function, the total value function and the displacement function, all adapted to our scenario. The following functions are showed below.

$$Linear = \frac{1}{(\mathrm{Wv} - \mathrm{Bv})}(\mathrm{Wv} - \mathrm{Dv}) \qquad TotalValue = \sum_{i=1}^{n} \alpha_i \mathrm{Linear}_i \qquad Displacement = |(Bv - Wv) * (1 - \alpha_i)|$$

Where:

Wv: Worst attribute value Dv: Desired attribute value α_i: Weight attribute
Bv: Best attribute value i: Content attribute Linear_i: Linear Function

The Linear Function is used to normalize the attributes due to different magnitudes. The function Total Value Function is used after a weight is given to the values of attributes. This function is defined as the sum of the result of the chosen Value Function (in our case, a Linear Function) multiplied by their respective attribute weight.

Finally, the Displacement Function is based on the model proposed by [3]. It is used to determine how the algorithm could vary the values of the attributes to suggest a new offer, so that it does not exceed the limits given initially by the negotiators, and represents a better offer than the ones offered so far by negotiators. Thus, the greater is the weight given to an attribute by the negotiator, the lower is the value of its displacement factor.

4.2 Offer Validation

Using the mathematical concept of combinatorics, the algorithm is able to generate new offers to be proposed to the buyers. To do this, the algorithm uses the offer that is desired by the sellers as base to perform variations and than generated new ones to the buyers.

As stated before, content can have several attributes, which can be classified as qualitative or quantitative. To generate a different offer, there must be a variation of at least one of the possible attributes of the content in negotiation.

Before performing a variation in an offer, it is necessary to know how much could be that variation of an attribute in a negotiation. To do so, it is necessary to calculate the Displacement Function. In our scenario, it is considered three attributes, generating 27 combinations of variations of these attributes to be processed and suggested to the participants of the negotiation.

Finally, a validation is still necessary, because the algorithm needs to generate an offer that is within the limits and interests of the buyer. As soon as an offer is generated, an analysis is done to ensure that all the offers fit the buyer's needs.

5 Experiments

The simulation described in this section creates a scenario where a negotiation of a video content with four participants involved takes place: one buyer and three sellers. To execute a negotiation, the PLANE system considers three quantitative attributes related to the content. In this case, for video, was considered price, number of views (how many times a video can be watched) and resolution (aspect ratio of video). For each attribute, it is mandatory to all buyers and sellers to fill three initial values to be considered in the offer and counter-offer: Best Value, Worst Value and Desired Value. Based on these values, the PLANE will decide whether an offer can be accepted or not. The Fig. 2 presents the configuration interface for the values of the attributes.

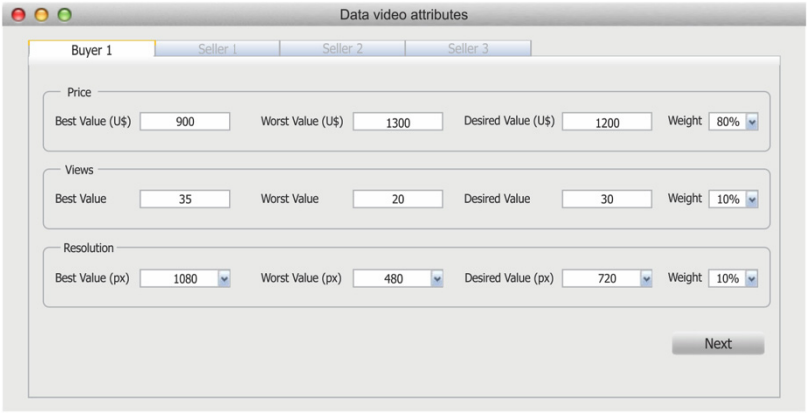


Fig. 2. Configuration interface for the attribute values and weights

After filling all the attributes values, the user requests to the PLANE system to generate a initial round of offers based on the informed values. An offer is generated using the following algorithm: first, it generates combinations where all attributes values increase, decrease or remain the same; second, for each possible generated combination of values, the algorithm uses as reference the Desired Offer value from the buyer and applies a variation in the attributes values of this offer, in order to generate a new one. Also in the second pass of the algorithm, to generate a new offer three important calculations should be applied for all buyers and sellers: the first is related to the variation and it is calculated using the Displacement Function; the second is related to the impact of the variation of the attribute values and it is done using the Linear Function; the third is using Function Total Value, which normalizes the attribute values due the fact these values could use different representation units. The third step of the algorithm tests if the new offer is within the Best and Worst values of the buyer. If it is within these limits, in the fourth step it tests whether the new offer is close to buyer than to the current one. Comparing the values generated from Total Value Function of the new offer and the current offer is possible to know how close they are from each other. If the value calculated by Total Value Function of the new offer is closer to the buyer's current value (calculated in the same way, but considering the Best Offer value), then the current one is replaced by the new offer. The comparison uses Best Offer value from buyer as the base value, as it is considered that the buyer is the decision maker in a negotiation of a content, demanding and choosing the best offer to fit its needs. The algorithm steps will be executed as many times as the user configured in the interface. An example of an execution of the algorithm in PLANE system can be seen in Fig. 3.

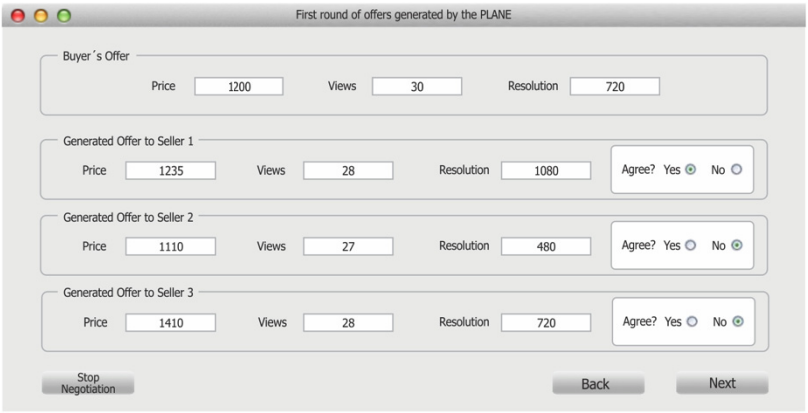


Fig. 3. Interface of the negotiation simulation using PLANE

The interface shows all the offers generated by sellers and the option to agree or decline an offer since the PLANE is a semi-automated negotiation system, depending on adjustments and final decisions taken by humans.

6 Conclusion and Future Works

This paper presents a proposal for a negotiation system that was applied for audiovisual content. This negotiation system is implemented in a semi-automated way, using multi-attribute functions and quantitative weighing of attributes to better negotiate the terms of a possible deal. Using PLANE can bring some advantages such as reducing the time to reach an agreement, semiautomatic negotiation allows for multiple participants. Sometimes the algorithm fails to generate valid offers or there may be biased depending on the values added to the product attributes in negotiation.

As a future research it is being investigated the use of other multi-attributes functions in order to increase the efficiency of the negotiations. Furthermore, it is being investigated how to choose the attributes dynamically at the time of negotiation, giving more freedom for both the buyer and the seller.

References

1. Glasman, K., Peregudov, A., Lichakov, V., Shetinin, B.: Television Broadcast Networks: from centralized automation control to distributed computing, http://www.Broadcasterpapers.com (Published in June 29, 2003)
2. Bellini, P., Bruno, I., Nesi, P.: An Architecture of Automating Production of Cross Media Content for Multi-Channel Distribution. In: First International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution (AXMEDIS 2005), Florence, Italy, pp. 123–133 (2005)
3. Faratin, P., et al.: Designing Responsive and Deliberative Automated Negotiators. In: Proceedings of Workshop of Automated Negotiation, Barcelona (1999)
4. Néri, E.L., Dahia, M.L.M.: Implementing a mechanism of integrative negotiation: challenges and results. Operational Research 24 (2004) (in Portuguese)
5. Raiffa, H.: The Art & Science of Negotiation. Harvard University Press, Massachusetts (1982)
6. Lung, N., Mak, P., Wan-cheng, N., Lian-chen, L., Cheng, W.: A Semi-automated Negotiation Process to Improve the Usability for Online Marketplaces. In: 7th IEEE International Conference on Computer and Information Technology (CIT 2007), October 16-19, pp. 253–258 (2007)
7. Tous, R., Carreras, A., Delgado, J., Cordara, G., Gianluca, F., Peig, E.: An architecture for TV content distributed search and retrieval using the MPEG query format (MPQF). In: Proceedings of the 2008 Ambi-Sys Workshop on Ambient Media Delivery and Interactive Television (AMDIT 2008). ICST, Brussels (2008)
8. OASIS. XACML 2.0 Specification Set (2005), http://www.oasis-open.org
9. ISO/IEC. ISO/IEC 21000-5/FPDAM 1- MPEG-21 - Part 5: Rights Expression Language, Amendment 1: MPEG-21 REL profiles
10. Public Broadcasting Metadata Dictionary Project (PBCore), http://www.utah.edu/cpbmetadata
11. TV-Anytime Forum, http://www.tv-anytime.org

12. Tiecher, A.: Identify product attributes in the smartphone market that add value as perceived by students of School of Management Federal University of Rio Grande do Sul (2011) (in Portuguese), http://www.lume.ufrgs.br/bitstream/handle/10183/33341/000787035.pdf
13. Ferreira, A.: The New Aurélio Dictionary of Portuguese Language, 3rd edn. Positivo (2004) (in Portuguese)
14. Baily, P., et al.: Shopping: Principles and Management. Ed. Atlas, São Paulo (2000) (in Portuguese)

Smart Land Record Application Using Web GIS and GPS

Muhammad Shahid Bhatti*, Muhammad Ajmal, Atif Saeed, Maqsood Ahmed, Rizwan Khalid, and Nouman Arshad

COMSATS Institute of Information Technology defence road off Riwind Road Lahore, Pakistan
{msbhatti,rajmal,asaeed}@ciitlahore.edu.pk,
{maqsoodciit,noumanarshad84,razi.professional}@live.com

Abstract. Land is a natural resource for the livelihood of human and basic element of wealth generation. The land plays an important role in agricultural, residential and environmental advancement of the country. Due to rapid growth rate in population, mortgage ability, cultivation and industrial areas enhanced the importance of land record management. The system is design to provide maximum accuracy, accessibility and portability. In our system GPS and GIS technologies used to manage land records. GIS and GPS technologies are used to preserve the land record that helps in reducing the boundary disputes and automatically localize land boundaries irrespective of the changing conditions of the physical location of land. This application eliminates illegal land shrinkages and expansions based on the moving physical boundaries (more often the case in rural areas and agricultural/orchards). These Land records are helpful for taxation and provide protection of the rights and ownership information of land owners.

Keywords: GPS, GIS, Centroid Based Search, Khewat.

1 Introduction

Internet and mobile technologies are growing rapidly advanced and used for information broadcasting. The Smart Land App is designed with such tools and features which are easy to understand and common on web pages. To make sure information is accessible anywhere, anytime the application sends the information to user through the internet. Smart Land app initially brings data manual to digital. It manages the data of land resources by using Database indexing, GPS and Web GIS.

The system automatically retrieves the patch information based on the physical position of the owner having GPS device and access rights. This system automatically generates the GPS coordinates of a boundary of patch of any shape and size when walking along the boundary while holding active smart app or a GPS device. The smart Land app helps to control illegal land shrinkages and expansions based on the

* Corresponding author.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 893–900.
DOI: 10.1007/978-3-642-36981-0_83

moving physical boundaries (more often the case in rural areas and agricultural land/orchards etc.). This Smart Land app assists users by providing maximum accuracy, accessibility and portability.

Our purposed system include advantages of the land record management systems which are unallocated land identification, duplicate ownership detection, resource allocation and assessment, reduce corruption, taxation and one centralized database for whole system.

1.1 Literature Review

To transforming the digitized system into an automatic computerized land management system Ehtesham [1] proposed a technique. In which images are scanned and then converted into scalable vector graphics with the help of image processing. Scalable vector graphics data are managed in an XML file format which is saved into a geospatial database with other associated data and searching is based on patch number. This system is technically exercisable with respect to updating and editing. Converting paper image to digital image using scanner is a time consuming process.

Hao Zhang [9] Google maps API and XML data-storage technologies are used to build a land information system which will be networked and released. Google API facilitates to draw a patch while latitude and longitude are provided. The Google map data is saved in the database in the form of points, lines and polygons. For querying a specific patch again Google API is used, local search and reverse address resolution. This system is convenient and open source but this system uses XML format to store data which has a data overloaded issue and searching becomes slow.

RosmadiGhazali [10] is using patches of land using remote sensing and vector polygons. After detection of patch changes are detected in it by using the remote sensing images. This system takes the information between old date and new date data to detect the change in the land. The data taken from remote sensing are images and coordinates of the surveyed land. This method compares the information of old and new patches after some period of time by using the threshold technique. The system detects the changes efficiently and increases the automation of land-use analysis. The mix patches which are partially changed interrupt the detection. This algorithm cannot detect internal change in pixel and cannot detect editing or tempering.

QingzhouLuo et el. [11] The Pangkalan Data Ukur Kadaster(PDUK) data comprises parcel information such as distance, coordinate, area and boundary marks. Handheld GPS devices are integrated with PDUK database. GPS device exchange the waypoints, routes and tracks with web. The map is available for surveyor on a handheld GPS device. This device shows all nearest lot parcels along with their coordinates. The cadastral surveys can create, define and subdivide the public land, especially to solve boundary disputes. The surveyor takes long time to find the work location for the survey. For the confirmation of the lot parcel they need land owner. Who verifies and guide at the time of the survey.

Shouyi [2] transfers the geographic data on the internet for the land management. Electric maps are collected through Geo-server. Attribute data and graphical data are

managed and stored independently into a Geo spatial database. For searching a specific patch information Ajax queries are performed. Web-GIS are designed in such a way which is user friendly. People can take and share geographic information from anywhere by using the Web-GIS. The main purpose of the system is Information inquiry, information visualization for efficiency and reliability of data. For system safety purpose administrative privilege and end user have different rights.

KasphiaNahrin [3] is a realistic method for developing a land administration and a rapid source to maintain and manipulate land records. Patches would be extracted through GPS surveys, associated attribute information could be downloaded from GPS devices directly into a GIS database. The overall procedure of Land Information System is to take Satellite images, conventional land record and then builds up a geodetic survey network and starts extensive GPS surveys. Land records will support in taxation, subdivision forecasting and Infrastructure development. Land information system accomplishes different analysis on the bases of previous records but the key issues are lack of administrative coordination and training for human resource.

In [4] deliver evidence for manipulating factors and assessment for the utilization of land. The basic purpose of the system is to make the comprehensive index. The process is to maintain the attribute database in which graphic operation, information query, statistical analysis and visualization and graphic output involve. In this system architecture GIS technology is used to perform visual expression, management and data acquisition. By the integration of GIS system we manage spatial and attribute data in an efficient way. Obtaining and processing the information is the complex factor of the system. LMDE system classified into graphic data, attribute data and multimedia data.

Yuxiang Cao [6] is a practicable technique that is accepted by different countries of Zhejiang and Jiangsu etc. LUPGS method ensures land record management and automation with the combination of ArcGIS platform, Oracle database, .NET development tools. This system can increase the efficiency of work, efficiently work in new situations and avoid substantial duplicated mechanical work. It produces information such as maps, reports and documents which can help in many other perspectives.

Abbas [7] is using to predict future trends of land variations like agricultural, water body or residential areas. GIS technique will help to analyze future Food crisis from existing trends. If the trend changes then it may predict erroneous. Ground surveys and GPS coordinates takes to confirm the satellite images and maps.

Zhang [8] technique analyzes the land information data and gives some suggestions. These value able suggestions give support to economy organization for planning new urban areas. It allows the organizations to place industry and research project to develop the construction of land according to the plan. Make the approval system stricter for acquirement of land. Vital objective is to guarantee the farmers' ownership to secure his life standard.

1.2 Existing Land Record Management Systems

The Land Record and Management systems in the subcontinent are operated manually. This system consisted of delegating power in correspondence of the local governance under supervision of the Board of Revenue. This system has two types of data.

1.2.1 Geographic Data (Maps)

Geographic data consists of handmade maps. These maps are a reflection of the land parcel map with "quantifiable measurements". There are of two types of maps respectively small maps and large maps, where small maps are developed at the time of settlement and cannot be changed before next settlement on the other hand large map is a reflection of all small maps of the particular village on a big cotton cloth.

1.2.2 Attribute / Alphanumeric Data (Registers)

Attribute Data contains Field book, Owner record register, Mutation register, Inspection register and daily dairy. Field book is an alphanumeric representation of the small maps. Owner record register contains ownership records, Mutation register dynamically changes at the time of mutation of land, Inspection register is crop inspection register and daily diary is a daily report register in which essential daily proceedings are written like land transfer or mutation, crop physical condition, pesticide attack on crop and other parallel activities.

1.3 Flaws of Conventional Land Record System

In Conventional Land Record System rights of land are recorded and managed in registers which is an old fashioned and inefficient way. Data is stored in decrepit conditions and unsafe. Manually generated records have poor accuracy. Because of this procedure land right holders and landowners can't access their record easily and mutation of land has become a time consuming process. There are more possibilities of fraud in the conventional land record system. A single person in the land record office can modify the record in South Asia.

To avoid from government taxes, peoples who are close to village accountant transfer their land directly without proper registration. Sometime peoples pay some cash to village accountants to bypass standard operating procedures to transfer ownership. Village land accountant is the key person to parcel boundaries and temper records without any check on him. Due to Special syntax for land parceling, intimate knowledge, sovereignty over system village accountants misused this system. They forced the landowners to pay under the table money to them for their right.

2 Proposed System

Our proposed system is based on 3 layer architecture which is shown in (Fig. 1).

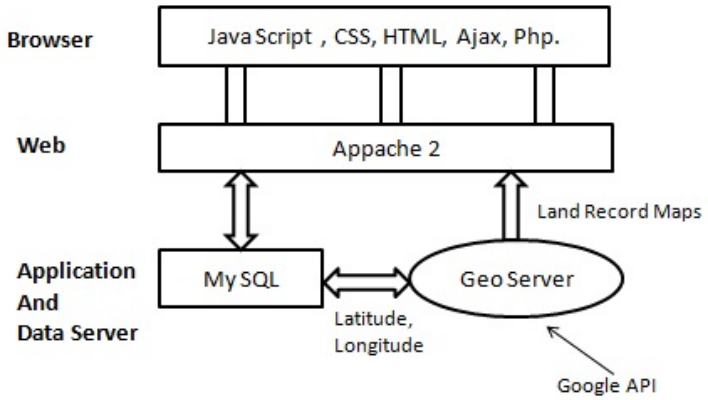


Fig. 1. Architecture of proposed System

At present Land Management System need to maintain a large number of land records on paper. District governments use ancient management skills to keep and maintain all land records in different registers. The present system makes it complex to search records and modify them in the register. The present system is slow, messy and unreliable. This system should be replaced with an automated digitized system.

For an automated digitized system, firstly we need to digitize the current data by entering it in a database. The most significant and important part in land record data is map. Currently these maps are available on clothes (handmade). For digitizing these maps we use GIS technology and Google API's. We use GPS survey data (latitude and longitude) for extracting patch boundaries and use GIS collectively with Google API's to draw latitude and longitude values on the digital map and save these coordinates into a Geo spatial database. At the same time we calculate centroid of each patch and save in database.

2.1 Patch Detection

To detect a patch and save it in the database surveyor move along with the boundary of a patch while holding a GPS device in the hands (Fig. 2 -A, B, C). The GPS device transmits the coordinate's values to the spatial database in a sequence. While surveying if boundary overlaps with any neighbor's patch then system will generate a warning to the surveyor by checking point in polygon problem method. After completing the one the system will automatically calculate the centroid of the patch and save it in spatial database as an index. At the end boundary of the patch is drawn on the map.

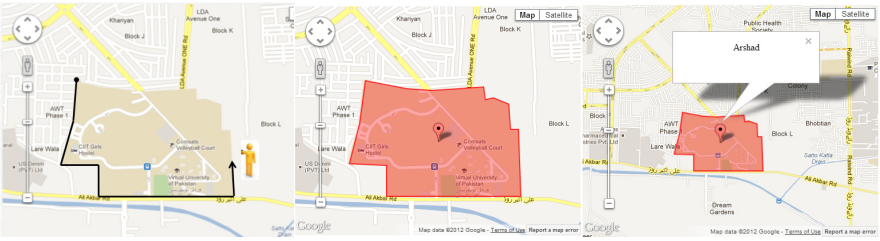


Fig. 2.

A) Extracting boundary of a patch by moving along the boundary While holding a GPS device.

B). Drawing a complete patch boundary where balloon representing Centroid.

C). As searching result Marker shows the Owner name.

2.2 Database Working

Database structure is given (Fig. 3) which is given below**.**

Geographic data and Attribute data are managed collectively in the database. Every patch and owner has a unique ID.

Patch ID	Owner ID	Kehwat No	Kahasra No	Land Type	Index Centroid–Lati	Index Centroid-Longi	Date
1	5	8	53	Commercial	31.53492	74.64263	12/11/2012
2	7	6	3	House	28.56322	74.38721	16/11/2012
54	8	34	5	Plot	35.3424	78.65321	22/11/2012

Fig. 3. Database table structure with centroid coordinates

2.3 Searching a Patch

Searching is based on the centroid indexing of polygons. When we use the smart app to search for any patch we give the manually values or go to that place to give the input value of the required patch. Smart app sends the coordinate values (latitude, longitude) to the system and start searching based on the coordinate value. The algorithm searches the centroids nearby to the required value. After this it searches the corresponding polygons of the centroids which are nearest to find the exact value. After identifying the required patch or polygon it is plotted the on the map and along with the detail of the owner of the patch. The indexing on centroid based optimized the search of any patch. Our technique is faster because we are using centroid based indexing for fast search. We are not searching that point in all polygons stored in our database. Instead we compare that point to nearest centroid.

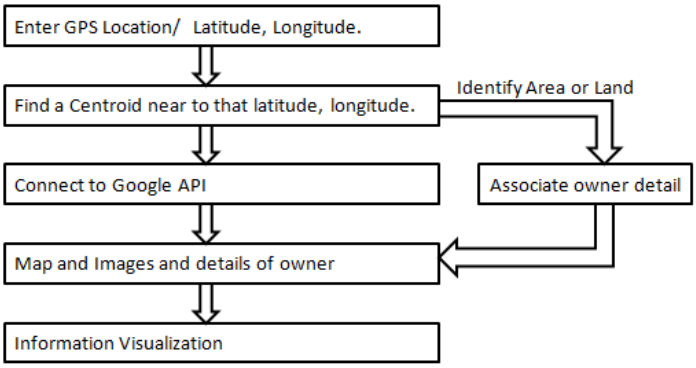


Fig. 4. Centroid based indexing fast searching procedure

3 Conclusion

Our research showed that GPS technology can help to achieve best system for land records. We used latitude and longitude for storing and searching a land patch/parcel. Moreover by the used of centroid as an index has make searching very fast. By using GPS technology we detected the difference between official documented record and physical data. This system makes very easy to manage land record to facilitate government for crop estimation, agricultural land taxation, to avoid fraudulent in tempering land record and beneficially transfer of ownership is very easy with absolute accuracy. In future we can use this system for toll collection using government administrated units and also for resource distribution.

References

1. Choudhury, E., Ridwan, M., Awal, M.A., Hosain, S.: A Web-based Land Management System for Bangladesh. In: ICCIT 2011, Dhaka, Bangladesh, December 22-24 (2011)
2. Lin, S., Xu, J., Li, Z., Yang, Y.: Design and Implementation of Lujiazui Land Management Information System Based on WebGIS. In: 2010 International Conference on E-Business and E-Government, IEEE (2010)
3. Nahrin, K., Shafiq-Ur Rehman, M.: Land Information System (LIS) for Land Administration and Management in Bangladesh. Journal of Bangladesh Institute of Planners 2 (December 2009)
4. Yang, D.-Z., Li, L.: Design and Analysis of Sustainable Land Management Evaluation System. In: Second International Conference on Computer and Electrical Engineering 2009 IEEE (2009)
5. Dacey, S., Barbour, R., Fernando, A.: Participatory Land Use Management of Public Spaces in New Zealand. Joint Urban Remote Sensing Event, Munich, Germany, April 11-13 (2011)
6. Cao, Y.: Management Information System Design and Implementation of Land Use Planning By GIS Technology. IEEE (2010)

7. Abbas, Arigbede, Y.A.: Sustainable land management in Zaria using Remote Sensing and GIS techniques. Journal of Ecology and Natural Environment 3(4), 131–138 (2011)
8. Zhang, W.-J., Tang, D.-S.: Empirical study on the utilization and management of collective construction land in the urban planning area. IEEE (2010)
9. Zhang, H., Li, M., Chen, Z., Bao, Z., Huang, Q., Cai, D.: Land Use Information Release System Based on Google Maps API and XML. Key Laboratory of Geoinformatics of State Bureau of Surveying and Mapping, No.200712 (2007)
10. Ghazali, R., Latif, Z., Rasam, A.R., Samad, A.M.: Integrating Cadastral GIS Database into GPS Navigation System for Locating Land Parcel Location in Cadastral Surveying. In: 2011 IEEE 7th International Colloquium on Signal Processing and its Applications (2011)
11. Luo, Q., Wang, P., Wang, Y.: Study on a kind of Land-use Change Detection Based on Patch Object. In: International Conference on Information Science and Technology, Nanjing, Jiangsu, March 26-28. Chinats of Remote Sensing Image (2011)

Polynomial Approximation of the Battery Discharge Function in IEEE 802.15.4 Nodes: Case Study of MicaZ

Odilson T. Valle[1,2], A. Milack[2], C. Montez[2,4], Paulo Portugal[3], and Francisco Vasques[4]

[1] IFSC, Federal Institute of Santa Catarina, Campus Sao Jose, Brazil
odilson@sj.ifsc.edu.br
[2] UFSC, Department of Control and Automation, Federal University of Santa Catarina, Brazil
{amilack,montez}@das.ufsc.br
[3] ISR / [4] IDMEC, FEUP,
University of Porto, Portugal
{pportugal,vasques}@fe.up.pt

Abstract. An important feature for the design of algorithms for Wireless Sensor Networks is the estimation of its lifetime. To design communication protocols and data collection algorithms with low power consumption, a key factor is the ability to measure the actual consumption in the sensor node. This ability enables the verification of theoretical/simulated models, allowing an accurate estimation of the network longevity. This paper proposes a system for measuring the energy consumption levels of sensor nodes in normal operation, allowing the estimation of the real lifetime for each node, enabling the optimization of energy aware protocols. Presented data is related to the energy consumption of a MicaZ node and to the behavior of the full discharge of a battery pack. A set of polynomial functions describing this energy discharge are also presented that can be easily inserted into the battery models of network simulators to obtain a more realistic behavior.

Keywords: Wireless Sensor Networks, Energy Measurement, IEEE 802.15.4.

1 Introduction

Wireless Sensor Networks (WSN) faced enormous development and applicability in recent years due to advances in microelectronic systems that have driven the development of smart and low cost sensors. These sensors allow monitoring a wide variety of physical phenomena [1], [2] and therefore may be used in various automation environments, such as building automation and industrial applications among others [3], [4].

Intelligent sensors are basically formed by a controller, memory, sensor/actuator, communication device and power supply (Fig. 1). Processor and memory support small programs related to monitoring and decision making algorithms. The sensor/actuators are directly associated with the application to which the node is designed, for example, temperature and pressure sensors may be used to monitor

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 901–910.
DOI: 10.1007/978-3-642-36981-0_84

these magnitudes in industrial environments. Moreover, they can be used to control on/off actuators in particular applications. The communication device typically use wireless communication over radio link, where one of the most widely used standards for the physical and data link layer is the IEEE 802.15.4 [5]. The energy source is typically provided by a set of batteries allowing the remote deployment of nodes, with consequent limitation for its working life since the stored amount of energy is finite.

Due to possible deployment in difficult locations, the cost associated with replacing the batteries is a major concern in the design of WSN. Therefore, it is of paramount importance to reduce the energy consumption of nodes. In this sense there are some proposals for the self-administration of the network in what concerns energy consumption, since the nodes are usually equipped with mechanisms for checking the voltage of its battery, as is the case of MicaZ nodes [6]. However, such energy monitoring mechanisms increases the energy consumption, which degrades the life of the network and interferes with correct measurement actions. Therefore, there is a need for the accurately determination of the energy consumption of any node without interferences to target program.

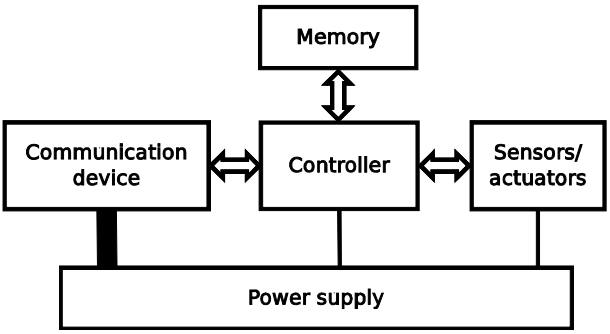


Fig. 1. Model of a node

An accurate energy consumption measurement system is crucial to assess theoretical models and/or simulated in experimental tests. This system should have low cost and be easy to build and operate. The system should also allow the collection and storage of data autonomously, to allow extensive testing without requiring the constant presence of any operator.

In this sense, the present work proposes a measurement system for energy consumption that can be coupled directly to any type of sensor node. The target is to proceed with the energy consumption measurement of a node during normal operation, without interfering with its operation and without the need of energy monitoring algorithms in the node itself.

The main contributions of this paper are: *i)* the proposal of a low-cost system for measuring energy consumption of sensor nodes without interfering with its normal operation, *ii)* the measurement of the power consumption a of MicaZ node in various operation modes, *iii)* the total discharge curve of a battery pack coupled to a node periodically collecting and transmitting data, *iv*) determination of polynomials which represent the discharge curve of a battery pack that may be used in network

simulators and *v)* a simple and fast method for error determination, that can be used to minimize measurement errors.

The remainder of this paper is organized as follows. The detailed description of the proposed measurement system is provided in Section 2. In Section 3 it is presented a case study where a MicaZ node is subjected to various operations, as well as one discharge curve of battery pack. In Section 4 it is presented an estimation of errors of the measurement system for our prototype. In Section 5 some related works are presented. Finally, Section 6 shows the conclusions.

2 Measuring System

The goal is to build a simple system for measuring energy consumption, with easy implementation and low cost. The main idea is that all components must be easily accessible and available in most electronics labs.

The proposed model for the measurement system is shown in Fig. 2, where the Node under consideration has its voltage (Vn) and current (In) physically measured, allowing the calculation of its power consumption ($P = V \cdot I$). The Acquisition Board samples these values, converts them to binary values and transmits them to the PC (Personal Computer). On the PC, a data collection system can program the sampling frequency and store these values associated with the time of each sample (as the Acquisition Board is connected to the PC, they share a common time base). In possession of stored data it is possible to use the system for data processing and, therefore, to determine the power and energy consumed by the node during a given time interval ($E = P \cdot \Delta t$) and/or a specific operation that the node is running.

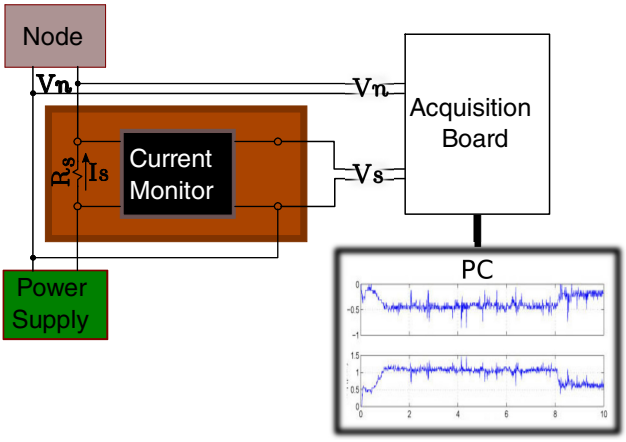


Fig. 2. Measuring system of energy consumption

The node under measurement is used without any modification, except for the access to the positive supply for the insertion of the monitoring probe. The power supply can be normally performed by a set of batteries.

2.1 Current Monitor Circuit

To allow easy data acquisition on a PC, where you can easily sample voltage values, we chose to use a high precision current monitor – ZXCT1022 [7]. This type of device converts current values into equivalent values of tension, through the resistor sensing. This allows the use of common data acquisition boards that in general can capture only voltage values.

The use of this type of device eliminates the need to interrupt the ground plane for monitoring a load current. Furthermore, the ZXCT1022 provides a fixed gain of 100 for applications where minimal sense voltage is required, as it is the case in the WSN node. The very low offset voltage enables a typical accuracy of 3% for sensing voltages of only 10mV, according to the manufacturer.

In our system we used an accuracy resistor sensing – R_s – with nominal value of 1Ω. For the case of MicaZ in sleep mode (critical case), its current is announced as 1μA, which gives a voltage of 0.1mV value that is above the minimum sensitivity of the acquisition board.

2.2 Data Acquisition Board

Using a data acquisition board is an option that has several advantages: *i)* low cost, *ii)* seamless integration with PC that facilitates the data acquisition, storage and manipulation and *iii)* the possibility of using multiple channel acquisitions that allows monitoring of multiple nodes simultaneously.

We selected the multifunction data acquisition board – NI PCI-6251[1] – which converts the analog voltage to digital values, with a bounded sampling rate up to 1MS/s, that is, a data sampling at every 1μs. This board has high speed 16-bit resolution and 16 analog inputs, which is more than enough for the acquisition of two voltages (V_n and V_s) levels in a node. It is sensitive to variations of up to 6μV, which in the test case is equivalent to current variations of 6pA. This acquisition card can be used to simultaneously monitor the power consumption of up to eight sensor nodes.

2.3 PC

On the PC, with the acquisition board properly installed and configured, it is easy to configure the software for data collection and storage. An easy approach is to use a spreadsheet for data processing, where one can compute the instantaneous power of each sample and then calculate the average power consumed during a particular time interval. After processing, it is possible to use a graphical utility to generate graphs and comparatives assessments according to specific requirements. A good example of such tool is Gnuplot[2]. Another possibility is to use the LabVIEW®[3] software, which

[1] http://sine.ni.com/nips/cds/view/p/lang/en/nid/14124
[2] http://www.gnuplot.info/
[3] http://sine.ni.com/nips/cds/view/
modelpopup/p/pcat/\\829/lang/en

allows to carry all the steps, from the acquisition, storage and manipulation of data through a simple and intuitive interface.

3 Case Study: MicaZ under IEEE 802.15.4

The purpose of this section is to describe the behavior of the energy measurement system and to display some relevant issues for their use. For this purpose we use a MicaZ node running TinyOS[4] micro-kernel, version 1.1.15. Over the TinyOS is running the OpenZB[5] version 1.2, which implements the protocol stack specified by IEEE 802.15.4. The system performed 100 data collections (voltage and current) per second.

Fig. 3 shows the current and voltage curves of a node set in various operational modes that are changed every 60s, interleaved by a sleep interval. The goal is to differentiate the various types of energy consumption for a node loaded with its standard embedded software.

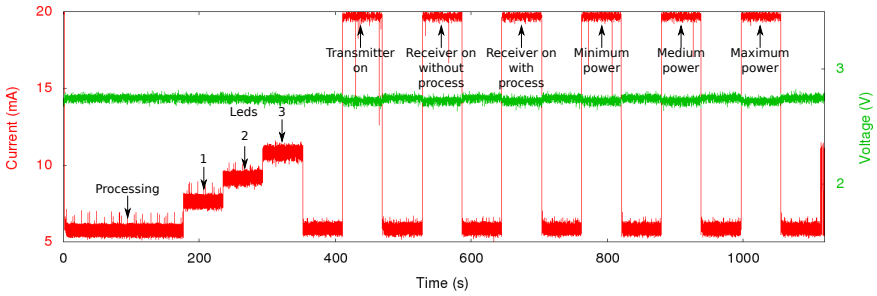


Fig. 3. Current and voltage of node in different states of functioning

What is striking is that the energy consumption of the node (in this configuration) does not meet perfectly all the commands and that there is no significant change in current. First of all, what we see is that there is no change in current when node is subjected to heavy processing operations, the explanation for this may have two origins: either the processor node is always active or operations consumption is negligible when compared to other consumptions.

It is also noticeable that each LED is responsible for the increase of about 1.6 mA or, otherwise, 4.3mW. Therefore, it is not recommended to use LEDs in applications target.

This strange behavior is observed when the radio is on, when it is clearly noticeable that whatever the power set, either in receiving or transmitting mode current drain is the same. Analyzing only the transmission modes (last three peaks), it is clear that the system does not respond to requests for adjusting the radio transmit

[4] http://www.tinyos.net/
[5] http://www.open-zb.net/

power (MCPS_DATA.setRFPower) which allows values between 0 and 31, where we used the values 0 (minimum), 15 (medium) and 31 (maximum).

Fig. 4 illustrates the behavior of a battery pack discharge in a MicaZ node. In this experiment, a node was programmed to light all three LEDs and to continuously transmit at full power. The battery pack is formed by a pair of new common batteries of manganese dioxide, carbon and zinc chloride. The purpose of using this type of battery is to shorten the test period. The system performed a collection of data every second.

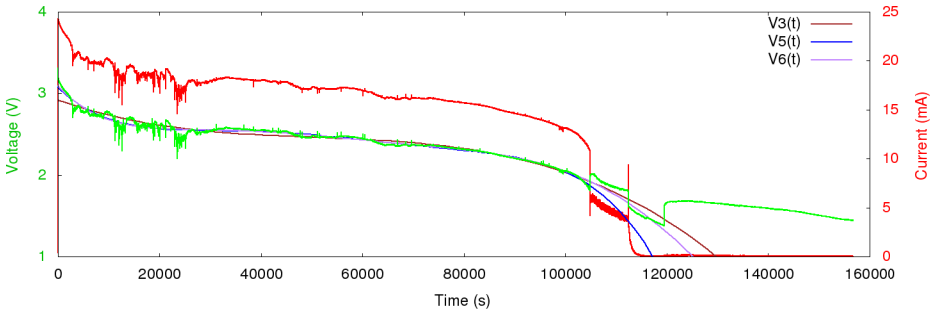


Fig. 4. Discharge curve of a battery pack under a node

The Figure shows that the node behaviour was normal until about 29 hours (105,000s), where the current abruptly dropped and the voltage jumped up, and then went into a mode of irregular operation. Around 31 hours (112,500s) latter, the node almost ceased to drain any current, where the voltage has taken a new leap and fell down slowly. In this figure we also illustrate the polynomial approximation for the voltage curve.

Analyzing the period of regular operation, we can draw the following conclusions. First, the node behavior is typically resistive, i.e., as the voltage drops, the current falls proportionately. This clearly indicates that there is no voltage regulator effect, which leads to undesired behavior of the type: the lower the input voltage, the smaller the radio power of the transmitted signal. This can be critical when the nodes are working at the limit of their transmission distance, leading to premature death of the network.

Another issue that draws the attention is the nonlinearity of the voltage drop. There are several studies that propose routing algorithms based on the residual capacity of the battery, such as those discussed in Section 11.3.1 of [8]. Other works use this type of factor to adjust the data fusion algorithms such as [9]. The way to measures the current energy state of the batteries is measuring its voltage level and this nonlinearity will cause unwanted oscillations in algorithms, making their convergence difficult. Evidently this type of nonlinearity is directly related to the type of used battery, one must make a preliminary survey of this behavior for used battery set or designing algorithms that are immune to it, for example, by introducing some degree of hysteresis.

The oscillation of the voltage and current in the initial period (up to ~25000 s) is typical in discharge curves of batteries of zinc chloride[6] type. For other types of batteries behavior can be distinguished.

The polynomial approximation of the voltage curve is aimed at obtaining functions that can be implemented in a network simulator and that faithfully represent the battery discharge curve. With the numeric processor Octave[7], we obtained polynomials of order 2-11 representing the functional part of the voltage curve, i.e., up to 102000s. In figure are shown curves obtained by polynomial of order 3, 5 and 6, which were the closest to the voltage curve obtained. These polynomials are:

$$V3(t) = -2.5446 \cdot 10^{-15} \cdot t^3 + 3.8635 \cdot 10^{-10} \cdot t^2 - 2.2062 \cdot 10^{-05} \cdot t + 2.9178$$

$$V5(t) = -1.8603 \cdot 10^{-24} \cdot t^5 + 5.2037 \cdot 10^{-19} \cdot t^4 - 5.4938 \cdot 10^{-14} \cdot t^3 + 2.6468 \cdot 10^{-09} \cdot t^2 - 5.9981 \cdot 10^{-05} \cdot t + 3.0704$$

$$V6(t) = 1.4807 \cdot 10\text{-}29 \cdot t6 - 6.3913 \cdot 10\text{-}24 \cdot t5 + 1.0456 \cdot 10\text{-}18 \cdot t4 - 8.3508 \cdot 10\text{-}14 \cdot t3 + 3.3753 \cdot 10\text{-}09 \cdot t2 - 6.7412 \cdot 10\text{-}05 \cdot t + 3.0885$$

By simple inspection of the polynomial curves, it can be perceived that V5(t) is the best choice and is easily deployable in a simulator. This type of curve can be used as a basis to generate a more realistic function for the battery discharge, and this function can be inserted into network simulators, aiming to improve the models and the results of the simulations. Obviously, one should get these curves and functions for each type of battery used in sensor nodes.

4 Errors of the Measurement System

The curves presented in Section 3 illustrate the behavior of a MicaZ node. However, for more accurate values of current and voltage, it is necessary to estimate the errors of the measurement system and, for the case of systematic errors, to calibrate the measurement system. The calibration can be made through the software for data analysis.

For our system, Table 1 presents data used to calculate the percentage of relative errors. To create this table, it was used a set of precision resistors, each of them replacing a node as in the system shown in Fig. 2. Moreover, a digital multimeter was placed in series, aiming to measure the exact value of the current. For each resistor, it is presented:

1. Its nominal value, obtained by reading their color bands.
2. Its value measured by a 6½-digit digital multimeter, model HP34401A[8] .

---

[6] http://data.energizer.com/PDFs/carbonzinc_appman.pdf
[7] http://www.gnu.org/software/octave/
[8] http://www.home.agilent.com/agilent/product.jspx?cc=\\BR\&lc=por\&ckey=1000001295:epsg:pro\&nid=-33228.536880933.00\&id=1000001295:epsg:pro\&cmpid=\\90131

3. The supply voltage V_s directly measured by the acquisition board (Section 2.2).
4. The current measured by the multimeter HP34401A.
5. The current obtained by the proposed system. More precisely, the voltage generated by ZXCT1022 (Section 2.1) is measured by acquisition board and then the equivalent current value is calculated in the PC (Section 2.3).
6. The percentage of relative error, using the formula: error = $\left|\frac{I_0 - I}{I_0}\right| \cdot 100$. Where I_0 is the "exact" value of current, obtained by the measurement via multimeter and I is the value measured by our system.

Table 1. Data used to calculate the percentage of relative errors

Resistor		**V_S**	**I_n**	**I_n**	**I_n**
Nominal (Ω)	**Measured (Ω)**	**Proposed System (V)**	**Measured (mA)**	**Calculated (mA)**	**Relative Error (%)**
105	114.74	2.4742558	26.945	24.74558	8.17
536	542.67	0.5959466	5.5711	5.959466	6.97
1.00 k	1.0060 k	0.3141877	3.0345	3.141877	3.54
5.11 k	5.1275 k	0.0625797	0.5967	0.625797	4.88
10.5 k	10.497 k	0.0308690	0.2913	0.308690	5.97
51.1 k	51.071 k	0.0064104	0.0599	0.064104	7.02

Analyzing the column with the errors, it can be seen that the correct choice of the resistor can reduce the measured error. With low values of resistors, the resistor placed in series – R_S – has a great influence on the voltage drop as a whole, directly influencing the value of the measurement. Moreover, with high values of resistors, the current in the system decreases considerably, leading to the measuring system to its limits.

5 Related Works

In [10], it was proposed the Sensor Node Management Device (SNMD), a complete energy monitoring system for WSNs. It was designed a device management sensor node that is attached to MicaZ and PC via a USB port, allowing monitoring of their energy consumption as well as the other features. Because all devices of WSN can be easily monitored, considering that it has a device for each node, it is possible to mount a bench test that monitors each node individually, and so the network. Because it is a more complete, complex and expensive, it also allows each node individually to simulate the discharge of batteries. One limitation of this system is that it needs a special interface for connecting the nodes: only nodes with this featuring could be used in the test, as is the case of MicaZ.

In work [11], it is proposed a system to check energy consumption in WSN. It is similar to our proposal but with two key differences: *i)* the measuring instrument is an oscilloscope that must be connected to a PC, which is not always usual in laboratories with older equipment and normally has a high cost; *ii)* the sensing resistor used is 10Ω, which is 10 times higher than used in our work and presents a voltage drop proportionally higher, impairing the accuracy of the measurement. Moreover, in that study the applications used in the tests was very specific. In our case, we measured the energy consumption of node in all operational modes without concern for any particular application.

In [12], the measurement system was not the main concern but the measure of energy consumption of MicaZ nodes in some operational modes of its processor and radio transceiver. With this goal, they used a digital oscilloscope including a DC/DC regulator. In contrast with this work, in [12] were not set battery discharge curves.

The main goal in [13] was to create a generic model for battery lifetime analysis, targeting any portable electronic devices. A series of analysis was performed, and the results were compared with a low-level simulator DUALFOIL and with measurements on a real battery. The obtained results showed that an analytical battery model allows a designer to predict the battery lifetime for a given load with a trade-off between the accuracy and the amount of computation performed.

6 Conclusions

In this paper it is proposed a system for energy consumption measurement of wireless sensor network devices. The system has low-cost and it is simple and easy to carry out measurements. It supports measurement of the energy consumption of WSN nodes in an autonomic way, allowing experiments of long duration.

It was presented a case study where a node MicaZ subjected to various operational modes in radio transceiver had its voltage and current monitored. Analyzing the results, some important conclusions about system´s operation was drawn.

Another significant result was the real discharge curve of a battery pack powering a MicaZ node. This result was important to give support to a better understanding of the energy consumption of in WSN nodes and also to construct polynomial curves. These curves could contribute to improving the estimated lifetime of nodes in simulators and also in improvements of energy aware algorithms.

As part of future work, we intend to study the curves of full discharge in other typically used batteries for wireless sensor nodes, such as those with compositions like: Mercury, Alkaline, NiCd (Nickel Cadmium), NiMH (nickel metal hydride) and LiIon (lithium-ion). From these curves can intend to generate approximation functions to be used in networks simulators such as OMNeT++[9] and NS-3[10].

[9] http://www.omnetpp.org/
[10] http://www.nsnam.org/

Acknowledgements. The authors acknowledge the support received from national funding agencies CAPES (International cooperation project CAPES/FCT) and FCT (Project DHT-MESH) and also from their institutions (IFSC, UFSC and FEUP).

References

1. Yoneki, E., Bacon, J.: A survey of Wireless Sensor Network technologies: research trends and middleware's role, Cambridge (2005)
2. Sharma, A.B., Golubchik, L., Govindan, R.: Sensor Faults: Detection Methods and Prevalence in Real-World Datasets. ACM Transactions on Sensor Networks 6(3), 1–39 (2010)
3. Gungor, V., Hancke, G.: Industrial wireless sensor networks: challenges, design principles, and technical approaches. IEEE Transactions on Industrial 56(10), 4258–4265 (2009)
4. Zurawski, R.: Embedded Systems Handbook: Networked Embedded Systems, 2nd edn. Taylor & Francis, San Francisco (2009)
5. IEEE, IEEE 802.15.4 - IEEE Standard for Local and metropolitan area networks - Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs), New York, vol. 2011, p. 314 (September 2011)
6. Crossbow, MicaZ Sensor Node, `http://www.xbow.com/`, `http://www.datasheetarchive.com/crossbow%20micaz-datasheet.html` (accessed in June 12, 2012)
7. Semiconductors, Z.: ZXCT1022 DataSheet, `http://www.diodes.com/products/catalog/detail.php?item-id=1670&popup=datasheet` (accessed in June 12, 2012)
8. Karl, H., Willig, A.: Protocols and architectures for wireless sensor networks, p. 497. John Wiley & Sons, Ltd. (2007)
9. Pinto, A.R., Montez, C.: Decentralized Power Aware Approach for Data Fusion in IEEE 802.15.4 Wireless Sensor Networks. In: 2012 Second Brazilian Conference on Critical Embedded Systems, pp. 64–69 (May 2012)
10. Hergenroder, A., Wilke, J., Meier, D.: Distributed Energy Measurements in WSN Testbeds with a Sensor Node Management Device (SNMD). In: 2010 23rd International Conference on Architecture of Computing Systems (ARCS), pp. 1–7 (2010)
11. Chang, C.-C., Nagel, D.J., Muftic, S.: Assessment of Energy Consumption in Wireless Sensor Networks: A Case Study for Security Algorithms. In: 2007 IEEE International Conference on Mobile Adhoc and Sensor Systems, pp. 1–6 (October 2007)
12. Krämer, M., Geraldy, A.: Energy measurements for micaz node. GI/ITG KuVS Fachgesprch Drahtlose Sensornetze, 61–68 (2006)
13. Rakhmatov, D., Vrudhula, S., Wallach, D.A.: A model for battery lifetime analysis for organizing applications on a pocket computer. IEEE Transactions on Very Large Scale Integration (VLSI) Systems 11(6), 1019–1030 (2003)

Enhancing PTN to Improve the QoS Provided by the IP Mobility Management

David Cortés-Polo[1], José-Luis González-Sánchez[1], Javier Carmona-Murillo[2], Fco. Javier Rodríguez-Pérez[2], and Javier Corral-García[1]

[1] Research, Technological Innovation and Supercomputing Center of Extremadura (CénitS), Cáceres, Spain
{david.cortes,joseluis.gonzalez,javier.corral}@cenits.es
[2] Department of Computing and Telematics System Engineering. GITACA Research Group, University of Extremadura, Cáceres, Spain
{jcarmur,fjrodri}@unex.es

Abstract. The integration of Mobile and Fixed Network is one of the most important topics in the Future Internet development and an open issue for network providers. In this case, mobility management and Quality of Service (QoS) are two of the main goals in this integration. These tasks are an important challenge if the mobility rate is very high. In this paper, we present a new architecture to provide both of them, QoS and high rate mobility using an IP-based mobility protocol and an MPLS-based protocol to deploy Packet Transport Networks (PTN). To achieve this, we had developed analytical models to evaluate registration updates cost, packet loss rate and buffer requirements of the architecture. Numerical results show that the proposed mechanisms can significantly reduce registration updates cost and provide low handoff latency and packet loss rate under various scenarios.

Keywords: Mobile IPv6, MPLS, MPLS-TP, Proxy Mobile IPv6, PTN, QoS, Registration Update Cost.

1 Introduction

Fourth-Generation (4G) wireless communication systems aims to integrate heterogeneous networks seamlessly, in order to satisfy the increasing demand of users in terms of QoS and bandwidth requirements [1].

In this sense, the integration of heterogeneous networks is solved using mobility management IP-based protocols. For wireless communication, there are two approaches; Mobile IP (MIP) [2] and Proxy Mobile IP (PMIP) [3]. Both of them are IP-based protocols and allow the mobility of the Mobile Node (MN), keeping the connectivity to the Internet while moving.

MIP was the first protocol developed to introduce the mobility on the MN. This protocol has some disadvantages such as, the MN has to implement MIP on its protocol stack to keep the connectivity to the Internet while moving, as well as, the respective consume of resources to do that.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 911–921.
DOI: 10.1007/978-3-642-36981-0_85

PMIP was developed to avoid those disadvantages. Some entities are included at the mobile network to resolve it. Those entities will track the movements of the MN and initiate the mobility signalling in order to set up the required routing state. This reduces the signalling of the MN and the hardware resources reserved to the protocol stack. However those protocols present some drawbacks, such as the long handoff latency or the large signalling load due to frequent registration updates.

Besides, 4G networks try to resolve the increasing need of the users to get better services with QoS and a steady bandwidth. This could be obtained using resource provisioning in the access network to provide certain QoS parameters to the communications.

There are three different architectures to provide network resources for QoS guarantees in the Internet: Integrated Services (Intserv) [4], Differentiated Services (Diffserv) [5] and Multiprotocol Label Switching (MPLS) [6]. The notable benefits of MPLS in terms of QoS, traffic engineering and support of advanced IP services, inspire some works on the use of this technology in the wireless infrastructure [7].

Over the past several years, fixed network carriers has moved from IP/MPLS approaches towards Multi-Protocol Label Switching – Transport Profile (MPLS-TP) as the protocol to converge the traditional fixed networks to the packet-based transport networks. MPLS-TP includes the resource provisioning capability and resilience offered by MPLS and adds the Operations, Administration, and Maintenance (OAM) layer [8] and protection capabilities offered by MPLS-TP to achieve a higher efficiency and lower operational cost.

To meet the requirements of next generation mobile networks, in this paper, a new approach called Integrated Proxy Mobile MPLS-TP is proposed. This approach improves the limitations of PMIP/MIP and combines the benefits of MPLS applied to the PTNs with MPLS-TP.

In order to gauge the effectiveness of the proposed approach, analytical expressions of the signalling cost, packet loss rate and buffer requirements are derived. Numerical results show that our proposal can significantly reduce the registration updates cost and provide low packet loss rate when compared to the existing schemes (PMIP[3], PMIP-MPLS [9]).

The remainder of this paper is organized as follows. Section II gives the related works of the mobility networks in conjunction with MPLS. Section III introduces our proposed architecture. In section IV, analytical models are developed to derive the signalling cost function of registration updates, packet loss rate and the buffer size required for all underlying protocols. Numerical results are given in section V. Finally; section VI contains our concluding remarks.

2 Related Works

Mobile IP is probably the most widely known mobility management proposal at the moment. One of the most important research trends is the micro-mobile extension

applied to the Mobile IP and the tunnelling mechanism using MPLS to deliver the packets through the visited network guaranteeing QoS requirements [10-13].

The tunnelling mechanism carries the traffic among mobility entities such as Home Agent (HA), Foreign Agent (FA), Mobile Node (MN) and other proxy nodes. As usual, IP-in-IP tunnelling is used as tunnelling and encapsulation mechanism. Mobile IP approaches need to modify the protocol included at the Mobile Node in order to add extra functions or new algorithms to the MN.

Proxy Mobile IPv6 provides several advantages compared to MIPv6. One of the main improvements is the network-based handover provided by the protocol. This provides local mobility management to a mobile node without requiring any modification.

Moreover, in both protocols, the handover latency and the packet loss rate are very high when a handover is produced.

Other work [14] has studied the convenience of MPLS in the access network to improve the QoS offered by the access network.

Xia and Sarikaya [9] analyse the impact produced in Proxy Mobile IPv6 the creation of the tunnel using MPLS. With this proposal, the Local Mobility Anchor (LMA) and Mobile Access Gateway (MAG) are interconnected with an LSP. This approach reduces the handover latency and optimizes it. With this model, all messages used are extensions of existing messages already defined in PMIPv6. On the other hand, QoS could be included on the backhaul of the network, but the packet loss and the handover latency is still very high with this approach.

In this paper, a new mobility management scheme for PMIPv6 is proposed, improving the actual PTN networks to integrate the Mobile and the Fixed Network in a unique architecture. The scheme combines the benefits of the network based mobility management and the MPLS-TP capabilities. Our approach reduces the cost of the registration updates of the original protocol, and includes QoS on the communication using MPLS-TP forwarding path to create the tunnels in the access network.

3 Proposed Approach: IPM-TP

In this section, our new approach called Integrated Proxy Mobile MPLS-TP (IPM-TP) is presented. The architecture is depictured in Figure 1.

We assume that an MPLS-TP access network exists between the Ingress Label Edge Router Gateway/Local Mobility Anchor (Ingress LLMA) and the Egress Label Edge Router/Mobility Anchor Gateway (Egress LMAG).

The network architecture is based on a two-level hierarchy. At the higher level is the Ingress LLMA that performs the role of an edge LER, filtering between intra- and inter-domain signalling.

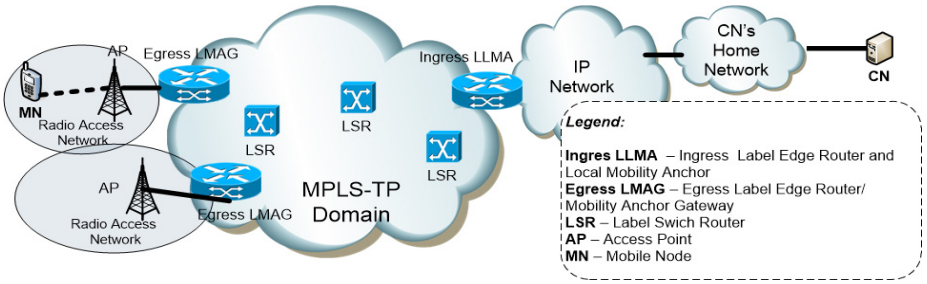


Fig. 1. Architecture of an IPM-TP wireless access network

At the same time, this network element has the functionality of a Local Mobility Anchor (LMA) and the second level is the LMAG connected to several access points (APs) that offer link-layer connectivity. We distinguish here between link-layer functionalities of the air interface, which are handled by the AP, and IP-layer mobility (L3 handoff), which occurs when the MN moves between subnets served by different LMAGs. Note that an LMAG is the first IP-capable network element seen from the MN.

Our approach is based on the MPLS-TP tunnelling mechanism instead of the IP or GRE tunnelling proposed by the standard. The MPLS-TP tunnel can be extended creating a set of forwarding LSP-tunnel between the previous LMAG (pLMAG) and the new LMAG (nLMAG). Each time that the Mobile Node moves to a new subnet, it sends the Mobile Node ID (MN ID) and the new LMAG ID (nLMAG ID). Both of them will be registered at the pLMAG instead of the LLMA. By this procedure, the existing LSP-tunnel between the LLMA and the pLMAG could be extended to new LMAGs if it is necessary. Packets travelling towards this Mobile Node will be sent by the LLMA across the path, taking advantage of the existing LSP between the LLMA and the first LMAG, and then forwarded along the path of LMAGs to the Mobile Node. It is easy to see that this approach may cause unacceptable delays due to long paths.

To avoid an excessive extension of the path, the MPLS-TP routers measure the most important QoS parameters in mobile communications. Those parameters are bandwidth, packet loss, delay and jitter (i.e. the bandwidth will be measured by the link-state protocol and the packet loss and tunnel delay will be obtained from the connection monitoring entity located in the OAM plane [15]).

With the data obtained of the MPLS-TP routers, the new LMAG can decide when the QoS is below a fixed threshold for each flow (using RSVP resource reservation signalling). If this situation is detected, the new LMAG must send a Proxy Binding Update to the LLMA using the common procedure described in PMIP standard. With this message, a new LSP-tunnel will be created from the LLMA to the new LMAG where the Mobile Node will be connected. To avoid packet loss when the handover is produced, a buffering technique is implemented in the LMAG. This avoids the packet delivery out of order too. Figure 2 shows the operation mode of the IPM-TP protocol.

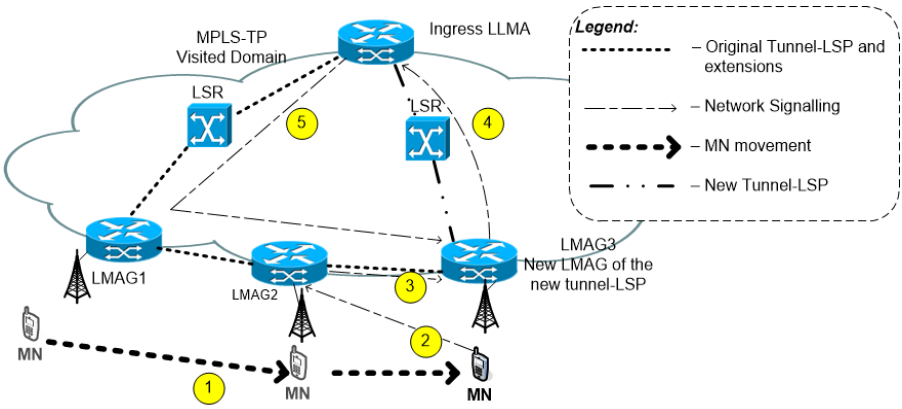


Fig. 2. Mechanism to maintain the QoS required by the communication

In this example, the Mobile Node already has been moved from subnet 1 to subnet 2 (step 1). All the movements of the Mobile Node are detected by the link-layer (L2 Trigger). In the example, the first movement of the Mobile Node is to the subnet 2; it reports the tuple [MN ID, new LMAG2 ID] to the previous LMAG.

Thus, data packets destined to the Mobile Node are sent from the ingress LLMA to the LMAG1 using the existing LSP. When the LMAG1 receives the handover notification, it will send a Handover Initiate message to the LMAG announced by the Mobile Node. With the message, the previous LMAG searches the path which satisfies all the QoS constraints using QoS management Algorithm (i.e. Constrained Shorted Path First protocol, CSPF). A full duplex LSP-tunnel between the LMAG1 and LMAG2 is created, and the packets will be forwarded using this LSP-tunnel and they will be buffered until the Mobile Node is attached.

When the Mobile Node moves to the subnet 3, as seen in the last paragraph, Mobile Node sends the tuple [MN ID, new LMAG ID] to the previous LMAG that is the LMAG2 (step 2).

This LMAG searches the best LSP-tunnel to extend the path. It chooses the LSP-tunnel that offers the best QoS parameters to use (step 3).

LMAG3 measures the QoS parameters of the set of forwarding paths. If the threshold negotiated for the flow is reached, the LMAG3 sends a Proxy Binding Update message to the LLMA (step 4). In this message it indicates that the MN has moved and the LLMA starts to create a new LSP-tunnel from LLMA to the LMAG where the Mobile Node will be attached. This procedure is described in the PMIP standard [3].

When the new path is created the LLMA send a PATH Tear message to the previous LSP-tunnel to free all the reserved resources. This will be the last packet sent by the previous path from the LLMA to the LMAG3 (step 5). This method closes the forwarding path.

4 Performance Evaluations and Analysis

In this section, analytical models are developed to derive the cost function of registration updates, packet loss and buffer size of the different approaches. We compare our proposal with respect to the PMIP [3], PMIP-MPLS [9] schemes. Below the parameters used in the analysis are introduced.

Parameters

- t_s Average connection time for a session;
- t_r Average stay time at a visited network;
- T_{ad} Time interval between Agent Advertisements messages;
- N_h Average number of level 3 handover in a session ($N_h = t_s/t_r$);
- N_f Average number of forwarding path during a session ($N_f = N_h/length\_forwarding\_path$);
- s_u Average size of a signalling message for record update;
- s_l Average size of a message for LSP establishment;
- $h_{x\text{-}y}$ Average number of hops between x and y in the wired network;
- B_w Bandwidth of the wired link;
- B_{wl} Bandwidth of the wireless link;
- L_w Latency of the wired link (propagation delay);
- L_{wl} Latency of the wireless link (propagation delay);
- λ_d Transmission ratio for a downlink packet;
- T_{inter} Time between arrivals of consecutive data packets.
- P_t Routing or label table lookup and processing delay;

Let *t (s,hx-y)*, the time spent for a packet with size *s* to be forwarded from *x* towards *y* across wired and wireless links. *t(s,hx-y)* can be expressed in the form:

$$t(s,h_{x-y}) = c + h_{x-y} \cdot \left(\frac{s}{B_w} + L_w \right) + (h_{x-y} + 1) \cdot P_t \tag{1}$$

$$where \quad c = \begin{cases} \dfrac{s}{B_{wl}} + L_{wl} & if \quad x = MN \\ 0 & if \quad x \neq MN \end{cases}$$

4.1 Signalling Cost

Total signalling cost of registration updates during a session is denoted by C_u. The signalling cost is the accumulative traffic load on exchanging signalling messages (*hop x message size*) during the communication session of the MN. For each movement into a new subnet, the Proxy Binding Update message with the LMA is performed in both PMIP and PMIP-MPLS. In IPM-TP, a forwarding path is created

between the previous LMAG and the new LMAG to forward the packets to the mobile node after that a Proxy Binding Update is sent to the LLMA. This forwarding path is extended while the QoS requirements are satisfied. The registration with the LLMA will be done when the threshold requirements is reached (i.e. the bandwidth or the delay are not satisfied). The expression of registration updates cost for all underlying protocols can be summarized as follows:

$$C_u(PMIP) = 2 \cdot s_u \cdot h_{MAG-LMA} \cdot N_h + 2 \cdot s_u \cdot h_{nMAG-LMA} \cdot N_h \tag{2}$$

$$C_u(PMIP-MPLS) = 2 \cdot s_u \cdot h_{pMAG-LMA} \cdot N_h + 2 \cdot s_l \cdot h_{nMAG-LMA} \cdot N_h \tag{3}$$

$$C_u(IPM-TP) = 2 \cdot s_u \cdot h_{MN-pMAG} \cdot N_h + 2 \cdot s_l \cdot h_{pMAG-nMAG} \cdot N_h + 2 \cdot N_f \cdot s_l \cdot \cdot h_{pMAG-LMA} + 2 \cdot s_l \cdot h_{LMA-nMAG} \cdot N_f \tag{4}$$

4.2 Total Packet Loss during a Session

The total packet loss P_{loss} during a session is defined as the sum of lost packets during all handoffs while the MN is receiving the downlink data packets.

In PMIP and PMIP-MPLS, all in-flight packets will be lost during the handoff time due to the lack of any buffering mechanism. In IPM-TP, in-flight packets are stored using the buffering mechanism. As mentioned in previous section, the L2 trigger is used in IPM-TP. It is assumed that the packet loss begins when the L2 handoff is detected.

$$P_{loss}(PMIP) = \left[\left(\frac{1}{2}T_{ad}\right) + T_c(PMIP)\right] \cdot \lambda_d \cdot N_h \tag{5}$$

$$P_{loss}(PMIP-MPLS) = \left[\left(\frac{1}{2}T_{ad}\right) + T_c(PMIP-MPLS)\right] \cdot \lambda_d \cdot N_h \tag{6}$$

$$P_{loss}(IPM-TP) = t(s_u, h_{MN-pMAG}) \cdot \lambda_d \cdot N_h \tag{7}$$

4.3 Buffer Size Requirement

The buffers used to store in-flight packets are located at the LMAG in IPM-TP. The buffering mechanism is activated when the Handover Initiate message is received. The packets will be forwarded to the MN when a Handover ACK message is received.

The buffer size requirement for IPM-TP is the maximum size to storage all in-flight packets while the MN does not have any connectivity, is derived as follows:

$$B_{size}(IPM-TP) = \left(\frac{1}{2}T_{ad}\right) + (t(s_u, h_{MN-pMAG}) + t(s_l, h_{pMAG-nMAG})) \cdot \lambda_d \tag{8}$$

5 Results

In this section, the protocols presented above are compared using an analytical approach. The following parameters are used in the analysis [16]: t_s=1000 sec, t_r=[5-50] sec (default 20), T_{ad}=1 sec, s_u=48 Bytes, s_l=28 Bytes, B_w=100 Mbps, B_{wl}=11 Mbps, L_w=1 msec, L_{wl}=2 msec, λ_d=64 Kbps.

Figure 3 shows an example where we can observe the path that follows a packet involved in the recovery mechanism. There are 8 hops between LMA and the pMAG and the distance between the LMA and the nMAG is 9 hops. There is one hop between pMAG and nMAG.

Figure 4 shows the comparison of registration costs. PMIP and PMIP-MPLS have a high registration cost due to the signalling sent to the LMA to update the binding cache. This can be observed when the cell resident time is between 5 - 15 seconds. We can consider that the mobility of the MN is high and the registrations to the LMA are frequent. With the IPM-TP approach the registration cost is reduced around 75% of the extra signalling. This value decrease around 36% of extra signaling when the cell resident time is higher.

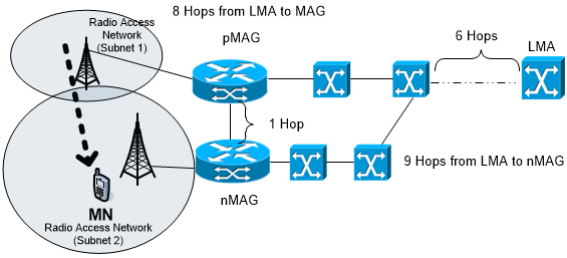


Fig. 3. Relative distances in hops in the simulated network

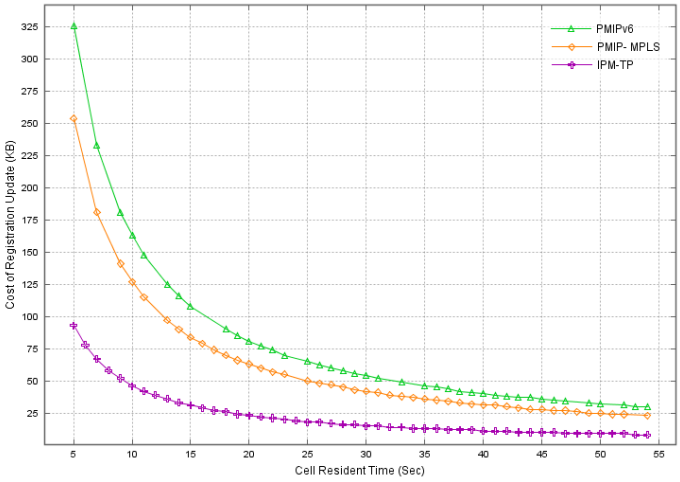


Fig. 4. Registration Updates Cost Comparison

Figure 5 shows the amount of packets loss during the whole connection session for different approaches. PMIP and PMIP-MPLS do not implement the buffering mechanism and they have the largest amount of lost packets. This is because all the packets sent to the MN from the LMA while the handover is produced will be lost. If the mobility of the MN is high, the numbers of handovers are increased and these handovers increment the packet loss as we can observe in the figure.

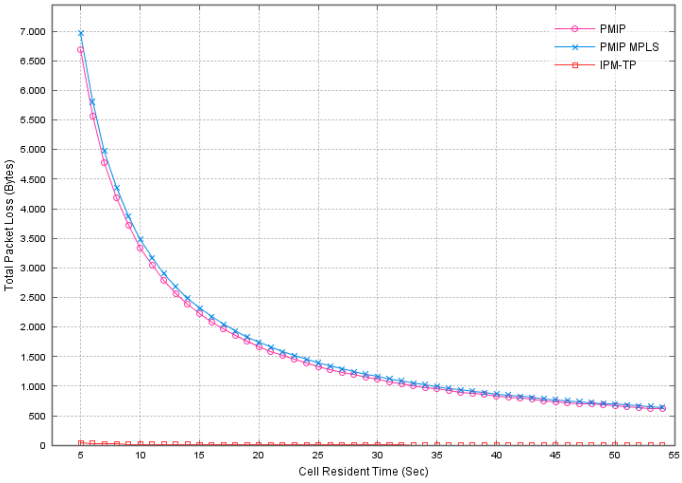


Fig. 5. Total Packet Loss Comparison during a Session

However, the IPM-TP approach, which initiates the buffering mechanism when the handover is detected, stores the packets and avoids the packet loss as shown in the figure.

The buffer size requirement for each MN using IPM-TP is about 2.7 KB. This is derived when the Cell Resident Time is minimal. This means that over 40 thousands of MNs can be provided with the buffering mechanism if a node is equipped with a memory of size 100 MB.

6 Conclusions

In this paper, we proposed an enhancement to the PTN access network to improve the QoS offered to the MN, called Integrated Proxy Mobile MPLS-TP that supports both mobility and Quality of Service management. In addition, this scheme is appropriate for Mobile Nodes with high mobility rate, where data packets must be forwarded quickly to their new locations.

We have described the basic operation to create the LSP before the MN moves and attaches into a new subnet in order to reduce service disruption by using the L2 functionalities. We have proposed to extend the LSP-tunnel created in the MPLS-TP

domain to track the MN movement. With this extension we reduce the handoff signalling messages; the registration updates cost and provides low handoff latency and small packet loss rate.

To achieve this, a comparison between our proposals and existing solutions (PMIP and PMIP-MPLS) is given. We analytically derived the registration updates cost, the packet loss during the handoff and the size of the buffer implemented in the nodes.

We also have described the architecture to implement this approach. We proved, through analysis, that our proposed mechanisms achieve a substantial signalling cost gain and improve the QoS offered by the network at the price of a slight increase of the link usage cost.

Acknowledgments. This work is financed in part with FSE Founds through the Research, Technological Innovation and Supercomputing Center of Extremadura (CénitS).

References

1. Frattasi, S., Fathi, H., Fitzek, F., Prasad, R., Katz, M.: Defining 4G technology from the users perspective. IEEE Network (2006)
2. Perkins, C.: IP Mobility Support for IPv4, RFC 3220 (January 2002)
3. Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K., Patil, B.: Proxy Mobile IPv6, RFC 5213 (August 2008)
4. Braden, R., Clark, D., Shenker, S.: Integrated Services in the Internet Architecture: an Overview. Internet Engineering Task Force, RFC 1633 (June 1994)
5. Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., Weiss, W.: An Architecture for Differentiated Services. Internet Engineering Task Force, RFC 2475 (December 1998)
6. Rosen, E., Viswanathan, A., Callon, R.: Multiprotocol Label Switching Architecture. Internet IETF RFC 3031 (January 2001)
7. Francesco, P.: An MPLS-based architecture for scalable QoS and traffic engineering in converged multiservice mobile IP networks. Comput. Netw. 47, 257–269 (2005)
8. Bellagamba, E., Andersson, L., Skoldstrom, P., Ward, D., Takacs, A.: Configuration of Pro-Active Operations, Administration, and Maintenance (OAM) Functions for MPLS-based Transport Networks using RSVP-TE, Draft IETF (April 2012)
9. Xia, F., Sarikaya, B.: MPLS Tunnel Support for Proxy Mobile IPv6. IETF Draft (work in progress) (October 2008)
10. Wenchao, M., Yuguang, F.: Dynamic hierarchical mobility management strategy for Mobile IP networks. IEEE Journal on Selected Areas in Communications 22, 664–676 (2004)
11. Hairong, Z., Chihsiang, Y., Mouftah, H.T.: DHMM: A QoS Capable Micro-Mobility Management Protocol for Next Generation All-IP Wireless Networks. In: Global Telecommunications Conference, GLOBECOM 2007, pp. 4989–4993. IEEE (2007)
12. Chiussi, F.M., Khotimsky, D.A., Krishnan, S.: Mobility management in third-generation all-IP networks. IEEE Communications Magazine 40, 124–135 (2002)

13. Langar, R., Bouabdallah, N., Boutaba, R.: A Comprehensive Analysis of Mobility Management in MPLS-Based Wireless Access Networks. Transactions on Networking 16(4) (August 2008)
14. Garroppo, R.G., Giordano, S., Tavanti, L.: Network-Based Micro-Mobility in Wireless Mesh Networks: Is MPLS Convenient? In: Global Telecommunications Conference, GLOBECOM 2009, November 30-December 4, pp. 1–5. IEEE (2009)
15. Frost, D., Bryant, S.: Packet Loss and Delay Measurement for MPLS Network. IETF, RFC 6374 (September 2011)
16. Lo, S.-C., Lee, G., Chen, W.-T.: Architecture for Mobility and QoS Support in All-IP Wireless Networks. IEEE Journal on Selected Areas in Communications 22(4) (May 2004)

User's Requirements in Internet Access Sharing

Conceição Tavares and Henrique Santos

Centro Algoritmi, University of Minho, Portugal
cbt@ua.pt, hsantos@dsi.uminho.pt

Abstract. Nowadays consumers dispose of Internet connections with increasing capacity in such a way that those connections are frequently underused. Sharing connections seems like an obvious attractive response, and it has being explored through several initiatives, e.g. OpenSpark, FON, with relative success although inherent risks. The related studies made so far concerning security, incentives, legality, etc., did not allow inferring a global model upon what some analysis and evaluation of this technology could be done. This is partially because there are several entities playing crucial roles in the realization of this model, e.g. users, service suppliers and authorities and aspects from several dimensions, e.g. technical, regulatory and business that must be considered. Aiming to design such a model, it was carried out an inquiry to potential users, in both perspectives of using and providing a shared Internet access. The preliminary study results are presented here.

Keywords: Broadband connection sharing, Internet access, security.

1 Introduction

Within our information society it is quite common to have a Wi-Fi network to support the domestic network and have that local network interconnected with the Internet. Actual residential Internet access subscriptions (e.g. xDSL, optical fibber) are already being shared among a restrict group of people that the owner of the Internet subscription trusts, e.g. family and friends visiting over. The Wi-Fi network also has the consequence of expanding the residential network outside the physical limits of the domestic property. Given this circumstance, it is possible to expand the access sharing of domestic networks to encompass any candidate person in the neighborhood (hereafter called a Visitor). Visitors could use domestic Wi-Fi networks deployed by others (hereafter called Micro-Providers, or MPs for short) to get temporary access to the Internet. The existing number of such configurations (Wi-Fi network with broadband Internet access) makes the idea attractive and the public usage of private residential subscriptions had already being identified as a business opportunity [1]. Several approaches to Internet access sharing, either studies or deployed initiatives, envisage the potential of solutions on that basis.

Although, generalizing the access sharing is very challenging. It creates new business opportunities, both for access subscribers (MPs can make profit from it, with direct revenue or obtaining any other kind of reward) and Internet service providers

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 923–929.
DOI: 10.1007/978-3-642-36981-0_86

(ISPs), but these may clash with established legal frameworks. Also, extending network sharing to Visitors raises several critical security issues, mainly because of the lack of trust relationships between MPs and Visitors. Bina and Giaglis pointed out some of the related research issues [2].

1.1 Security Issues in Internet Access Sharing

Currently, the Internet subscriber implicitly trusts people with whom he/she explicitly shares the subscription. The subscriber trusts that such people will not do any illegal activity, will not misuse the domestic infrastructure, i.e. will not attempt to access resources to which they are not authorized to, will not run denial of service initiatives, etc. Finally, they (subscriber and known users) all trust each other in terms of privacy, and assume that others do not inspect or trace their network activities and, of course, identity disclosure is not supposed to be an issue.

Once a domestic network is opened to a Visitor, all these trust assumptions no longer holds. As there is no trust relationship between the MP and the Visitor, MP would not risk being responsible for any abuse activities conducted by a Visitor, mainly because of ethical or legal implications it may have. On the other hand, a Visitor may not trust on the honesty and correctness of an MP, assuming he/she can eavesdrop or tamper communications. Finally, in terms of privacy, a Visitor should be as anonymous as possible to a MP otherwise colluded MPs could trace the location and profile a Visitor.

1.2 Legal and Regulatory Framework

Policies, rules, responsibilities and obligations definition are crucial for security. Internet access sharing is not specifically foreseen by the law and therefore lacks from this basis. In order to put connection sharing in practice, the surveyed initiatives described below circumvent legal issues or impose usage restrictions. There is not actually a legal framework that establishes the conditions and rules for the sharing.

The core of the EU regulatory frameworks related to Internet access came up in the year 2002 with the Telecommunications Package. This package comprehends the Framework Directive 2002/21/EC[1], Authorisation Directive 2002/20/EC[2], Access Directive 2002/19/EC[3], Universal Service Directive 2002/22/EC[4], Privacy and Electronic Communications Directive 2002/58/EC[5], Directive 98/34/EC[6] and e-Commerce Directive 2000/31/EC[7]. Still, many issues related to Internet access continue being subject of discussion. An example of an open issue is the usage of IP address to identify users. It is not guaranteed that IP will be accepted by the law to

[1] OJ L 108, 24.4.2002, p. 33–50.
[2] OJ L 108, 24.4.2002, p. 7–20.
[3] OJ L 108, 24.4.2002, p. 21–32.
[4] OJ L 108, 24.4.2002, p. 51–77.
[5] OJ L 201, 31.7.2002, p. 37–47.
[6] OJ L 204, 21.7.1998, p. 37.
[7] OJ L 178, 17.2.2000, p. 1–16.

user identification, e.g. in Portugal [3] and in EUA [4] and neither is clearly listed and characterized what are considered illegal activities [3]. Other missing aspect is the possibility of the MP to profit from the sharing.

It is important a carefully study of legal framework not only to assure the lawfulness of the process but also at the time of laying down stakeholders' requirements. For instance, accountability may be required by users only if they are afraid of being charged for illegal activities.

1.3 New Models – Current Business Initiatives

There are some initiatives that promote domestic Wi-Fi sharing, e.g., FON[8], NetShare[9], SofaNet[10]and OpenSpark[11], in which Wi-Fi access points and Internet access are owned by individuals. These initiatives differ significantly in terms of business models, in terms of security and in the service they provide.

OpenSpark's and FON's initial objective was to provide Internet access through others' connections by simply share own Internet access, but currently, in both cases, users can also profit from the sharing of their own connections. Netshare allows subscribers to legally profit from their sharing. The Sofanet is a service provided by T-Com operator and thus restricted to Germany. For every account acquired through the shared connection it is offered a certain volume of free traffic valid for a certain period of time. The Sofanet service is supported by a special architecture: the MP has two separated DSL links, one for personal usage and other intended to be shared by Visitors through an ordinary AP. Data exchange takes place via established PPTP (Point-to-Point Tunneling Protocol) tunnels between each Visitor and the ISP. The solution also incorporates mechanisms to prevent bandwidth exhaustion.

OpenSpark's MPs sign a Terms and Conditions (T&C) agreement for transferring the responsibility of misuses of the broadband connection to the Visitors and implements confidentiality based on encrypted tunnels. It also implements authentication processes that contribute to lower the risk of impersonation attacks.

Part of FON's strategy is based on partnerships with other enterprises, e.g., Skype, Google, Index Ventures and with operators and Internet Service Providers (BT, Neuf, Comstar, Zon, Time Warner Cable, T-Com). These FON associations aim to achieve a faster wide spreading of the FON community and bypass the unwillingness of Internet Service Providers (ISPs) in allowing subscription sharing. Confidentiality and authentication are supported based on a specific AP configured with two SSIDs and through a Captive Portal. Some FON services are unavailable in some countries due to legislation (and not technical) constrains, e.g. profit from the sharing. FON claims that frees the MP to become liable of any illicit activity through shared connection but it is not clear in what extent disputes over it would be FON competent to solve.

Concerning Netshare, Speakeasy – the operator – remarks clearly that "Administrators are also responsible for the security and integrity of their shared network".

---

[8] FON, `http://www.fon.com/` (2011)

[9] Speakeasy. WiFi Netshare service, `http://www.speakeasy.net/netshare/` (2011)

[10] Sofa networks, `http://www.sofanet.de/` (2011)

[11] MP-Master Planet. OpenSpark, `http://www.sparknet.fi/` (2011)

2 Users Survey on Internet Sharing

The overview of connection sharing in terms of legal, business and technical aspects reveal that for many issues it is essential to promote further discussion and reflection to find proper and effective solutions. To start, it is desirable to grasp the users perspective and shape this service (also or even mainly) to their expectations increasing the likelihood of a successful solution. To capture user requirements within this project we decide to promote a survey. The survey allowed answering the following key questions:

- What is the overall trend in own Internet access sharing?
- What are the main motivations and quid pro quos for the sharing?
- What are main barriers and discouragement factors to engage in a shared Internet access?

The inquiry was applied between the 5th of September, 2011 and the 28th of December, 2011. We were able to perform statistical analysis on 310 questionnaires using SPSS[12]. The age of respondents vary from 18 to 62 years with the medium being 31 years old. People pooled are predominantly graduated and employees.

The inquiry was structured in three major sections: individual characterization (personal and related to Internet usage habits); assessment of user's perspective about using connection sharing and; assessment of user's perspective about sharing their own connection.

3 Results Presentation and Analysis

The study showed that respondents frequently use Internet from several places, e.g. their homes, school, workplace and even public places, to perform a wide range of important activities for the day life, e.g. find information, buy online, home banking, mail, instant messaging, social networking, telecommuting, etc., which in principle justifies improve Internet access in the broadest sense. Still, most of the activity is performed from home. This could be because there is no need or because there are no conditions for doing it in other locations.

3.1 Major Concerns in Connection Sharing

To assess the most important factors that could hinder people from using Internet shared access, respondents were asked to express their opinion using a three-level scale, "Irrelevant", "Relevant" or "Very relevant"; an option to "Do not have an opinion/Do not understand"; and seven possible options: "View confidential data", "Forging data or connections on your behalf", "Being accounted for other's traffic or actions", "Disclosure of your identity", "Knowing personal data even without identity disclosure (e.g. risk in profiling or tracing in Internet usage)", "Too complicated

[12] IBM Corp.'s statistical software (Statistical Package for Social Sciences).

and/or lengthy procedures", "Not having a trusted entity to manage the model (e.g. a system managed solely by users)". The results showed that people are more concerned with data confidentiality and the accounting of activities than with privacy or eventual technical demanding. People are willing to accept additional difficulties operating the communication system. Table 1 shows the surveyed items ranked according the importance level. The three most valued factors were "View confidential data", "Forging data or connections on your behalf" and "Being accounted for other's traffic or actions" with respectively 79%, 78.4% and 61% of the respondents choosing "Very relevant". The less valuable item was "Too complicated and/or lengthy procedures" having being pointed as "Irrelevant" by 25.2% of the respondents.

Table 1. Factors that could hinder people from use shared Internet access (%)

(Option)	1 (Irrelevant)	2 (Relevant)	3 (Very relevant)	Without opinion/ do not understand
View confidential data	4,5	12,9	79	1,9
Forging data or connections on your behalf	6,5	10	78,4	2,9
Being accounted for other's traffic or actions	11,3	23,2	61	2,6
Disclosure of your identity	17,7	40,3	37,7	2,3
Knowing personal data even without identity disclosure (e.g. risk in profiling or tracing in Internet usage)	15,5	44,5	33,9	4,8
Procedures too complicated and / or lengthy	25,2	41	25,8	6,5
Not having a trusted entity to manage the model (e.g. a system managed solely by users)	13,5	36,1	38,7	9,7

3.2 Willingness towards Sharing Own Internet Access

Approximately half of the respondents have their own subscriptions (50,3%). 57.4 % of the respondents are not willing to share Internet access (whether possessing or no subscriptions of their own). 23.5 % are willing to share their subscription and 12.9 % selected "Other situation (e.g. only if a trusted entity manages the process)".

The inquiry listed ten specific problems to be classified in five-level scale (1: not important; 5: very important) and an option, "Without opinion", for those with no opinion about the subject (Table 2). For obvious reasons this part of the survey was not prompted to the respondents that are not interested on sharing their Internet access. Users seem willing to incur some inconveniences and do some effort to engage in connection sharing as long as security aspects are assured.

Table 2. Barriers on own Internet access sharing (%)

Option	1	2	3	4	5	Without opinion/do not understand
Identity theft	4,4	3,5	4,4	11,5	74,3	0,0
Unauthorized access to equipment on my network	3,5	4,4	11,5	22,1	54,9	1,8
Loss of privacy (generally)	7,1	4,4	13,3	20,4	54,0	0,0
Loss of quality of service. (e.g. bandwidth)	0,9	5,3	12,4	27,4	48,7	1,8
Bandwidth exhaustion	2,7	4,4	18,6	19,5	47,8	4,4
Not knowing who uses the connection	9,7	12,4	16,8	23,9	34,5	1,8
Having to make investments (e.g. purchase of equipment)	4,4	5,3	28,3	24,8	33,6	2,7
Administration of the equipment done by third parties	8,8	5,3	22,1	32,7	23,9	6,2
Not knowing the usage data * (e.g. unable to check their use of my access, including time and bandwidth)	6,2	10,6	15,0	34,5	23,9	8,8
Having to change service provider (ISP)	12,4	14,2	23,0	18,6	22,1	8,8
Other	0	0	0	0,9	0,9	8,8

3.3 Motivations on Sharing Own Internet Access

The inquiry included a question to classify according the importance given possible forms of reward expressed through a five-level scale and a “Without opinion” option.The forms of reward presented were: “Participation in a digital city”, “Internet access through other´s subscription”, “Discount on subscription service”, “Obtaining other services”, “Direct monetary retribution”, “Other”. “Participation in a digital city” was the option which fewer respondents elected as of the maximum importance (26%). The option which has been given more importance was “Discount on subscription service” (54% of the answers). The second most elected option as an important reward was “Internet access through other´s subscription” (Table 3).

It is not straightforward to account for the difference that respondents gave to the importance of “Discount on the service” and to “Direct retribution”. It would be expected similar ones due to the common nature of reward - money saving. It is worthwhile to notice that social nature motivation, “Participation in a digital city”, was appointed less important than other types of reward.

Table 3. Incentive evaluation on own Internet access sharing (%)

Option	1	2	3	4	5	Without opinion/do not understand
Discount on service	0,9	8,0	9,7	22,1	54,0	3,5
Have access to the Internet through the subscriptions of others	0,9	4,4	18,6	27,4	43,4	3,5
Obtaining other services	5,3	6,2	17,7	28,3	33,6	6,2
Direct monetary retribution	8,8	7,1	23,0	20,4	32,7	5,3
Participation in a digital city	14,2	5,3	22,1	20,4	25,7	9,7
Other	0	0	0	0	0	6,2

4 Conclusions

The survey confirmed that security aspects are of paramount importance concerning shared access. Users seem willing to incur some inconveniences and do some effort to engage in connection sharing as long as security aspects are assured.

To design a model for Internet access sharing, not only the identified variables must be considered but also their interplay. Furthermore, it is necessary to consider other issues, namely ISPs' strategy and regulatory authorities. Their interplay, possible contradictory objectives and tradeoffs have to be investigated. The uncertainties about how legal and security issues will be managed in Internet access sharing could have some influence in users' opinion at the time of identifying barriers. Uncertainties on the different relationships among MP, ISPs and Visitors to be established (derived from legal framework) is probably strong correlated to the lack of trust in shared access.

Acknowledgments. This work was funded by FEDER through Programa Operacional Fatores de Competitividade – COMPETE, and by national founds through FCT – Fundação para a Ciência e Tecnologia, under project: FCOMP-01-0124-FEDER-022674.

References

1. Francis, J.C., Nils, E., et al.: Business opportunities of open broadband wireless access networks. In: International Workshop on Broadband Wireless Access for Ubiquitous Networking (2006)
2. Bina, M., Giaglis, G.: Emerging issues in researching community-based WLANS. J. of Computer Information Systems 46(1), 9–16 (2005)
3. http://exameinformatica.sapo.pt/noticias/mercados/2012/09/26/ministerio-publico-diz-que-e-legal-copiar-musicas-e-filmes-na-net
4. http://www.pcworld.com/article/255061/judge_throws_out_mass_john_doe_porn_copyright_lawsuits.html

A Multidimensional Model for Monitoring Cloud Services

Nuno Palhares, Solange Rito Lima, and Paulo Carvalho

Centro Algoritmi, Departamento de Informática, Universidade do Minho,
Campus de Gualtar, 4710-057 Braga, Portugal
nunopalhares89@gmail.com, {solange,pmc}@di.uminho.pt

Abstract. The complexity of monitoring cloud environments and the lack of standards so far urge for a careful analysis, systematizing and understanding of key points involved when assessing the services provided. In this context, this paper proposes a layered model for Cloud Services monitoring, identifying the multiple dimensions of monitoring, while combining the perspectives of service providers and customers. This process involves the identification of relevant parameters and metrics for each monitoring dimension, focusing on monitoring of resources, quality of service, security and service contracts. Taking a stratified view of the problem, this study contributes to achieve a clearer and more efficient approach to cloud services monitoring.

Keywords: Cloud Computing, Cloud Services, Monitoring, Management.

1 Introduction

The provision of services based on Cloud Computing is becoming a trend and reality. This is mainly due to the decrease in capital and operational costs associated with this technology, combined with the potential advantages cloud computing brings.

In short, *Clouds* can be viewed as a large pool of virtualized resources (e.g. hardware, development platforms and services) easily usable and widely accessible. These resources can be dynamically reconfigured to be adjustable to variable loads, allowing enhanced and transparent resources utilization. The set of resources is typically available based on a *pay-per-use* business model, in which the infrastructure provider offers guarantees through customized SLAs (Service Level Agreements) [1].

The deployment of cloud services may follow distinct service models, commonly defined as IaaS (Infrastructure as a Service), PaaS (Platform as a Service) and SaaS (Software as a Service), and implementation models, either in the private or public sector, demanding different approaches for monitoring. In *Private Clouds*, resources and relevant data are typically under control and maintained within the organization premises. *Public Clouds* impose additional monitoring requirements mainly due to the wide geographical coverage and the large set of resources involved, requiring extra flexibility, scalability and security concerns. In particular, security issues can affect monitoring between cloud service providers, limiting interoperability.

A relevant topic to be taken into account in cloud monitoring is energy issues. The challenge in Green Cloud Computing involves minimizing the use of resources and

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 931–938.
DOI: 10.1007/978-3-642-36981-0_87

still meeting the required quality and robustness of the service, contributing to a reduction in operational costs and environmental impact [2], [3].

From a customer/provider perspective, offering cloud services also rises economic and contractual issues according to the negotiated SLAs, and corresponding QoS and QoE compromises. SLA compliance is the first step to a profitable interaction between customers and service providers, being a raw mechanism for mutual control. The role of monitoring is therefore reinforced to ground mutual SLA auditing.

In this context, articulating the various aspects cloud monitoring rises, this paper proposes a stratified approach to cloud services monitoring. For each layer, the model identifies the main parameters and metrics to consider, thereby creating an integrated approach for monitoring the different dimensions and perspectives of participating entities. Our aim and main contribution is therefore gathering, clarifying and systematizing major issues involved in cloud monitoring in order to ground and foster the development of comprehensive and flexible monitoring services.

This article is organized as follows: related work is discussed in Section 2; the proposed stratified monitoring model for cloud services is presented in Section 3; and the main conclusions and future work are included in Section 4.

2 Related Work

One of the constant concerns of service providers is related to monitoring and management of cloud services. Table 1 includes monitoring tools currently available to sustain these tasks, being here classified according to the technique and paradigm followed. As shown, the monitoring location can either be local or remote. In the latter, monitoring tools are distributed and scalable systems supporting high performance computing, such as cluster or grid platforms. Monitoring is usually complemented using web-based management platforms.

Table 1. Monitoring tools

Type	Examples
Local	Sysstat (Isag, Ksars), Dstat.
Remote	Nagios, Ganglia, GroundWork, Cacti, MonALISA, GridICE.
Web Management Platforms	RightScale, Landscape, Amazon CloudWatch, Gomez, Hyperic/Cloud Status, 3Tera, Zenos, Logic Monitor, Nimsoft, Monitis, Kaavo, Tap in systems, CloudKick, Enstratus, YLastic, TechOut, ScienceLogic, Keynote, NewRelic.

In addition to the tools mentioned above, examples of relevant ongoing projects are *Lattice* [4] and *PCMONS* [5]. Lattice is a framework designed primarily to monitor resources and services in virtual environments. Lattice uses a probe-based monitoring system to collect data for the management system. This framework was developed and implemented in conjunction with the RESERVOIR project. RESERVOIR is a cloud service that distinguishes service providers and infrastructure providers and aims to increase the efficiency of computing, enabling the development of complex services. Both geographical, quality and security issues are covered.

PCMONS assumes that monitoring can take advantage of concepts and tools already present in the management of distributed computing. Its main objective is to implement a monitoring system for Private Clouds and IaaS model, using open source software (e.g. Nagios). The architecture of the monitoring system comprises three layers and equates to a centralized model based on client/server connections. The base layer includes infrastructure components, the middle layer (Integration Layer) is responsible for abstracting the details of the infrastructure, allowing the system to be adaptable and extensible (plug-ins) to other scenarios/tools and the top layer provides an interface to assess the compliance with established policies and SLAs.

From the literature review it became evident that there is no consensus in defining monitoring solutions that satisfy all requirements of complex cloud environments. Through a layered view of the problem, we expect to contribute toward an efficient management and optimization of cloud services deployment.

3 Stratified Monitoring of Cloud Services

The proposed model for cloud services monitoring is stratified into four main layers, which are then divided into categories. As shown in Figure 1, the main layers are: Infrastructure; Network; Service/Application; and Customer/Provider. The Infrastructure layer covers monitoring of both physical and virtual resources involved in the cloud computing environment. Apart from the need to monitor distinct components that compose an entire infrastructure, there are other components that should be monitored at this level, namely energy and security. Aspects related to the IP service, throughput, performance and reliability are covered at Network layer. The Layer Service/Application is focused on assessing the availability, efficiency, reliability and safety of a service. Finally the relationship Customer/Service Provider is considered, targeting SLA auditing and accounting (usage and cost) of a specific service. The sections below discuss the four layers in more detail.

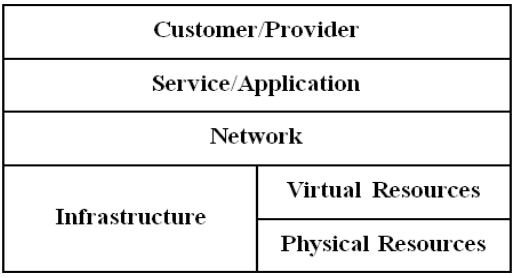


Fig. 1. Proposed Stratified Model for Monitoring Cloud Services

3.1 Infrastructure

As the foundation for the cloud computing architecture, the physical infrastructure is the major focus of monitoring. All physical components, from processing and storage

devices to network equipment, should be monitored. Most studies in the area agree in considering the percentage of CPU utilization, RAM, storage memory, and statistics of physical network interfaces as the most relevant metrics [2], [3], [5], [6], [7]. As mentioned, network devices must also be monitored, as problems in switches, routers or communication links may affect the cloud topological connectivity. An unstable topology may cause problems which influence a whole range of aspects, such as traffic engineering, throughput, service availability, SLA fulfillment, economic issues, among others. Table 2 summarizes the metrics to consider at Infrastructure level.

Regarding energy consumption, it is known that high temperatures reduce the lifetime of the devices, influencing the reliability and availability of the system. In turn, energy management procedures can affect the system performance in a complex way, since the overall rate computation results from the speed and coordination of multiple elements in a system [8]. According to [8], where ecological and performance issues of resource management system are taken into account (metrics, techniques, models, policies, algorithms), the power consumption is considered an adequate metric to address (see also [3]). In [2], metrics for temperature control and backup power systems (generators, UPS) are also proposed. The purpose is mainly to assess and optimize the use of energy and reduce the emission of carbon monoxide.

Concerning the infrastructure security in terms of physical resources, several restrictions and audits to cloud security are recommended in [9], [10]. Based on work carried out by Cloud Security Alliance (CSA), the cloud is modeled in seven layers, namely: *Facility, Network, Hardware, OS, Middleware, Application and User*. From this model, the first three levels (*Facility, Network* and *Hardware*) need to be considered at physical resources level. The resulting metrics are shown in Table 2.

Table 2. Sample metrics for Physical Resources layer

Layer	Category	Sample Metrics
Infrastructure Physical Resources	Components	CPU (usage, number of cores), RAM (usage), memory storage (usage, speed of reading and writing), network interface statistics, topology connectivity.
	Energy	Energy consumption, temperature, generator state.
	Security	Fire alarms/sensors, surveillance, access control, IDS and IPS monitoring, firewalls, authentication systems.

Regarding *Facility*, security is mainly handled at physical level, involving the implementation of access control through video surveillance, authentication systems, alarm systems and sensors, among others. The main objective is to prevent malicious intrusion and data manipulation, ensuring the integrity of facilities and components. At *Hardware* level, security metrics are in line with those adopted in the premises where security protocols should be followed up. Regarding *Network* level, which can be described as the boundary between customer data and the customers, mechanisms such as firewalls, Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS) can be adopted.

Within the Infrastructure layer, virtual resources assume a crucial role, increasing transparency, dynamics and scalability, therefore. Virtualization processes involve

operations such as suspend/resume/migration and start/stop of Virtual Machines (VMs). Common metrics at this level are mainly related to the percentage of CPU usage [8], RAM and memory storage of VMs (see Table 3). Statistics on the network interfaces of VMs are equally relevant. Operations related to creation and migration of VMs or number of active instances are also useful information [2], [3], [5], [6], [7].

The security of virtual infrastructure resources can be associated with *OS* and *Middleware* layers [9], [10]. In this case, the metrics should be extracted from monitoring OS-level events and system calls between VMs and hardware. The purpose is mainly to prevent copy and data violations. The *Middleware* layer is considered a potential weak point [10], due to its location between OS and Application layers, involving many components according to the service and architecture. At this layer, the metrics should then be related to monitoring of virtualization and safety systems in heterogeneous cloud architectures.

Table 3. Sample metrics for Virtual Resources layer

Layer	Category	Sample Metrics
Infrastructure Virtual Resources	Components	CPU (usage, number of cores), RAM (usage), memory storage (usage, speed of reading and writing), statistics of VM interfaces, VM migration, number of active instances.
	Security	Monitoring events and OS at call system level between VMs and hardware.

3.2 Network

In this layer, the relevant metrics are mainly at IP service level. As illustrated in Table 4 these metric types are classified as: Throughput, Performance, Availability and Reliability/Efficiency. The metrics involved derive from telecommunications and computer networking areas, resulting mainly from standardization efforts within ITU-T and IETF IP Performance Metrics (IPPM) workgroup.

Throughput is considered an essential parameter in cloud monitoring [11], [12]. Apart from its relevance for traffic engineering decisions, the verification of SLA fulfillment involves the assessment of throughput related metrics [13]. When analyzing traffic volumes per time unit, monitoring at service class level can bring benefits, particularly for the optimization of network utilization, identification of configuration problems within service classes, etc. Bandwidth quantifies the volume of data that a link or path is able to transfer per time unit. The available bandwidth, thus represents a metric variable in time, where the available capacity is identified, taking into consideration the current load. The capacity, representing the upper bound on available bandwidth, is also a metric that fits in this category.

In [5], [6] statistics related to network traffic are also identified as important sources of monitoring data. This information can also be useful at network layer in addition to physical and virtual infrastructure layer, as mentioned previously.

Performance metrics related to the network level include traditional QoS metrics such as packet duplication, packet loss (OWPL - One-way packet loss, OWLP - One-way

loss pattern, IPLR - IP packet loss ratio), delay (OWD - one way delay, RTT - round-trip time, IPTD - IP packet transfer delay, IPDV - IP packet delay variation), IP packet error ratio (IPER) and Spurious IP packet ratio (SPR) [7].

Regarding Availability, a network can present downtime periods caused by problems in network components, routing configurations, among other aspects. Thus, it is important to monitor the (un)availability of a network, as well as connectivity.

For Reliability/Efficiency assessment, the response time to a network configuration can be a relevant indicator. Upon the occurrence of a network failure, the mean time between failures or the average time to recover are common reliability indicators.

Table 4. Sample metrics for Network layer

Layer	Category	Sample Metrics
Network	Throughput	Traffic volume per time unit, used and available bandwidth, capacity.
	Performance	Packet duplication, packet loss (OWPL, OWLP, IPLR), delay (OWD, RTT, IPTD, IPDV), IPER, SPR.
	Availability	Uptime, (un)availability of the network connectivity (one or two-way).
	Reliability/ Efficiency	Response time (average/ maximum), mean time to repair upon failure, mean time between failures.

3.3 Service/Application

In the Service/Application layer, the nature of the monitored parameters and how they should be collected depends essentially on the software being monitored and not on the cloud infrastructure per se. One of the main concerns to be taken into account is the availability of a Service/Application. Measuring availability includes registering the periods of time during which a service is running and when it is unavailable. This topic also involves economic issues because in case of unavailability of a Service/Application, SLA violations and subsequent penalties at supplier side may occur. Apart from Availability, relevant metrics for this layer are classified in Reliability/Efficiency and Security (see Table 5). Regarding Reliability/Efficiency, the response time of a given service is a common indicator of efficiency. For instance, in [13], the average and maximum response time are defined as metrics for an online games scenario in Cloud Computing. In case of service failures (due to service unavailability or to QoS degradation), the time to repair should be provided to customers or third-parties responsible for monitoring. The time interval between occurrence of failures is also a measure of efficiency.

The insecure nature of the environment where services and applications are offered turns security into a fundamental aspect to control. As indicated in Table 5 the number of security vulnerabilities is a relevant metric, since it is necessary to monitor behavior to detect possible violations. Other aspects to be monitored and safeguarded are mostly digital certificates, private keys, etc.

The user behavior may also be considered at this layer, including relevant metrics such as login processes, access patterns and associated IPs. Monitoring should also

focus on managing passwords, controlling the format of passwords and how often they should be renewed [11]. Specific metrics for each Service/Application type should also be considered. Furthermore, it is useful to maintain a history, which may contain the IP addresses accessing the service and the login times for each client.

Table 5. Sample metrics for Service/Application layer

Layer	Category	Sample Metrics
Service/ Application	Availability	Uptime, service (un)availability.
	Reliability/ Efficiency	Response time (average/ maximum), mean time to repair upon failure, mean time between failures.
	Security	Number of security vulnerabilities, access patterns, login processes, password management.
	Others	Login times and IP access records (historic), specific metrics depending on service/application.

3.4 Customer/Provider

In technical terms, the Customer/Provider relationship relies on SLA negotiation. Formally, an SLA is a service contract specifying administrative and technological issues for the type of services provided, and a complete description of each service regarding QoS, uptime, security, privacy, backup procedures, responsibilities and compensation of both parties, among others [14]. A further reference may exist to issues related to geographic location of resources (e.g. datacenters) according to national and international laws. This is an important decision criterion for companies planning to invest in cloud-based solutions [15].

Regarding the management of services, an SLA acts as a valuable auditing instrument both for clients and service providers. The verification of SLA compliance is a cross-layer task spanning all the layers described above. For example, in [5], a metric for average SLA violations is obtained based on the average of CPU usage that was not allocated to an application when requested. Monitoring cloud services usage is also relevant due to the elastic nature of cloud environments, associated with the business model "*pay-as-you-go*", therefore, measuring use and cost become vital aspects [7], [13]. The accounting of services and corresponding revenue allows service providers to adapt pricing and business strategies according to market needs. The sample metrics for the Customer/Provider layer are summarized in Table 6.

Table 6. Sample metrics for Customer/Provider layer

Layer	Category	Sample Metrics
Customer/ Provider	Auditing	Monitoring SLA violations, penalties.
	Accounting	Monitoring of usage and cost, revenue.

4 Conclusions

Cloud monitoring is a recent and active research area where the lack of related standards is evident. This fact is particularly important and complex when trying to

perform monitoring of cloud services across multiple clouds, involving geographical, quality and legal issues. Contributing to the efforts toward modeling and standardization, this paper has proposed a stratified approach identifying and suggesting parameters, metrics and best practices for efficient monitoring of cloud services and environments. Future work includes validating and tuning the proposed model resorting to an experimental scenario and forthcoming activities in the area.

Acknowledgments. This work is funded by FEDER through the Competitiveness Factors Operational Programme - COMPETE and Portuguese National Funds through FCT - Foundation for Science and Technology under the Project: FCOMP-01-FEDER-0124-022674.

References

1. Vaquero, L., Merino, L., Caceres, J., Lindner, M.: A Break in the Clouds: Towards a Cloud Definition. SIGCOMM CCR 39(1), 50–55 (2009)
2. Werner, J., Geronimo, G., Westphall, C., Koch, F., Freitas, R.: Simulator Improvements to Validate the Green Cloud Computing Approach. In: 7th Latin American Network Operations and Management Symposium (LANOMS), pp. 1–8 (October 2011)
3. Beloglazov, A., Abawajy, J., Buyya, R.: Energy Aware Resource Allocation Heuristics for Efficient Management of Data Centers for Cloud Computing. ELSEVIER, Future Generation Computer Systems 28, 755–768 (2012)
4. Chaves, S., Uriarte, R., Westphall, C.: Toward an Architecture for Monitoring Private Clouds. IEEE Communications Magazine, 130–137 (December 2011)
5. Clayman, S., Galis, A., Chapman, C., Toffetti, G., Merino, L., Vaquero, L., Nagin, K., Rochwerger, B.: Monitoring Service Clouds in the Future Internet, pp. 115–126. IOS Press (April 2010)
6. Peng, Y., Chen, Y.: SNMP-Based Monitoring of Heterogeneous Virtual Infrastructure in Clouds. In: 13th APNOMS, pp. 1–6 (September 2011)
7. Choi, T., Kodirov, N., Lee, T., Kim, D., Lee, J.: Autonomic Management Framework for Cloud-based Virtual Networks. In: 13th APNOMS, pp. 1–7 (September 2011)
8. Sheikhalishahi, M., Grandinetti, L.: Revising Resource Management and Scheduling Systems. In: 2nd Int. Conf. on Cloud Computing and Services Science, pp. 121–126 (2012)
9. Spring, J.: Monitoring Cloud Computing by Layer, Part 1. IEEE Security & Privacy Magazine 9(2), 66–68 (2011)
10. Spring, J.: Monitoring Cloud Computing by Layer, Part 2. IEEE Security & Privacy Magazine 9(3), 52–55 (2011)
11. Chaves, S., Westphall, C., Lamin, F.: SLA Perspective in Security Management for Cloud Computing. In: 6th Int. Conf. on Networking and Services, pp. 212–217. IEEE Computer Society (March 2010)
12. Sousa, F., Moreira, L., Santos, G., Machado, J.: Quality of Service for Database in the Cloud. In: 2nd Int. Conf. on Cloud Computing and Services Science, pp. 595–601 (2012)
13. Patel, P., Ranabahu, A., Sheth, A.: Service Level Agreement in Cloud Computing. Technical Report (September 2009)
14. Cloud Computing Use Case Discussion Group: Cloud Computing Use Cases, white paper v4.0. Technical Report (July 2010)
15. Stamou, K., Morin, J., Gateau, B., Aubert, J.: Service Level Agreements as a Service - Towards Security Risks Aware SLA Management. In: 2nd International Conference on Cloud Computing and Services Science, pp. 663–669 (2012)

Effects of NGNs on Market Definition

João Paulo Ribeiro Pereira

School of Technology and Management, Polytechnic Institute of Bragança (IPB), Portugal
jprp@ipb.pt

Abstract. Historically, electronic communications networks were built to support specific services. For example, fixed and mobile telephone networks were developed to support voice telephone calls, whereas cable networks, satellites, and over-the-air broadcasts were built to support television services (unidirectional linear video). However, these communications networks are currently incorporating new technology and are rapidly evolving into multi-service networks that support voice, video, and data over a single, fully integrated communications platform. NGNs provide to customer access to a large range of services, leading to the increase of the bandwidth demand - For example, if customers encounter their demand on a single network, a triple play product, the bandwidth demand for that network will increase. Moreover, the migration to NGN may require upgrades to the infrastructure to provide sufficient service quality. The entry of new competitors can be based on the resale of services from the incumbent, on building up their own infrastructures, on renting unbundled infrastructure from incumbents, or, on the combination of the above elements. The availability of these options to competitors and price definition are generally determined by regulatory policies. So, the introduction of NGNs by telecommunication network operators obligates the national regulators adapt their access regulation regimes to the new technological conditions. Regulation and/or promotion of competition by regulatory measures need to be analyzed and compared. So, in this paper we explore the role of competition policy and regulation.

Keywords: NGNs, Broadband Access Networks, Telecommunication network operators, policy and regulation.

1 Introduction

New applications and greater Internet use have increased the demand for broadband connections. The traditional copper-based access networks will not be capable of supporting this increase in traffic for much longer. There will come a time when existing access networks can no longer meet increasing customer expectations. As a result of the introduction of NGNs, operators need to upgrade their access networks because in several cases existing access networks can no longer meet increasing customer expectations. Evolving consumer expectations will require changes to the existing access network – next generation access. However, existing technologies faces some difficulties and are not ready for large-scale roll-out yet [1, 2]. In the case

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 939–949.
DOI: 10.1007/978-3-642-36981-0_88

of DSL technologies, the great majority of operators with copper networks are improving their networks, making investments to deploy fiber optics closer to customers and offering higher-speed access, which is required for new emerging services (reducing the distance between fiber and the users.). However, the bandwidth of DSL technology is depend of the distance, and is currently limited to 16 Mbps for ADSL2 solution and 52 Mbps for VDSL solution - Where loop lengths are sufficiently short, either to the central office or to the cabinet, VDSL will often represent the preferred price/performance choice in the near to medium term. HFC operators of cable TV networks have to invest in return link capabilities, bandwidth shared by users. Like DSL, HFC technology need to reduce the distance between fiber and the users. Fiber solutions, particularly Passive Optical Networks (e.g. FTTCab, FTTC, FTTB, FTTH), provide downstream bandwidths up to 622 Mbps and high distance range (up to 20km). However, the deployment involves relative high construction costs (particularly trench and ducts – civil works). PLC uses the power supply system, but there are some unsolved problems with interference and range.

Additionally, the power network operators would have to invest significant sums in their infrastructure. Mobile telephony networks currently are being improved by setting up UMTS (384 kbps), HSDPA (2 Mbps), and LTE (up to 100 Mbps). However, UMTS and HSDPA technologies will not provide sufficient bandwidth in the near future to be regarded as a competitive network for triple play bundles and other applications demanding high bandwidth. Some broadband wireless access technologies offer promising opportunities with speeds comparable with fixed solutions.

2 Challenges to Regulation

The selection of the optimum technological solution for an NGA network depends upon the business model and ambitions, as well as the current position of the operator [3-6]. New technologies are creating new possibilities for last-mile competition, although the last mile continues to represent a market segment with high initial cost and low marginal cost in which only a limited number of telecommunications companies find it cost-effective to create and maintain network infrastructure [7].

The access network is usually the most expensive component in terms of capital investment (specifically passive infrastructure) and OA&M costs. Of the several costs, civil engineering costs are greatest when it is necessary to run a new fiber or copper connection to the cabinet, building, or home. Moreover, access to existing infrastructure, such as the ducts of the incumbent or other market players or sewage pipes, is critically important to avoid digging. For [8], a local loop network can be divided into three main layers or segments: a service layer and two infrastructure layers (see Figure 1.). Layer 1 includes passive infrastructures, such ducts and cables, and requires the greatest investment. Layer 2 consists of active infrastructures, such as the technical installations at the end of the fibers that send, receive, and manage the optical signals. Layer 3 includes several services that consumers buy from telecommunication operators.

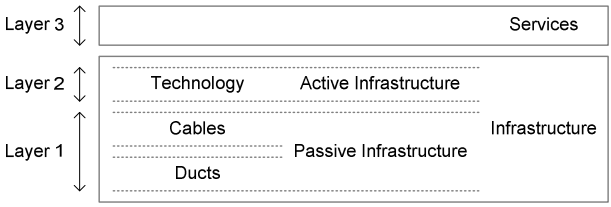


Fig. 1. Network layers [8, 9]

The arrival of NGAN has implications for the competitive conditions in access markets. In the access network, the challenges posed to the implementation of loop unbundling by the migration to VSDL and by FTTB/FTTH raise serious questions regarding how it might best be achieved that remain unanswered [7]. Experimentation in different countries is likely to provide insight into whether some combination of access to street cabinets (sub-loop unbundling), rights of way, ducts, and/or building wiring might be sufficient to maintain the effectiveness of loop unbundling and shared access as a ladder-of-investment complement to bitstream access.

Typically, without regulation, the resulting market outcome is likely to be an under provision of broadband services in sparsely populated regions and an overprovision in metropolitan areas [8]. Under provision may result from the high fixed costs associated with the network deployment if demand is too low. On the contrary, overprovision might be due to network duplication, which is a likely result in regions where deployment costs are low.

2.1 Competition in Next Generation Access

The emergence of technologies enables the introduction of new services, and opens up new revenue sources. While previously the main value of telecom was to realize simple communication between people, nowadays several new elements are added including mobility, personalization, portability, higher quality, etc. [10]. So, the offer of new services such as content, games and other broadband services, involves a number of business players (also known as actors). These players include not only traditional telecom roles (service operator, network operator), but also players from other industries (e.g. content, IT, consumer devices, etc.). Each player can represent several roles (for example, in the first scenario presented in Figure 2, the network operator acts as a service operator and in the second scenario they are different players. These changes have also raised the significance of the regulatory role.

Figure 2 show two basic service provisioning models. In the first scenario, the network operator (which acts as service operator) and the content provider are the two main business players. The network operator is responsible for the correct provisioning of contents (e.g. downloads of music, games and videos, data, …) to the subscriber. In this business model structure, the contents are created by the content producer and sell to the subscriber by the network operator. The network operator is responsible for the provisioning of the bandwidth that the service operator offers to the subscriber. In the second scenario, the user buys services from the service operator

and thus subscribes for telecommunication services. Service operator acts as the main responsible player towards the subscriber. In order to reach its customers, and provide them with services (like voice and video telephony, Internet access, value-added services etc.), service operator needs to buy network access and transport services from the network operator. Network operator is a player who operates both access and core portions of a network infrastructure. In this scenario, network operator is the responsible for the network equipment purchase and maintenance cost, and service operator for the management costs.

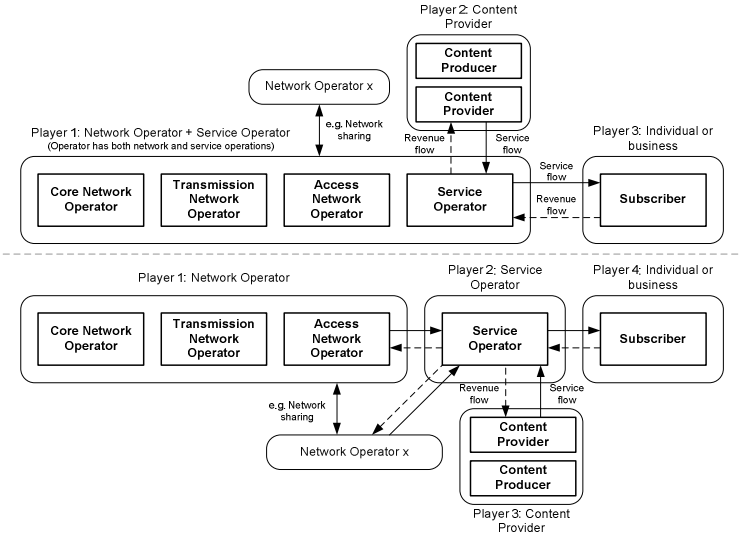


Fig. 2. Business models: basic service provision [11]

For example, [12] argues that European mobile operators must separate the network operator and service operator businesses to face the diminution voice revenues (because of the increasing competition and consequently the reduction of the prices). Network operators are extending their network coverage and capacity, whereas service operators promote new services that are more profitable and more attractive to the customers (e.g., providing new and advanced value-added data and content services such as rich video and audio, games and m-commerce).

2.2 Market Situation

Despite the key objective of the EU is to promote investment in new infrastructures [13], this goal has not been fulfilled in Europe. A high level of investment is a strong prerequisite for building networks that permit the provision of emerging broadband services. Therefore, whereas entrants have the theoretical alternative of building copper networks, it is not likely to happen. The strategic selection scope for deployment is, therefore, limited to the decision of which type of FTTH to deploy. A key decision in FTTH deployment is how far from the subscriber the access

node/switch should be placed. Under structural separation, the incumbent generally sells wholesale services to other providers, who then market final services to the users. That is, the infrastructure access charges shape the final prices [14].

Table 1 shows the broadband access lines market share. The market share of the incumbent fixed operators since July 2003 has followed a downward trend which is now stabilizing around 44% of the broadband market.

Table 1. Fixed broadband lines - operator market shares at EU level and PT, 2006 - 2010 (Source: EC)

Level	Operator	Jul-06	Jan-07	Jul-07	Jan-08	Jul-08	Jan-09	Jul-09	Jan-10	Jul-10
EU	Incumbents	47,7%	46,9%	46,8%	46,1%	45,7%	45,5%	45,2%	45,0%	44,0%
	New entrants	52,3%	53,1%	53,2%	53,9%	54,3%	54,5%	54,8%	55,0%	56,0%
PT	Incumbents	72,1%	71,1%	70,1%	39,7%	39,1%	40,6%	41,9%	43,0%	45,9%
	New entrants	27,9%	28,9%	29,9%	60,3%	60,9%	59,4%	58,1%	57,0%	54,1%

Figure 3 illustrates the DSL market share evolution by type of operator (incumbent and new entrants) in the DSL market (Portugal and EU) - trend in the % of DSL lines provided by incumbent fixed operators. The incumbents' share of the DSL market stabilizes around 55% in the EU level, decreasing since 2005. However, the figure and the report of EC [15] shows that in Portugal the market share of the incumbent operator continued to increase since July 2008. The new entrants' market share decreased in the last year for both non-DSL lines and DSL lines.

In EU cable market, new entrants share was 96.5 % in January 2011 (96.9 % in 2010) and incumbents market share only 3.5 % (3.1 % in 2010). The Market share of the incumbent and new entrants in other internet technologies (not DSL or cable) is: 90.6 % share for new entrants and 9.4 % for incumbent operators.

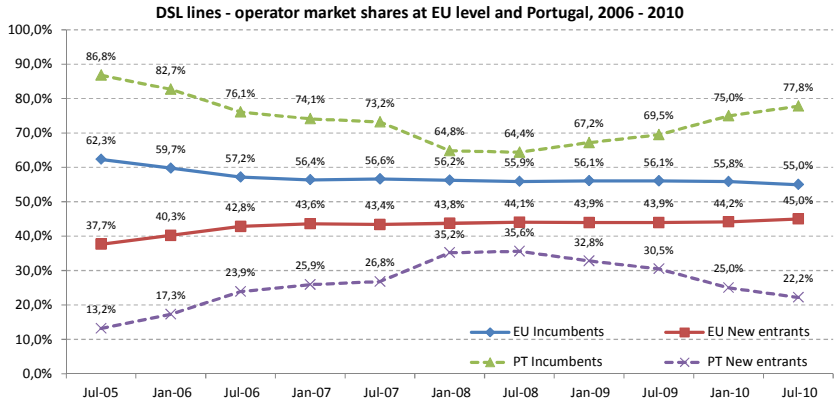


Fig. 3. DSL lines - operator market shares in DSL market: EU level and PT, (Source: EC)

3 Regulatory Options/Strategies for Access Networks in NGNs

The advent of NGN (new network technologies, access infrastructures, and even services) has changed the concept of telecommunication networks and has profound implications for operators and regulators. The definition of policies and regulations for competition in the access networks constitute one of the most debated issues in telecommunications today. The regulation of telecommunications networks and services is seen as a necessary requirement in most countries to meet government objectives and to ensure public interest [16]. Regulation is fundamental to generate positive welfare effects where markets alone would not tend to perfect competition.

But, as referred by [17], the major problem is how to measure these welfare effects, as they can occur as consumer surplus, producer surplus, societal gains (e.g., increased tax income, better working conditions, etc.). Their empirical study uses price situation to examine the welfare effects measured by the state of competition. They assumed that the increase of competition reduces prices in the market and that competition can also increase consumer welfare without reducing prices (achieved by innovation). Public policies should promote an efficient investment and competition in all markets (see Figure 4).

Regulators face substantial challenges in dealing with the evolution of the technology of electronic communications networks from circuit switching to packet switching based on the IP. As would be expected, NGNs have different configuration which will have competitive and regulatory implications. However, operators and countries are approaching the migration to NGN differently. The migration to NGN changes the character of competition substantially and brings about the decoupling of the service (provision) from the network (transport) [18, 19]. Specifically, in an IP-based NGN, any network can provide any service, any network can simultaneously carry multiple services, and a service provider does not have to be a network operator (and vice versa). Regulators face the question of how to adapt to this changing competitive and dynamic environment in which many incumbents have announced or begun the migration to NGN as well as the implementation of optical access networks [18, 20]. The objective with the NGA access regulation has been to foster investment and innovation in a new and enhanced infrastructure while preserving a strong market competition [21].

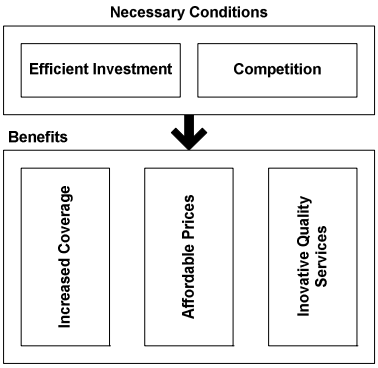


Fig. 4. Policies effects [22]

The two main economic reasons that have been used to justify interventions in access networks are the beliefs that access networks constitute a natural monopoly for which competition is not feasible in principle and that regulation is, therefore, necessary to control monopoly power and to achieve universal service in which all (or most) users have the opportunity to affordably access the services of the network. The challenge of telecom operators to provide a profitable deployment of broadband services depends if is a high or a low competition area. In areas with high competition already exists competition between broadband network operators, and the main question is know the market share of all intervenient. However, in low competition areas high investments cost must be incurred to promote broadband. [3] argued that national or regional policy concerns can also affect NGA roll out. Without some type of intervention, there is the risk for a new digital divide, with urban customers on short loops being able to receive IPTV/multi-media services and HDTV while those in rural areas might not be able to receive such services. Therefore, the access network poses serious challenges to the regulator [7].

The question then becomes whether it is more important to stimulate investment or to ensure competition. Investment in network quality is important for consumers because it provides access to both better quality and speed to services, such as Web browsing and email, and services that require more bandwidth, such as video. Investment in network quality also improves the service value for consumers and attracts new consumers to the market. The promotion of competition in the telecommunications market means supporting competition not only among incumbents' competitors, but also between incumbents and their competitors [23].

Therefore, there are two major options for access regulation [3, 20]: temporary or permanent deregulation (i.e., the removal of sector-specific rules and regulations) or mandated access (i.e., the obligation to grant access to bottleneck facilities at a regulated price and quality). Deregulation increases investment incentives, as it overcomes the "truncating problem" and allows above-normal profits. However, in the absence of alternative infrastructures or in areas of low population density under limited competition or the threat of entry into the upstream market, an integrated incumbent might leverage its market power to competitive downstream segments. Normally, the deregulation in NGNs may be applicable in competitive normal-speed markets. In the case of high-speed markets, deregulation incentives investment, but exist the risk that operators' market power is likely to increase market concentration in the long run.

For NRAs, one request of decisive importance is if they must foster service-based competition in the first phase of liberalization or to focus on infrastructure-based competition. This decision (infrastructure or service-based) would lead to lower prices, more differentiated and innovative products and improved services for consumers [17].

When access is available at different levels of the incumbent's network, new entrants will be able invest in the infrastructure gradually as sufficient economies of scale became achievable [24] - This concept is the ladder of infrastructure competition. This concept defends that new entrants (or access seekers) may enter the market offering broadband access by reselling the wholesale services of the

incumbent operator (requires least investment) where they only cover minor elements of the value chain (Figure 5). When the number of customer grows and financial means become available, the operator move on to higher rungs of the ladder [17, 25]. Next, new entrants need to building their own infrastructure and acquiring only the residual infrastructure from the incumbent's wholesale department. This includes a move for the operators from service to infrastructure-based competition.

The migration to NGAN has raised a range of issues related to building wiring and infrastructure sharing. The deployment strategies for operators and entrants are completely different. In addition, parameters, such as existent infrastructure, geographical characteristics, infrastructure renting costs, and consumer willingness to pay, influence the definition of the strategy. So, telecommunication operators can select among a set of deployment strategies that are characterized by path dependency and diminishing usage of the legacy copper loop [16]. The range of the selection space is based upon how much of the copper they use and, consequently, how far toward the customer they deploy new fiber. In the final step, operators replace all of the copper with FTTH. Within that scenario, FTTH can be implemented as either active Ethernet or passive optical networks, although most incumbent operators tend to select PON.

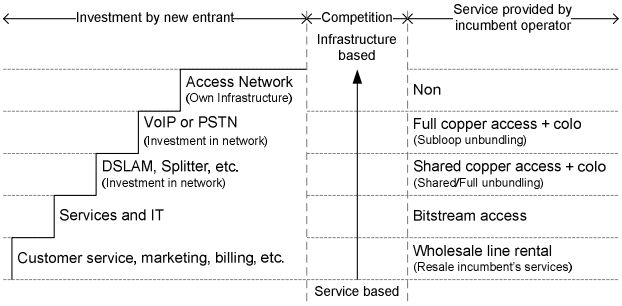


Fig. 5. Ladder of investment [24, 26]

The migration to NGAN has raised a range of issues related to building wiring and infrastructure sharing. The deployment strategies for operators and entrants are completely different. In addition, parameters, such as existent infrastructure, geographical characteristics, infrastructure renting costs, and consumer willingness to pay, influence the definition of the strategy. So, telecommunication operators can select among a set of deployment strategies that are characterized by path dependency and diminishing usage of the legacy copper loop [16]. The range of the selection space is based upon how much of the copper they use and, consequently, how far toward the customer they deploy new fiber. In the final step, operators replace all of the copper with FTTH. Within that scenario, FTTH can be implemented as either active Ethernet or passive optical networks, although most incumbent operators tend to select PON.

The strategy of a new entrant in an access network that does not have an infrastructure can be one of the following three alternatives [13, 16]: (1) Renting

infrastructure (i.e. conduit, cable, equipment, etc.) from other operators and offering only services (infrastructure sharing); (2) Deploying a new infrastructure; or (3) Not participating at all (see Figure 6). Figure 6 also shows the deployment strategies commonly used by incumbent and new entrants.

Regulators must decide whether to promote competition on the basis of a single infrastructure with regulated access (service competition) or to encourage the build-up of competing, parallel infrastructures (infrastructure competition) [27]. Then, is important create the right incentive for operators to make an efficient build/buy choice and define the appropriate pricing principles. To obtain economic efficiency, a regulator should [24]: (1) Encourage the use of existing infrastructure of the incumbent operator where this is economically desirable, avoiding inefficient duplication of infrastructure costs by new entrants (incentive to buy); and (2) Encourage investment in new infrastructure where this is economically justified by (1) new entrants investing in competing infrastructure, and (2) the incumbent operator upgrading and expanding its networks (incentive to build).

In this context, the cost models are fundamental in the determination of the access price that can be used by regulators in the definition of wholesale prices.

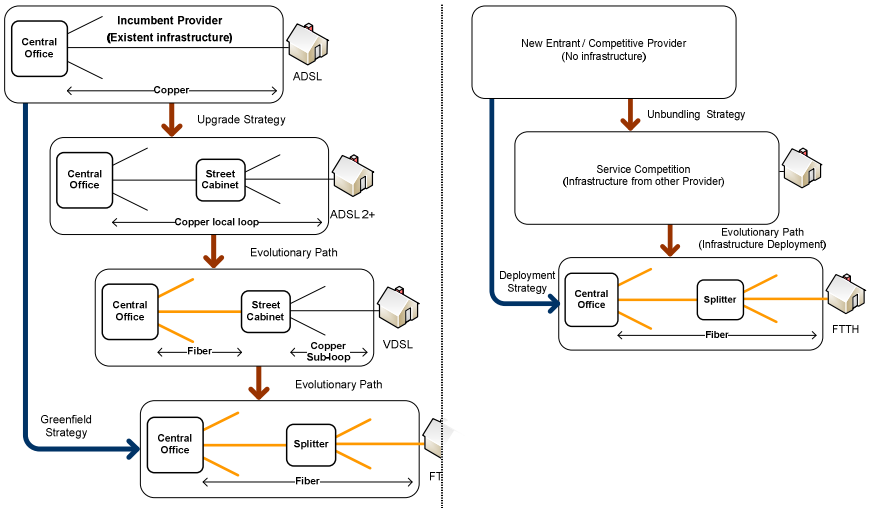


Fig. 6. Deployment strategies for incumbent operators and new entrants [9, 11]

4 Conclusions

The analysis of the broadband market suggests that where infrastructure competition exists, as in DSL and wireless broadband, service providers will more aggressively price their offerings, driving down the access price for consumers. However, in the case of limited infrastructure competition, broadband access price remains high for consumers. Infrastructure competition between DSL, Cable and wireless solution, had a significant and positive impact on the broadband penetration. We verify that

opening access networks (and network elements) to competitive forces increases investment and the speed of development. Despite increasing competition, incumbents are maintaining their dominant position. More than 60% of all broadband subscriptions make use of incumbent's broadband access infrastructure. In countries/regions where alternative technological platforms are not developed, the deployment of the DSL technology depends on the use of the networks infrastructures that are propriety of incumbent operators. To facilitate market entry of new competitors and develop competition in the access market, the regulatory authorities are focused on unbundled access to the local loop (fully unbundled local loop and shared access to the local loop) and on different forms of network access (bitstream and resale).

References

[1] ITU, Ruling the New and Emerging Markets in the Telecommunication Sector. Presented at the ITU Workshop on What rules for IP-enabled NGNs? Geneva, Swiss (2006)
[2] Pereira, J.P., Ferreira, P.: Next Generation Access Networks (NGANs) and the geographical segmentation of markets. In: The Tenth International Conference on Networks (ICN 2011), p. 6. St. Maarten, The Netherlands Antilles (2011)
[3] Collins, H.: Next Generation Networks - Creating a Dedicated Cost Model. InterConnect Communications Ltd., United Kingdom (2009)
[4] Pereira, J.P.: Telecommunication Policies for Broadband Access Networks. In: The 37th Research Conference on Communication, Information and Internet Policy, Arlington, USA, pp. 1–13 (2009)
[5] Pereira, J.P., Ferreira, P.: A Cost Model for Broadband Access Networks: FTTx versus WiMAX. In: 2007 Second International Conference on Access Networks (AccessNets 2007), Ottawa, Ontario, Canada, pp. 1–8 (2007)
[6] Pereira, J.P., Ferreira, P.: Access Networks for Mobility A Techno-Economic Model for Broadband Access Technologies (2009)
[7] Marcus, J.S., Elixmann, D.: Regulatory Approaches to NGNs: An International Comparison. Communications & Strategies 69, 21 (2008)
[8] Jaag, C., Lutzenberger, M.: Approaches to FTTH-Regulation: An International Comparison. In: Second Annual Conference on Competition and Regulation in Network Industries, Brussels, Belgium, p. 23 (2009)
[9] Pereira, J.P., Ferreira, P.: Game Theoretic Modeling of NGANs: Impact of retail and wholesale services price variation. Journal of Communications (2012)
[10] Kaleelazhicathu, R.K., Gjerde, I.G., Smura, T., Hämmäinen, H., Tesch, T.: Business models in telecommunications. In: ECOSYS 2004 (2004)
[11] Pereira, J.P., Ferreira, P.: Infrastructure Sharing as an Opportunity to Promote Competition in Local Access Networks. Journal of Computer Networks and Communications, 11 (2012)
[12] Harno, J.: Techno-Economic Analysis of Beyond 3G Mobile Technology Alternatives. In: 6th Conference on Telecommunication Techno-Economics, CTTE 2007, pp. 1–9 (2007)
[13] Chanclou, P., Gosselin, S., Palacios, J.F., Alvarez, V.L., Zouganeli, E.: Overview of the optical broadband access evolution: a joint article by operators in the IST network of excellence e-Photon/One. IEEE Communications Magazine 44, 29–35 (2006)

[14] Laffont, J.-J., Tirole, J.: Competition in Telecommunications. The MIT Press, Cambridge (2001)
[15] European-Commission, Portual: Main market developments, European Commission (2011)
[16] Sigurdsson, H.M.: Techno-Economics of Residential Broadband Deployment (PhD Thesis), Technical University of Denmark - Center for Information and Communication Technologies (2007)
[17] Kittl, J., Lundborg, M., Ruhle, E.-O.: Infrastructure-based versus service-based competition in telecommunications. Communications & Strategies, 67 (2006)
[18] Marcus, J.S., Elixmann, D., Wernick, C.: Next Generation Networks (NGNs). European Parliament, Brussels (2009)
[19] Elixmann, D., Figueras, A.P., Hackbarth, K., Marcus, J.S., Nagy, P., Pápai, Z., Scanlan, M.: The Regulation of Next Generation Networks (NGN): Final Report, Budapest (2007)
[20] Kirsch, F., Hirschhausen, C.V.: Regulation of Next Generation Networks: Structural Separation, Access Regulation, or no Regulation at all? In: First International Conference on Infrastructure Systems and Services: Building Networks for a Brighter Future (INFRA), Rotterdam, The Netherlands, pp. 1–8 (2008)
[21] Stordahl, K.: Mobile broadband subscription and traffic evolution. Presented at the International Telecommunication Society, 18th Biennial Conference, Tokyo, Japan (2010)
[22] Oliveira, A.: Next Generation Access Networks: A key factor of development. In: NGON Seminar – ISCTE, Lisbon, pp. 1–33 (2009)
[23] KPMG, Comparison of the Australian Fixed Network Cost Model and the TEA Model (2009)
[24] Andersen. Management. International, Pricing Shared Access in Sweden. Post-of Telestyrelsen, Sweden (2004)
[25] Xavier, P.: Geographically Segmented Regulation for Telecommunications. OECD (2010)
[26] ERG, Report on ERG Best Practices on Regulatory Regimes in Wholesale Unbundled Access and Bitstream Access. European Regulators Group (2008)
[27] Höffler, F.: Cost and benefits from infrastructure competition. Estimating welfare effects from broadband access competition. Telecommunications Policy 31, 401–418 (2007)

Ambient Assisted Living and the Integration and Personalization of Care Services

Alexandra Queirós[1] and Nelson Pacheco da Rocha[2]

[1] Health Sciences School, University of Aveiro
[2] Health Sciences Department, IEETA, University of Aveiro
Campo Universitário de Santiago, 3810-199 Aveiro, Portugal
{alexandra,npr}@ua.pt

Abstract. Ambient Assisted Living (AAL) can contribute for the integration and personalization of care services, which are important issues in the required reorganization of the care system to face their sustainability. This paper argues about the importance that AAL can assume as a centrepiece of the digital health continuum, from lifestyle management to hospitalization, if there are adequate information models to potentiate the integration of AAL systems-generated data and user-generated data with institutional information.

Keywords: Ambient Assisted Living, Electronic Health Record, Electronic Social Record, Personal Health Record.

1 Introduction

The challenges faced today by health and social care systems are related to theirs sustainability: with public budgets at strain the systems cannot afford to do less because demands and expectations are increasing, namely due to the demographic ageing.

Significant changes in organizational terms are required, which means, among other things, the opportunity to move from organization-centred care to a paradigm focused on the general welfare of the citizens [1]. Moreover, given the current pressure resulting from the cost of the health and social care systems, it is necessary, in a context of uncertainties and rapid changes, structural measures to increase the effectiveness and efficiency of these systems [2].

To adequately meet the demand patterns, ensuring quality and equal access, the health and social systems need to promote the integration and continuity of care. This leads towards the provision of services that are less organization-centred and more focused on the real needs of the citizens [3, 4].

The concept of personalized care can be understood as the individual customization of diagnosis and therapy. For instance, personalized medicine is a medical model based on information processing to customize decisions and therapeutics to the individual patient (*e.g.* using genetic or other information) by integrating medicine

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 951–958.
DOI: 10.1007/978-3-642-36981-0_89

with information technologies, namely, in terms of biomedical engineering, bioinformatics or medical informatics [5]. On the other hand, the empowerment of the citizens, namely the possibility of being actively involved in theirs health and care pathway, is also personalized care.

2 Ambient Assisted Living

The *eHealth* paradigm can be understood as the individualized provision of care services independent of the time these services are delivered as well as the physical location of actors and resources involved in the care process [4]. It can contribute both to the sustainability of health and social care systems and to the empowerment of the citizens.

Considering the idealized goal of the integration and personalization of care services, *eHealth* services are essential to provide better resources allocation and management, in accordance with the citizens' needs and those of the organizations providing care services [2].

Within the *eHealth* paradigm, Ambient Assisted Living (AAL) emerges as an important instrument within the required strategies to face the societal challenges related with the demographic ageing. The general goal of AAL solutions is to enable people with specific needs (*e.g.* elderly) to live longer in their natural environment. For that AAL systems and services should contribute to increase or maintain the individuals' performance in a broad spectrum of activities and participation: personal care, self-administration or scheduling, nutritional advice, maintenance of the house, getting involved, on a daily basis, in different leisure activities as well as engaging in social interactions [7].

A significant range of AAL systems is related with biomedical devices including [1]: mobile and wearable sensors able to monitor physiological parameters, infrastructures for the communication of these subsystems (*e.g.* body area network, wireless in-house network or mobile phone networks), distributed functions for data processing (*e.g.* alert management, cumulative registration or processing and presentation of multiple parameters) or decision support systems, namely for recognition and treatment of emergency situations.

Furthermore, AAL systems should acquire context information and combine multiple sources of information (*e.g.* environmental, physical or social information) and make pattern-based predictions to be able to track the individual's activities and behaviours (*e.g.* localization, presence of other people or personal and social actions). Effective AAL solutions should not be seen merely accordingly to the technological perspective. AAL systems should provide services inherently dynamic and able to complement, interactively, the human care providers. AAL solutions should combine the forces from both the technological and the societal ones, considering the integration and personalization of care, and they must be contextualised within a very complex ecosystem.

3 Concept

AAL systems have enormous potential, namely in terms of the generation and management of information required for the mediation between users, carers and institutions. In particular, AAL systems should contribute to optimize the cooperation between multiple care units and care providers. The various AAL care and assistance services must be integrated into collaborative environments allowing an effective and efficient cooperation among multiple actors, both in long term care and in specific time-limited missions.

The AAL systems can be, in terms of information for high quality services and for the empowerment of the citizens, a centrepiece, if there is the possibility to have information services that integrate both user-generated data (*e.g.* Personal Health Records - PHR) and AAL systems-generated data with institutional Electronic Health Records (EHR) [8] and Electronic Social Records (ESR) [9]. Therefore, it can be foreseen an important role of AAL within the digital health continuum that should be supported by *eHealth* services where health and social evidence and personal data can be used in the entire range from lifestyle management to hospitalization or institutionalization (see Fig. 1) [10]. For that adequate information models are required.

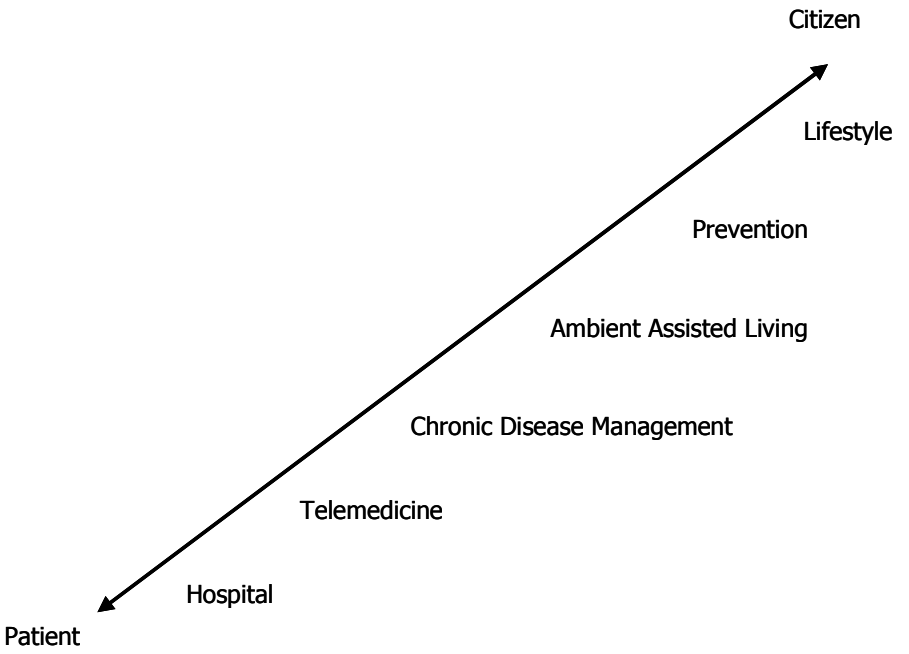


Fig. 1. The *eHealth* continuum

Such information services should contribute for the integration of heterogeneous compilations of resources to be acceded as a uniform conglomeration of information and knowledge. Therefore, efficient mechanisms to exchange and share information are essential. The availability of these mechanisms requires solving a broad range of technical and non-technical issues, such as interoperability, identification, security, governance, privacy and financial support to the common operations [11].

Due to the fact that current information management technology relies on a paradigm where real automation is only possible with formal, rigid processing rules based on the use of pre-designed structures for information containment [12], it is important the development of adaptable domain models separated from the implementations [12]. In this respect, it is required information management models that follow open management paradigms with two modelling levels for the information structure: the information model and the knowledge model.

The first level comprises the fundamental model for the technical implementation that should provide the required data types and be stable over time in order to be maintainable.

In the second level, domain and application specific concepts are modelled (archetypes) with constraint rules to specialize the underlying information model. According to this approach, archetypes are instances of an object oriented system implementation, which means they can be created and manipulated by adequate tools and alter as desired, without changing the underlying technical specifications of the information system [13].

A schematic view of the services architecture to implement the proposed information management model is presented in Fig. 2. Briefly, the data layer is responsible for the persistence of the data repository (*i.e.* the information model) accordingly to Health Level 7 (HL7) Reference Information Model (RIM) [14] and the common services layer provides resources to the top layers to ensure that they are able to communicate with the data repository with security and reliability (*e.g.* authentication, privacy and security, persistence, transport or management history, among other services).

The business layer allows the construction of complex operations, including those necessary for the conversion of messages to ensure interoperability in terms of the information that is imported or exported, and the knowledge layer has the required functions, in terms of archetype management, to shape the HL7 RIM to specific application domains (*i.e.* the knowledge model). Finally, the application layer is related with the domain specific applications.

4 Validation

The specification of the data model is conforming to HL7 RIM. Thus, the data layer must be a generic container and it was necessary to evaluate whether this container is adequate to accommodate all information objects that can be associated with AAL ecosystems.

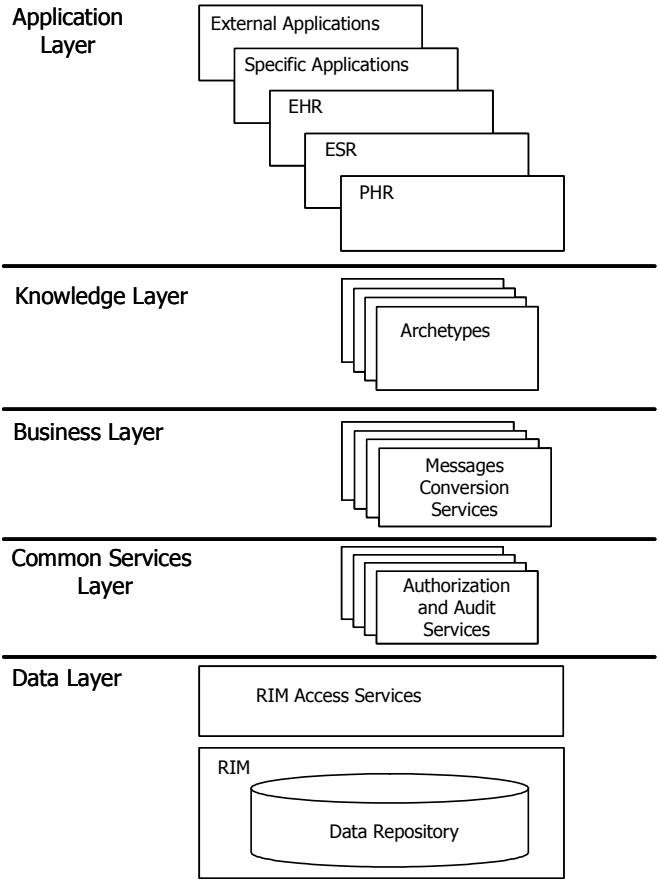


Fig. 2. Schematic view of the architecture to implement the information management model

The followed validation procedure had two steps. In the first step the data types of the HL7 version 2 [15] were mapped with the generic container of the data layer. The architecture of HL7 version 2 is completely different from the one of HL7 RIM and it is roughly characterized by a comprehensive description of different messages corresponding to different types of clinical information. Despite that, the validation demonstrated that the business layer (message conversion services) was able to convert the data types defined by HL7 version 2, so that the data layer can support them. This means that a significant range of data types of the health care field can be supported.

The data types of HL7 version 2 are clearly oriented for health care systems. However, a requirement of the presented information management model is that it should support other types of information besides the ones usually associated with EHR. The requirements are not only to consider EHR but also ESR, not only consider the information demanded by formal care providers but also by informal care

providers, including information provided by the citizen (PHR) and information generated by AAL systems.

Therefore, a second step was considered for the validation procedure. This consisted in the creation of scenarios, out of the health care field, and in the verification if all the information objects required by these scenarios were supported by the information management model. Furthermore, these scenarios emphasized the use of ESR and PHR and AAL generated information.

Keeping in mind that the service models used by health care providers and social care providers are different, there are important differences between EHR and ESR. Healthcare records should focus on a single patient often with considerable detailing and depth, and strongly protect the confidentiality of the individual. On the other hand, social care records should place the individual in their daily family life context as well as in other informal situations, recording the attitudes and effects seen on each situation, so as to ensure mental support and understanding [16].

Considering these requirements, it was possible to specify basic ESR archetypes with different types of entries (observation, evaluation and instruction) and contextual information (namely, procedural context and time, where the information was created, as well who is the author of the information). The created archetypes are compatible with the data layer and allow answering six basic questions linked to each record: what, when, where, who, why and how [17].

The basic PHR structure was based in the Health Information Form for Adults of the American Health Information Management Association (AHIMA) and the required archetypes (related with observations of daily living, namely, nutritional and physical activities) are compatible with the generic container of the data layer.

Finally, a scenario with mobile and wearable sensors to monitor physiological parameters has been used to verify if the data layer was able to accommodate AAL generated information.

The validation has shown that the generic container of the data layer is prepared to support information structures not only of EHR but also of ESR, PHR and AAL generated information. Therefore, the information management model is able to organize the information according to different requirements related with the provision of care services.

5 Conclusion

AAL systems should be able to complement the human careers. The AAL systems ability not only to acquire vital signals but also to acquire and combine multiple sources of information to track the individual's activities and behaviours must be considered in terms of care personalization.

Therefore, it was introduced an information management model to create and maintain a wide range of information objects within AAL environments. The proposed model, which is being implemented under an on-going QREN project, can contribute for the integration of EHR, ESR, PHR and AAL generated information.

Complex human and organizational factors can either hinder or accelerate the integrated and personalized care vision. Many challenges still exist to deployment of interoperable EHR, ESR, PHR and AAL generated information. Therefore, further research is required. Multiple stakeholders (*e.g.* citizens, providers, employers, payers, governments or research organizations) must play key roles in developing interoperable technology and to overcome the barriers to widespread adoption.

Acknowledgements. This work was supported by COMPETE - Sistema de Incentivos à Investigação e Desenvolvimento Tecnológico, Projectos de I&DT Empresas em co-promoção, under QREN TICE.Healthy.

References

1. Codagnone, C.: Reconstructing the Whole: Present and Future of Personal Health Systems, PHS2020. European Commission, Brussels (2009)
2. Dias, A., Queirós, A.: Integração e Continuidade de Cuidados. Plano Nacional de Saúde, 2011-2016. Alto Comissariado para a Saúde, Lisboa (2010)
3. Camarinha-Matos, L., Afsarmanesh, H.: Collaborative Ecosystems in Ageing Support, Adaptation and Value Creating Collaborative Networks. In: 12th IFIP WG 5.5 Working Conference on Virtual Enterprises, São Paulo, pp. 177–188 (2011)
4. Blobel, B.: Architectural Approach to eHealth for Enabling Paradigm Changes in Health. Methods of Information in Medicine 49, 123–134 (2010)
5. Bos, L. pHealth. In: 9th International Conference on Wearable Micro and Nano Technologies for Personalized Health, Porto, pp. 3–13 (2012)
6. Eysenbach, G.: What is e.health? Journal of Medical Internet Research 3 (2001)
7. Genet, N., Boerma, W., Kringos, D., Bouman, A., Francke, A., Fagerström, C., Melchiorre, M., Greco, C., Devillé, W.: Home Care in Europe: a Systematic Literature Review. BMC Health Services Research 11 (2011)
8. Redesigning Health in Europe for 2020. eHealth Task Force Report. European Commission, Brussels (2012)
9. Defining the Electronic Social Care Record. Information Policy Unit - Social Care Department of Health, London (2004)
10. Boye, N.: Co-production of Health Enabled by Next Generation Personal Health Systems. In: 9th International Conference on Wearable Micro and Nano Technologies for Personalized Health, Porto, pp. 52–58 (2012)
11. Heeks, R.: Health Information Systems: Failure, Success and Improvisation. International Journal of Medical Informatics 75, 125–137 (2006)
12. Yli-Hietanen, J., Niiranen, S.: Towards Open Information Management in Health Care. The Open Medical Informatics Journal, 42–48 (2008)
13. Beale, T.: Archetypes: Constraint-Based Domain Models for Future-Proof Information Systems. In: 11th OOPSLA Workshop on Behavioral Semantics: Serving the Customer, Boston, pp. 16–32 (2002)
14. ISO/HL7 21731: Health informatics - HL7 Version 3 - Reference Information Model - Release 1. International Organization for Standardization, Geneva (2006)

15. Health Level Seven Standard Version 2.7.1 - An Application Protocol for Electronic Data Exchange in Healthcare Environments. Health Level Seven International, Ann Arbor (2012)
16. Rigby, M.: Integrating Health and Social Care Informatics to Enable Holistic Health Care. In: 9th International Conference on Wearable Micro and Nano Technologies for Personalized Health, Porto, pp. 41–51 (2012)
17. Queirós, A., Rocha, N., Carvalho, S., Pavão, J.: Integrated Care of the Elderly and the Continuous Development and Adaptation of Information Systems. In: 12th IEEE International Conference on e-Health Networking, Application and Services, Lyon (2010)

A Mobile and Web Indoor Navigation System: A Case Study in a University Environment

Sara Paiva

Viana do Castelo Polytechnic Institute
Technology and Management High School, Viana do Castelo, Portugal
sara.paiva@estg.ipvc.pt

Abstract. The paper presents an indoor navigation system, both mobile and web, applied to a university environment. The case study follows a research methodology. We describe the prototype development, routing algorithm and current achieved results.

Keywords: indoor navigation, educational mobile application.

1 Introduction

Almost all Universities and High Education Institutions buildings worldwide are huge enough for anyone new to easily get lost. As a teacher in a High Public Education Institution, the author quickly realized the big confusion it is to find spaces: too many corridors and sometimes not very accurate indications. That difficulty is mainly felt by new students in the beginning of each scholar year and also for anyone that needs to come to the Institution for meetings or any other purpose for the first time. Plenty of times sitting at their offices, teachers have someone knocking at the door and asking "Excuse me. Can you tell me where room X is please?" It is probably a reality in all Institutions.

New students are with no doubts the main targets; starting from day 1, for enrolments and from then on, for attending classes. This last point assumes more relevancy when students frequently arrive late for classes because they can´t find the classroom or laboratory. This aspect, for its importance, was subject of a questionnaire - that will be detailed in this paper – and served as a strong motivation for developing a mobile application for indoor navigation in school environments.

From the moment mobile applications aroused, navigation applications are a major field for developers, namely navigation inside buildings, known as *indoor navigation.* As with all types of navigation application, localization is necessary and this is the main difficulty with indoor applications, as they can´t make use GPS for that purpose. Different strategies must be used in this type of applications.

In this paper, we cross the fields of indoor navigation with the need to help students in their way within schools and present a prototype which is currently being used in a Portuguese High Education Institution. Other applications that cross these two fields exist, such as [1] or [2]. The presented solution differentiates on the

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 959–968.
DOI: 10.1007/978-3-642-36981-0_90

algorithm and strategy for indoor navigation, strongly restricted by the environment and reduced costs.

The rest of the paper is organized as follows: the next chapter presents the concept of indoor navigation, challenges and strategies. Section 3 presents the case study starting with the research methodology explanation and afterwards the detail of each stage of action research.

2 Indoor Navigation

Navigation has become a major field for development in what mobile applications are concerned. There are mainly two types of navigation: outdoor and indoor. The first type is probably the most common or at least known by the majority of people, and an example is car navigation systems. Outdoor navigation applications rely on GPS (Global Positioning System), a satellite navigation system, capable of offering to a given mobile device precise coordinates of its location. As easily understood, this is only available in places where communication with the satellite which usually does not work inside buildings. Indoor navigation uses, therefore, other strategies to allow navigation, which starts by knowing the user´s current position. Some of those strategies are: RFID [2], [3], Ultrasonic sensors [4], [5] or WiFi [6].

3 The Case Study

3.1 Research Methodology

The methodology we used in this study was action-research, as proposed by Olesen and Myers [7]. This type of research is characterized by some form of collaboration by researchers and practitioners of which should emerge some new knowledge important for both parts. The authors also refer, as an important aspect of the method, the analysis of the results it provides, even in case of failure. In order to implement the action research proposed by Olesen and Myers [7], the following steps should be followed:

- Diagnosis: identify the problem;
- Action Planning: determine the actions to solve the problem;
- Actions taken: choose and implement a given course of action;
- Evaluation: analyse and study the consequences of the course of action;
- Learning obtained: document the knowledge obtained during the project.

3.2 Development Stages

3.2.1 Diagnosis

Motivated mainly by the need to help students find spaces (classrooms, laboratories, administrative spaces, etc.) in a High Education Institution, we started by running a

questionnaire to understand the magnitude of the difficulty and also to be aware of the relation it has with students arriving late to classes. The questionnaire was made to 100 students and is composed of 5 questions:

1. How would you rate the ease with which he could locate classrooms and laboratories needed in your first month?
2. Have you ever been late to any class because you could not find the room?
3. If you answered yes to the previous question, indicate how often.
4. Do you find useful to implement a mobile application for orientation inside ESTG?
5. Do you have a mobile phone?

Regarding question number 1, more than 90% of the students felt some kind of difficulty in finding classrooms. Of those, more than 80% have already arrived late to a class because they couldn´t find the laboratory or classroom. Again, considering those who already arrived late to classes, more than 20% say to have arrived late more than 6 times and more than 50% answered at least 5 times.

Questioned on the usefulness of a mobile application that could help them locate spaces within the school, more than 80% find it useful or very useful. Finally, and because, it is a fundamental requirement, 99% of the students answered positively when asked if they had a mobile equipment. Fig. 1 shows the results for all questions of the questionnaire.

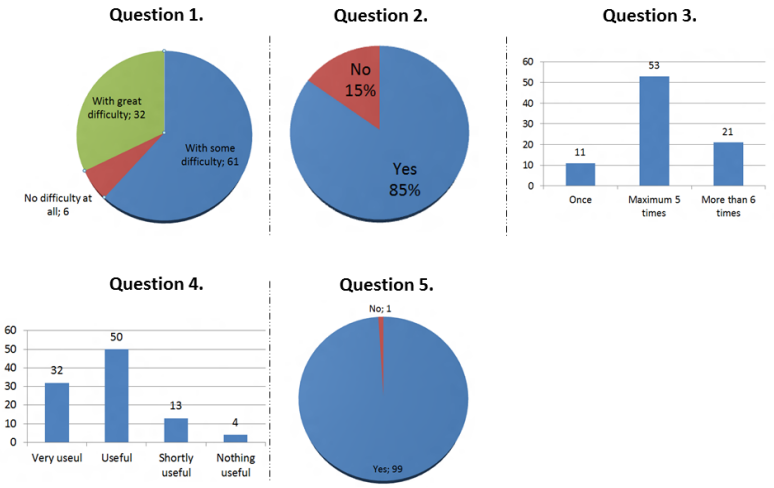


Fig. 1. Charts representing the answers obtained to the five questions of the questionnaire done to 100 students, regarding their opinion on the difficulty to find spaces within the school and the importance of a mobile application that helped them with that issue

The results obtained from the questionnaire allowed us to conclude such an application would be of great value to the students and could possibly reduce the number of times students arrive late to classes.

3.2.2 Action Planning

Having defined the purpose of the project, the following actions were planned:

- define the technology to use and the way to determine the user´s location.
- gather data about the spaces in school
- define the routing algorithm
- implement and deploy the prototype

3.2.3 Actions Taken

3.2.3.1 Determining User´s Location

A first prototype was developed using the Access Points (AP) placed along the school´s corridors. As a space can receive signal of more than one AP, a map/list was elaborated (Fig. 2) where for each space we should have the AP MAC address and the signal strength, for each AP that covers that space. After that, zones were created with spaces that have the same or a close value of signal strengths from the APs.

With this information, the strategy for determining a specific user location would be to capture the signals received in a specific place and map it to a zone which would be the starting point of the routing.

This approach had, however, to be put aside as APs frequently change location, either because they need replacement or because new equipment is added in order to improve the network coverage. Regardless of the reason, every time changes occur in APs location, all the values calculated in Fig. 2 had to be recalculated what would not be viable solution in a production scenario.

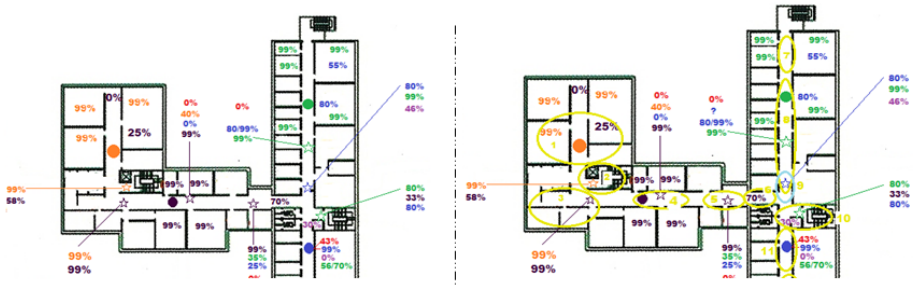


Fig. 2. Excerpt of the signal range of each AP. Three AP´s exist in the Figure (orange, green and blue). Each space is marked with the signal strength of the several Aps whose signal targets the space. The Figure on the right also has zones depending on signal strengths.

Another possible approach was to install RFID tags in several places but the costs to this solution are considerable and not viable to the school. Also, not all mobile devices can read RFID tags. With technology and costs limitations the decision was made to ask the user intervention to determine his location. For that, we defined some strategic *control points* (CP) (Fig. 3), which are easy to locate spots within the school.

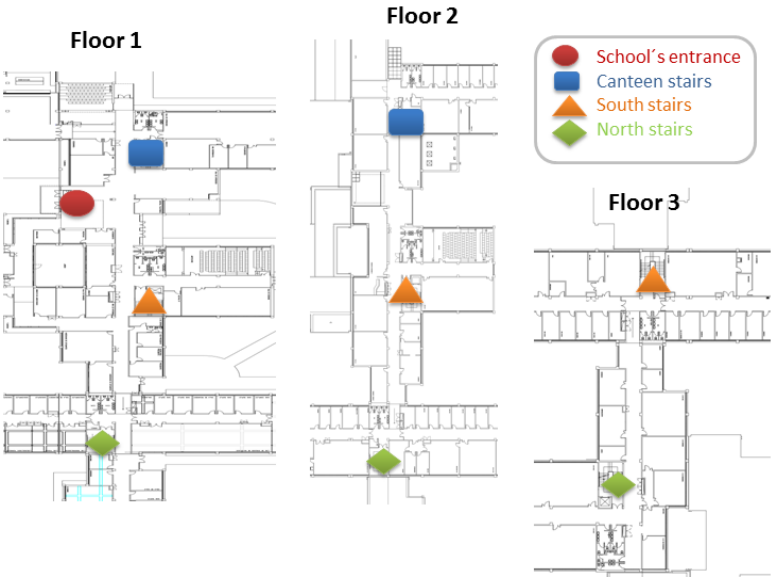


Fig. 3. Excerpt of the plants from the three floors of the schools and the defined control points that represent the possible sources of the routing algorithm, whenever a student asks for directions

They are either the reception or stairs. The red circle CP represents the school´s entrance and an easy spot for anyone to find. The blue rectangle CPs represents the canteen stairs that connect floors 1 and 2. The orange triangles CPs represent the south stairs that connect floors 1, 2 and 3. Finally, the green diamonds CPs represent the south stairs that also connect floors 1, 2 and 3.

Therefore, when choosing its current position, the student will have to choose among a CP, depending on the floor he is in. Given the restrictions we faced, we believe this to be a completely viable solution as those are easy to find places.

3.2.3.2 The Routing Algorithm and the Spaces Database

Considering the way current location is going to be obtained, we defined some assumptions for the routing process. First thing is that each space should be connected to all CPs in its visual range (Fig. 4). This is important because a student can go to

Space 3, in Fig. 4, either from North (Orange triangle) or from South (Green Diamond). If, for instance, a student wants to go to Space 2 from South, as there is no direct route (CP is not at visual range of the space), two routes are calculated: first, go from the source (South CP) to the CP that serves the destination (orange CP serves space 2); second: having reached the CP that serves the destination, there must be a direct route from one place to the other. Considering this algorithm, routes must exist between each CP. A route between a space and a CP or between two CPs is a sequence of pictures and a message.

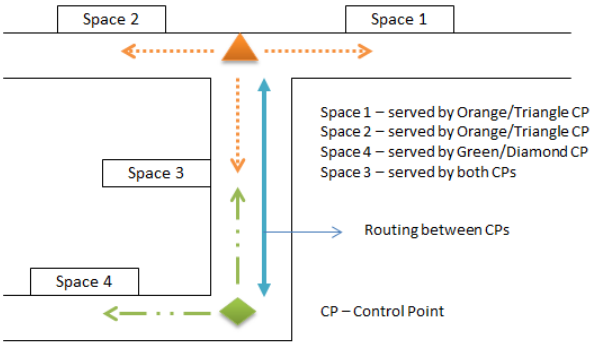


Fig. 4. For the routing algorithm, one assumption made was that each space should be connected to all control points in its visual range

When source and destination are in different floors, the algorithm has some more steps, as can be seen in the complete routing algorithm in Fig. 5.

3.2.3.3 The Prototypes

The prototype was developed in the Android OS and made available in Google Play [8], as it is the most common among students for its lower price, when compared to iOS or Windows Phone. A web platform was also developed, with the same algorithm and functionalities, for those who don´t have an android equipment or for those who want to access the system in a desktop computer, at home. The functionalities of both systems are pretty much the same. They allow the student to:

- pick a destination (1st picture in Fig. 6 and in Fig. 7) where filters can be done by floor, space type (laboratory, classroom, administrative spaces, etc. - 3rd picture in Fig. 6 and 1st picture in Fig. 7).
- pick a source (2nd picture in Fig. 6 and in Fig. 7) after choosing the floor where they are in. A picture of the control point´s location is shown so the student can assure that is his current location.
- After choosing source and destination, a sequence of picture/message is shown (4th picture in Fig. 6 and 3rd in Fig. 7).

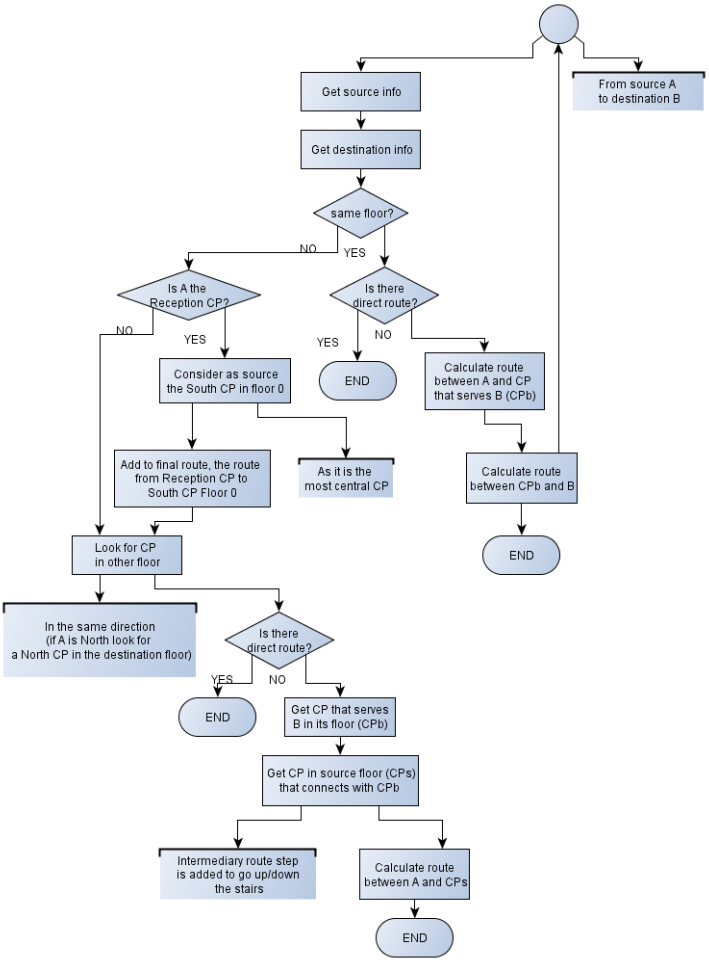


Fig. 5. Routing algorithm to help students find spaces within their school

Fig. 6. Android Mobile Application

The web platform [9] has, additionally, some search functionalities that allow students to go directly to a teacher´s office (shown in 1st picture in Fig. 7).

Fig. 7. Web Platform

3.2.4 Evaluation

Regarding the number of users, Fig. 8 shows the number of users from both versions that used the application in the first month of its publishing (7th September). A big increase is noticed around 25th September, when classes actually started.

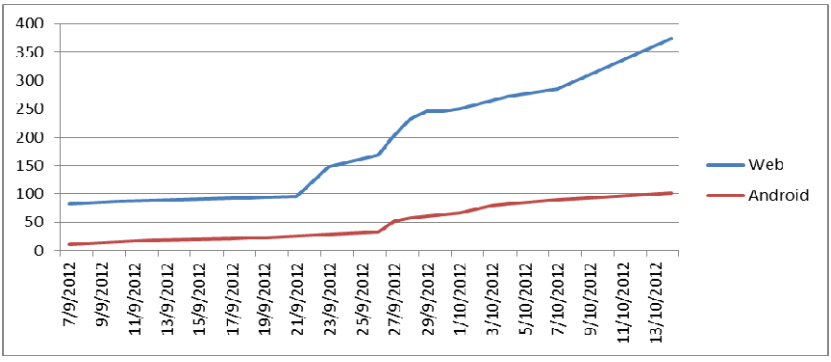


Fig. 8. Numbers of users in the first month of the prototype launch, from both mobile and web versions

These numbers are not yet totally satisfactory as we aim to target more students. More commitment on advertising the application in school entrance will be made in the next version launch.

We also made a questionnaire from where we concluded that there was a greater use of the web application compared to the android application. This cab ne explained for only 32% of the respondents have an android device (windows phone follows with 23%, Symbian and others with 18% and 9% with iOS). When asked what they would

change in the application, the functionalities "change how current location is obtained" and "routing method to destination" were chosen by 55%. Other suggestions were made such as use a 3D map of the university, use GPS (not viable) and develop the same application for different mobile OS.

3.2.5 Learning Obtained

From the evaluation, and for a first prototype, we think the results are positive, also judging for the feedback we had on the corridors. Yet we want to target more students and we learned more efforts have to be done in advertising in the launch of the next application version. One of the aspects we wanted to fight was the fact that students arrive late to classes. Unfortunately, that cannot yet be measured efficiently as few students are making use of the application. We also believe that is a questionnaire that should be done after the application is being used for at least one scholar year.

4 Conclusions and Future Work

This paper presents a mobile application for indoor navigation in a school environment. We presented our case study using a research methodology by Olesen and Myers, with diagnosis, action planning, actions taken, evaluation and learning obtained. The diagnosis was based on a questionnaire that proved us right in the need for such an application, namely because not knowing the location of classrooms and laboratories is one reason for students arriving late for schools. We planned and executed the following actions: define the technology to use and the way to determine the user´s location, gather data about the spaces in school, define the routing algorithm and implement/deploy the prototype. We concluded that the first version of this prototype has already helped a few students but a bigger audience should be reached and advertisement of the application should be reinforced for the next version. Regarding future work, several features are already taking place. First, use QR Codes in each lab/classroom and use it to determine where the user is, replacing the control points (all stairs have pretty much the same look). Second, we will incorporate videos and maps to show the route from source to destination, beside photos, as well as voice direction messages. Third, we will incorporate a search functionality based on semantic search [10] that will allow users to find, for example, computer labs without knowing their exact names. Also, we will include an option for people with disability that should always be directed to the elevator whenever source and destination are in different floors. Incorporating the functionality for the student to have its classes schedule in the mobile device can help him not to memorize where is the next class as the program, depending on the date/time will automatically take him there.

Acknowledgments. The author would like to thank students Miguel Maciel and Mickael Abreu for their help on the first stage of the project, namely on the study on how AP´s could be used to determine user´s current location and also to Tiago Monteiro for the help with the questionnaire.

References

[1] World's first university with indoor navigation app, NTNU - Norwegian University of Science and Technology, http://www.ntnu.edu/news/campus-guide
[2] NYU WayFarer, N. Y. University, http://www.behance.net/gallery/NYU-Wayfarer/4528801
[3] Chumkamon, S., Tuvaphanthaphiphat, P., Keeratiwintakorn, P.: A blind navigation system using RFID for indoor environments. In: 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, pp. 765–768 (2008)
[4] Kim, H.S., Choi, J.S.: Advanced indoor localization using ultrasonic sensor and digital compass. In: International Conference on Control, Automation and Systems, pp. 223–226 (2008)
[5] Nam, Y.J., Park, Y.G.: Efficient Indoor Localization and Navigation with a Combination of Ultrasonic and CSS-based IEEE 802.15. 4a. In: 4th International Conference on Ubiquitous Information Technologies & Applications, pp. 1–6 (2009)
[6] Paul, A., Wan, E.: Wi-Fi based indoor localization and tracking using sigma-point Kalman filtering methods. In: IEEE/ION Position, Location and Navigation Symposium (2008)
[7] Olesen, K., Myers, M.D.: Trying to improve communication and collaboration with information technology: An action research project which failed. Information Technology People 12(4), 317–332 (1999)
[8] Where am I in ESTG/IPVC?, Google Play Application, https://play.google.com/store/apps/details?id=spaiva.estg.whereami&feature=search_result#?t=W251bGwsMSwxLDEsInNwYWl2YS5lc3RnLndoZXJlYW1pIl0
[9] Web Platform 'Where am I in ESTG/IPVC?' http://apps.estg.ipvc.pt/orientacao/default.aspx
[10] Paiva, S., Ramos-Cabrer, M., Gil-Solla, A.: GSSP – A Generic Semantic Search Platform. In: Proceedings of the Conference on Enterprise Information Systems, vol. 5, pp. 388–396 (2012)

Towards a Security Solution for Mobile Agents

Djamel Eddine Menacer[1], Habiba Drias[2], and Christophe Sibertin-Blanc[3]

[1] National Computer Science School of Algiers (ESI)
Algiers, Algeria
[2] Houari Boumedienne Sciences and Technologies University of Algiers (USTHB)
Algiers, Algeria
[3] Université Toulouse 1 Sciences Sociales
Toulouse, France
d_menacer@esi.dz, hdrias@usthb.dz, sibertin@irit.fr

Abstract. The current distributed systems are mainly based on the client – server model. The growing use of the Internet and the development of information and communication technologies has yielded more needs in terms of performance, reliability and security of information. By its synchronous nature, the client – server model no longer seems able to meet these new demands. Several new approaches have been proposed, for example the mobility of the code, involving the emergence of the mobile agents' paradigm (MA). This paradigm certainly holds great promise, though questions of security arise. In this paper, we propose an extension of the MA model called Seller – Buyer model (SB). This extension is made by making the server component mobile and using market mechanisms between the client agent (the buyer) and the server agent (the seller). While most of works in the area propose asymmetric security solutions, the SB model proposes a symmetric security approach for the MA. The SB model will give the developers a way to use, safely and efficiently, MA to build distributed applications on large-scale networks such as the Internet.

Keywords: Mobile Agent (MA), Seller–Buyer model, Distributed systems, Security.

1 Introduction

The importance of the quality of service (QoS) in distributed applications (non functional aspect) is becoming critical. The quality of service includes (Coulouris et al., 2001): the performance, the security and the safety of functioning (or reliability). While most of the distributed systems are based on the client – server model, new paradigms arise to deal with the growing QoS requirements. Mobile agents (MA) are a promising paradigm. The technical aspect of the MA is the main argument for their use (Lange and Oshima, 1999; Rouvrais, 2001). However, the security problem blocks their development (Harrison et al, 1995; Papaioannou, 1999).

A MA needs a specific execution environment on each of the sites that constitute its itinerary. The execution environment is provided by a *MA platform*. A lot of MA platforms have been developed around the world, such as the Aglets from IBM

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 969–979.
DOI: 10.1007/978-3-642-36981-0_91

(Lange, 1998; Lange and Oshima, 1999; ASDK) or the Mole from University of Stuttgart (Baumann et al., 2002). Many efforts are made to provide MA systems that could be more accessible by the industry. FIPA (FIPA, 2009) and MASIF (Milojicic et al., 1998) are the two most known standards. Few MA platforms provide security features, and none of the existing FIPA and MASIF MA platforms proposes the protection of the agents.

Most researchers in the area are seeking a better solution, and there is no general methodology suggested to protect agents. In the meantime, developers of MA systems, who are non security experts, have developed their own methodologies according to their own needs.

The remainder of this paper is organized as follows. Section 2 describes the security issues for the MA and presents some existing security solutions. A new MA model is described in Section 3, and its security improvement is discussed. In section 4, we discuss the security evaluation of the SB model. Finally, section 7 concludes the whole paper.

2 Overview of Existing Security Solutions for Mobile Agents Systems

2.1 Mobile Agents' Security Issues

On large-scale public networks such as the Internet, security is an essential concern. In the case of MA, the hosts receive the agents and give them an execution environment. The agents can then use the local resources of the host. The MA security problem is then *dual*: we have to secure not only the visited hosts but also the visiting agents. MA security issues can be categorized into four streams (Gulyás et al., 2001; Jansen et al., 1999):

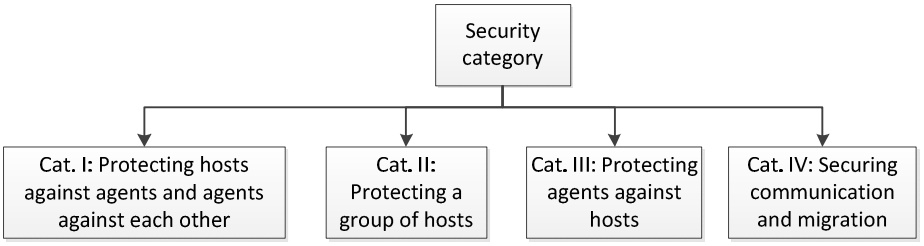


Fig. 1. Security Categories

The figure 2 shows, in detail, the different threats in each category. Figure 2 shows also that category I and category IV have good solutions already used in traditional systems. However, category III still does no complete solution. Indeed, the main unsolved security problem concerns the protection of visiting agents against a malicious host: a receiving agent platform can easily isolate and capture an agent and attack it by extracting information, corrupting or modifying its code or state, denying

requested services, or simply reinitializing or terminating it (Jansen et al., 1999). Moreover, an agent is very dependent on the agent platform and may be corrupted by the platform responding falsely to requests for information or service, or delaying the agent until its task is no longer relevant.

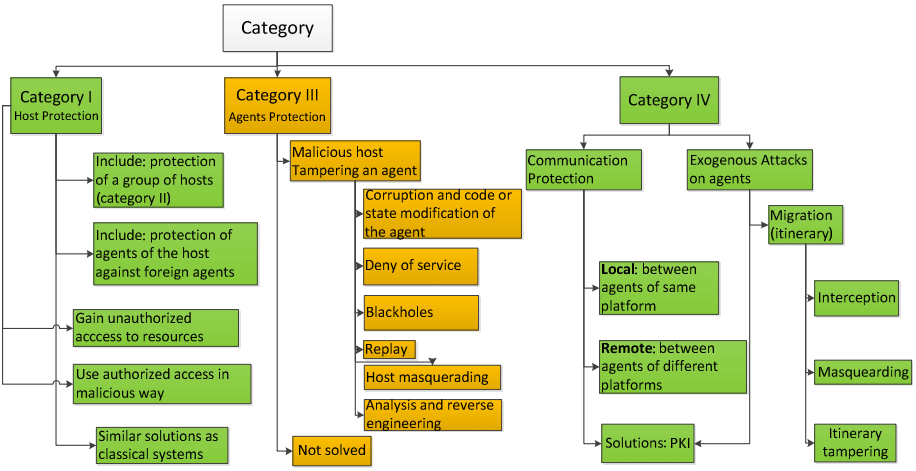


Fig. 2. Security Category analysis

2.2 Security Solutions for Category III

There are a number of solutions proposed to protect agents against malicious hosts which can be summarized into three streams (Tschudin, 1999; Sander and Tschudin, 1998), as shown in figure 3:

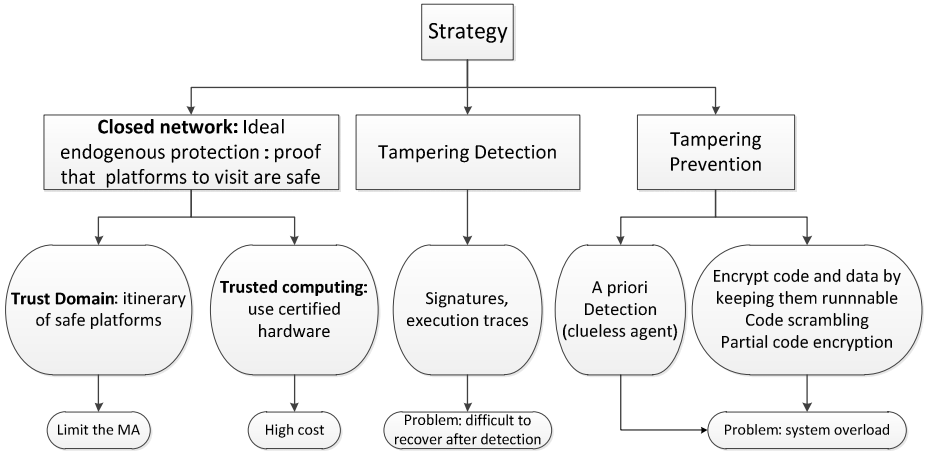


Fig. 3. Most known solutions for category III

2.3 Examples of Works on Security

Recently, some works focus on the protection of agents (category III) but propose solutions for just specific threats. We can summarize these threats in two lines: risk on the *agent's code* and risk on the *agent's itinerary*. For the first type of threat, the authors in (Garrigues et al., 2008b) propose a solution for protecting MA from external replay attacks. This kind of attacks is performed by malicious platforms by dispatching an agent several times to a remote host, thus making it to re-execute parts of its itinerary. The protection of the agents' itinerary by encrypting the itinerary prevents this attack but decreases the system performance to a great extent. Another solution in (Chan et al., 2000) consists in leaving the itinerary in clear but merging it with an encrypted chain at each visited site, and a recent solution in (Garrigues et al., 2008a) consists in a cryptographic scheme for shielding dynamic itineraries from tampering, impersonation and disclosure. For the second type of threat, the authors in (Pechoucek et al., 2005) propose a MA platform called SECMAP and its security infrastructure. SECMAP presents abstractions which ensure the protection of agents and system components through a *shielded agent model*. This solution has a cost: on each execution of an agent, data must be decrypted in memory. The solution increases also the processor payload in the visited agents' platforms. The logging process, though useful, should make the whole system very heavy. Moreover, the presence of untrusted nodes is not discussed.

Finally, a new approach known as *trusted computing* introduced by (Arbaugh et al., 1997) has received interest in (Muñoz et al., 2009). In a Trusted Computing scenario, trusted applications run *exclusively* on top of trusted supporting software. This approach provides, therefore, a secure environment for agent execution (category III). However, this solution is generally expensive and spreading trusted environments on Internet is not possible.

2.4 Discussion

The main reason that blocks MA technology relies on security. Indeed, the protection of agents against hosts, known as endogenous attacks, remains unsolved. However, we can notice that while security issues of mobile agents are *symmetric* (they concern the protection of both the agents and the hosts), in contrast to this, the research efforts concerning these issues seem to be *asymmetric* (focus on host security).

Another important fact is that all proposed solutions rely on **conventional techniques** (techniques already used in traditional systems), while MA belongs to non conventional systems (mobile systems).

We then feel that the MA technology needs non conventional techniques to be fully used.

3 The Seller – Buyer Model

In the following, we will define a new approach for securing MA. To do this, we propose to extend the classical mobile agent model (MA) towards a more suitable model, the Seller-Buyer model (SB) (Menacer et al., 2011).

The SB model allows reducing the mobile agent's security problem. This is done by preventing an agent from migrating *directly* to a host and a host from receiving agents. This is possible if we dissociate the rendering of services and the hosting of visiting agents; then a MA representing the *client* meets, on *intermediate sites*, a service *provider* representative. As the client agent, the server agent process becomes mobile.

The client agents are *delegated* by clients to *buy* a service (buyer agents) and the server agents are *delegated* by remote services to *sell* a service (seller agents). MA may migrate *only* on special meeting sites: the *market places* or *MP* (figure 4).

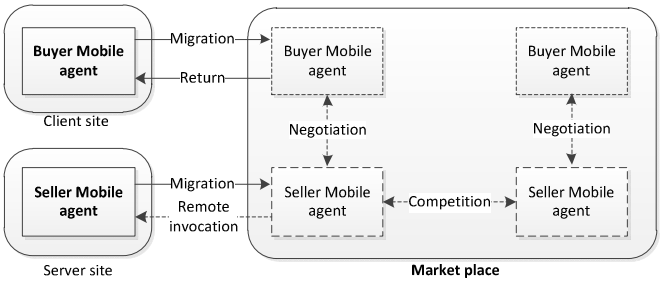


Fig. 4. The seller-buyer model

Seller agents should only carry the **minimum resources** from their providers. In order to complete the service rendering, seller agents can run remote invocations with their providers (figure 4). This is the key of **service mobility** in SB.

3.1 SMA in SB Model

In the SB model, we adapt market mechanisms to distributed systems. The buyer agents perform *negotiation* with seller agents in the MP and the seller agents are in *competition*. The negotiation is based upon a price p that can be the QoS. Facilitator agents are used to ease the migration of MA. Figure 5 shows the different interactions of the *SB multi-agents system*.

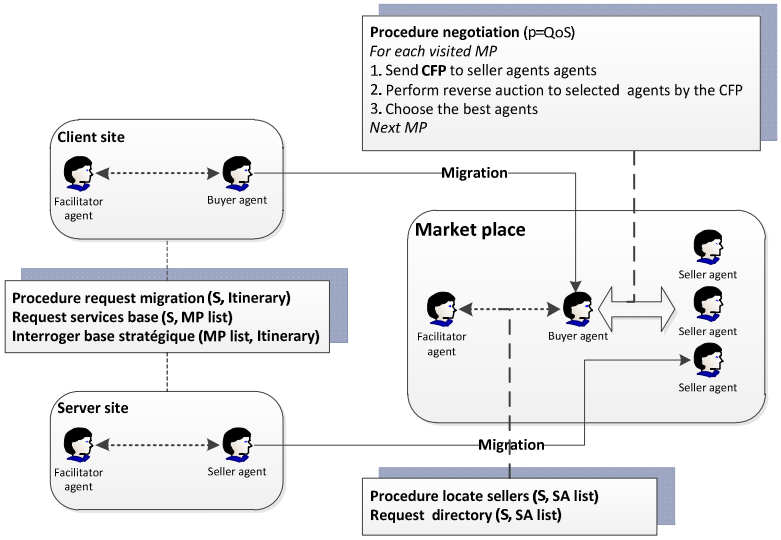


Fig. 5. The SB multi-agents system

3.2 SB Security Model

The SB model reduces the endogenous attacks against the agents by a symmetric approach: *only trusted agents (cat. I) can migrate on trusted MP (cat. III)*. The SB security model relies on two levels (see figure 6):

Level 1 (Top Level): Non Conventional Level. In this level, we operate a translation from category III issues to category I issues. This is done as follows:

- **Service delegation:** delegate services as well as requests by means of seller agents
- **Service mobility:** Move server components as well as client components my making mobile the seller agents
- **Trust:** the interactions between buyer agents and seller agents must take place only in the market places (MP)
- **Mixed strategy:** If the agents cannot move on MP (must interact with non-trusted hosts), then they must use only remote invocation

Level 2 (Bottom Level): Conventional Level. In this level, we use traditional techniques such as trust and PKI. This level is used under the top level and implement solutions for the category I issues.

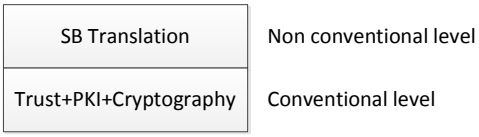


Fig. 6. The SB security model

3.3 SB Security Architecture

We can summarize the SB security model in the architecture shown in figure 7.

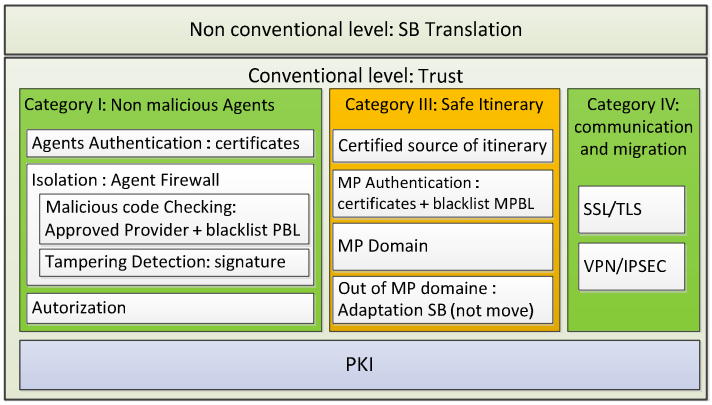


Fig. 7. The SB security architecture

3.4 Recapitulation

The SB model allows reducing the security issues in the MA model, by removing the category III issues (endogenous attacks) as shown in the figure 8.

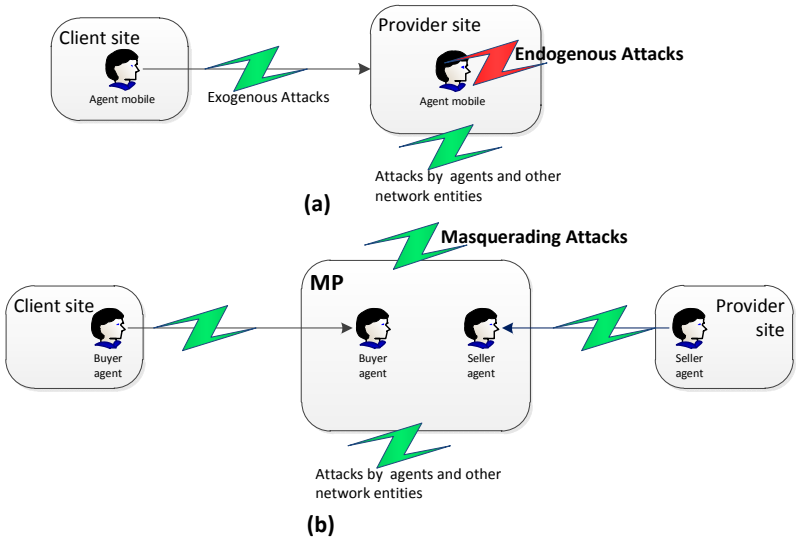


Fig. 8. Improvement of the MA (a) security by the SB model (b)

4 Evaluation of the SB model

4.1 Threat Analysis

Threats. There are two main threats in SB model:

1. **Threats against agents :** there are three threats :
   a. Masquerading of places: by IP spoofing or hijacking
   b. Legitimate but contaminated market places
   c. Tampered itinerary
2. **Threats against market places :** there are two threats :
   a. Malicious providers giving malicious code on seller agents
   b. MArket place identity usurpation

Threats Analysis. We can notice the *symmetry* of threats as follows (see figure 9):

- 1b (cat 3) ⇔ 2a (cat 1)
- 1a (cat 3) ⇔ 2b (cat 1)
- 1c (cat 3) ⇒ presence of malicious place in the itinerary ⇒ 1a or 1b

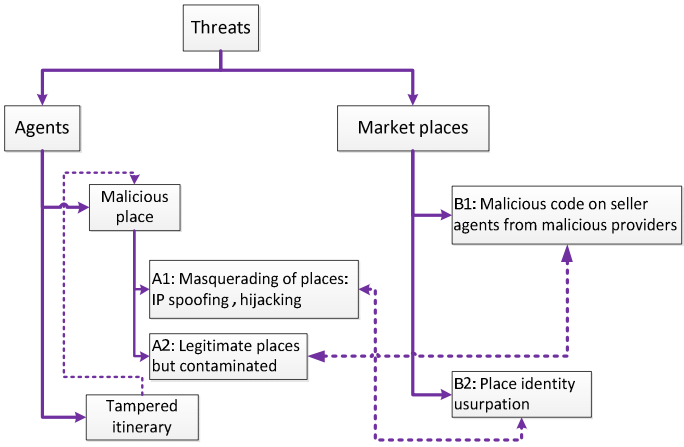


Fig. 9. The SB security threat symmetry

4.2 Countermeasure Evaluation

Protection. We can protect both agents and places by category I means as follows:

1. **Agents protection**
   a. Fight against masquerading of market places
   b. Fight against malicious code (Service code)
2. **Places protection**
   a. Fight against malicious agents
   b. Fight against masquerading of market places

Protection Analysis. We can notice the *symmetry* of protections as follows (see figure 10):

- 1a (cat 3) ⇔ 2b (cat 1)
- 1b (cat 3) ⇔ 2a (cat 1)

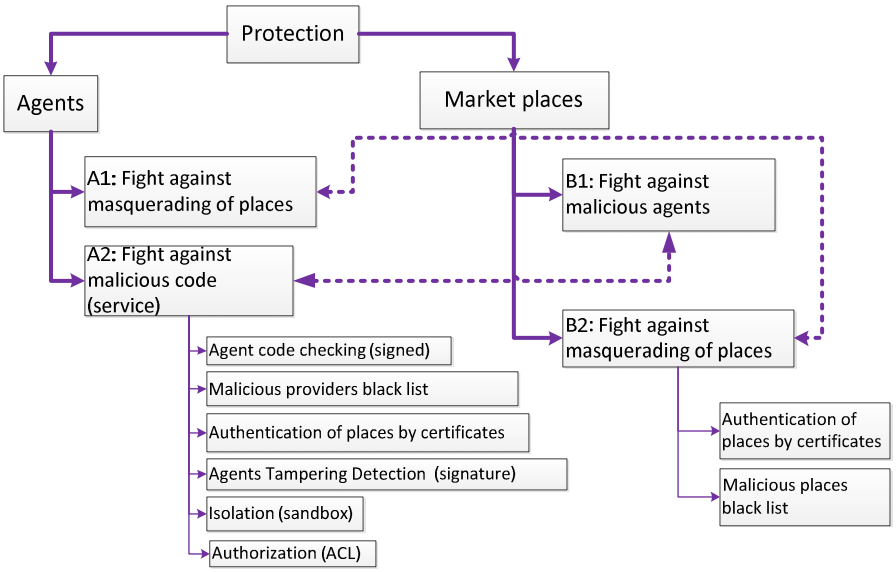


Fig. 10. The SB security solution symmetry

Protection Mechanisms. Here are some mechanisms that can be used for the protection:

1. **1a and 2b.** Each agent can use PKI and certificates as follows:
   - Authentication of places by certificates
   - Using malicious places blacklists (MPBL) for each suspect place.
2. **1b and 2a.** In each place, an *agent firewall* acts as follows:
   - Perform seller agent code checking (signed code)
   - Using malicious providers blacklists (PBL) for each suspect provider
   - Authentication of visiting agents by certificates
   - Using certificates for agents tampering detection (signatures)
   - Isolation (using java sandbox for example) of agents before running them
   - Giving appropriate **authorization** for resources (using ACL) to safe agents

4.3 Impact on the Performance

The SB model uses *mixed strategy*: seller agents can make remote invocations with their providers to render the full service while buyer agents can decide whether they

move or interact remotely. Moreover, the SB model allows reducing the number of migration of the MA. The SB security model does not affect the performance and can even improve the MA model as shown in (Menacer et al., 2011).

4.4 Discussion

We can notice that, in SB model, we have reduced the number of threats to just two main threats:

1. **Masquerading and usurpation of the market places**, that can be considered as well as category I and category III issues.
2. **Attacking a market place by malicious seller agents**, which is category I issue. This attack can contaminate the place and makes it malicious place, and can be then considered as category III issue.

A *tampered itinerary*, which is category III issue, is also caused by both those threats. Since both threats are category I issue, the security issue in SB model is easier than in MA model.

5 Conclusion and Future Works

In this paper, we have proposed an extended MA model in order to reduce the security problem of the MA: the SB model. The originality of our approach resides in making the server component mobile and using market mechanisms between the client agent (the buyer) and the server agent (the seller). While most of MA systems provide asymmetric security solution, the SB model proposes a symmetric security model by making a translation from the category III to the category I security issues (known as non conventional security level) and using PKI mechanisms under this translation (known as conventional level). The SB model is also more efficient than the MA model since it uses a mixed strategy and reduces the number of migrations.

Through this work, we have given an answer to the question "how to use safely MA in large-scale networks environments such as the Internet?" The proposed model uses market mechanisms that are not detailed in this work. We feel that these mechanisms can optimize the QoS in a MA system. We then project to detail these market mechanisms through a framework based upon the SB model.

References

1. Arbaugh, B., Farber, D., Smith, J.: A secure and reliable bootstrap architecture. Paper presented at the IEEE Symposium on Security and Privacy (1997)
2. ASDK. IBM Aglets Software Development Kit Homepage, `http://www.trl.ibm.co.jp/aglets` (accessed in July 2010)

3. Chan, A.H.W., Wong, C.K.M., Wong, T.Y., Lyu, M.R.: Securing Mobile Agents for Electronic Commerce: an Experiment. In: Proceedings of the IFIP TC11 Fifteenth Annual Working Conference on Information Security for Global Information Infrastructures. Kluwer, B.V. Deventer (2000)
4. Baumann, J., Hohl, F., Rothernel, K., Straβer, M., Theilmann, W.: Mole: A mobile agent system. Journal of Software-Practice and Experience (SPE) 32(6), 575–603 (2002)
5. FIPA, Official site of the Foundation for Intelligent Physical Agents (2009), Obtained through the Internet: `http://www.fipa.org` (accessed June 15, 2009)
6. Garrigues, C., Robles, S., Borrell, J.: Securing dynamic itineraries for mobile agent applications. Journal of Network & Computer Applications 31(4), 487–508 (2008)
7. Garrigues, C., Midas, N., Buchanan, W., Robles, S., Borrell, J.: Protecting mobile agents from external replay attacks. Journal of System Software 82(2), 197–206 (2009)
8. Gulyás, L., Kovács, L., Micsik, A., Pataki, B., Zsámboki, I.: An Overview of Mobile Software Systems. MTA SZTAKI Technical Report. Department of Distributed Systems, MTA SZTAKI, Computer and Automation Research Institute of the Hungarian Academy of Sciences (2001)
9. Harrison, C.G., Chess, D.M., Kershenbaum, A.: Mobile Agents: Are they a good idea? IBM T.J. Watson Research Report. RC 19887 (1995)
10. Jansen, W., Mell, P., Marks, D., Karygiannis, T.: Mobile agent security. National Institute of Standards and Technology (NIST) Special Publication – 800-19, Gaithersburg, MD 20899, USA (October 1999)
11. Lange, D.B., Oshima, M.: Seven Good Reasons for Mobile Agents. Communications of the ACM 42(3), 88–89 (1999)
12. Menacer, D.E., Drias, H., Sibertin-Blanc, C.: The MP Architecture: towards a secure framework for mobile agents. International Journal of Agent-Oriented Software Engineering 4(4), 390–414 (2011)
13. Milojicic, D., Breugst, M., Busse, I., Campbell, J., Covaci, S., Friedman, B., Kosaka, K., Lange, D., Ono, K., Oshima, M., Tham, C., Virdhagriswaran, S., White, J.: MASIF: The OMG Mobile Agent System Interoperability Facility. In: Rothermel, K., Hohl, F. (eds.) MA 1998. LNCS, vol. 1477, pp. 50–67. Springer, Heidelberg (1998)
14. Papaioannou, T.: Mobile Agents: Are They Useful for Establishing a Virtual Presence in Space? In: Agents with Adjustable Autonomy Symposium, Part of the AAAI 1999. Spring Symposium Series (1999)
15. Ugurlu, S., Erdoğan, N.: SECMAP: A Secure Mobile Agent Platform. In: Pěchouček, M., Petta, P., Varga, L.Z. (eds.) CEEMAS 2005. LNCS (LNAI), vol. 3690, pp. 102–111. Springer, Heidelberg (2005)
16. Rouvrais, S.: Construction de services distribués: une approche à base d'agents mobiles. TSI-HERMES / RSTI – TSI 21(7), 985–1007 (2002)
17. Sander, T., Tschudin, C.F.: Protecting Mobile Agents Against Malicious Hosts. In: Vigna, G. (ed.) Mobile Agents and Security. LNCS, vol. 1419, pp. 40–60. Springer, Heidelberg (1998)
18. Tschudin, C.: Mobile Agent Security. In: Intelligent Information Agent: Agent Based Information Discovery and Management in the Internet, pp. 431–446 (1999)

X-Band Radar Sensor for the Landslide Risk Mitigation

Sandra Costanzo, Giuseppe Di Massa, and Marco Salzano

DIMES – University of Calabria
87036 Rende (CS), Italy
costanzo@deis.unical.it

Abstract. A low cost Stepped Frequency Continuous Wave sensor is proposed in this work as radar module for the long range monitoring of environmental scenarios subject to landslide risk. A radar configuration based on the adoption of a variable scanning antenna array is proposed for the accurate detection of land movements. Furthermore, a X-band radar module prototype is realized and tested to demonstrate the high resolution capabilities of the proposed configuration.

Keywords: Radar, sensor, early warning, environmental risk.

1 Introduction

The adoption of low-cost and compact radar systems with high resolution capabilities can be strongly helpful in the framework of risk issues management, with a special focus on the monitoring, early warning and mitigation of landslide risk, where the accurate detection of the displacements of large areas is required. When considering standard real aperture radars, limited resolution capabilities, typically of the order of tens meters at high microwave frequencies, can be achieved. To further improve the resolution, synthetic aperture radar techniques and/or interferometric methods are usually adopted, but at the expense of increased signal processing complexity. Most of commercial radar modules work in the millimeter range of frequency for automotive applications. They don't need a high transmitted power, as operating at short distances. As a consequence of this, they are characterized by a limited frequency bandwidth which precludes a higher resolution in distance and does not allow velocity measurements. A simple and low cost solution is presented in this work, by proposing the use of an own customized Stepped Frequency Continuous Wave (SFCW) radar module, able to achieve range resolution of the order of centimeters. Furthermore, a radar configuration based on the use of a variable scanning antenna array [1], [2], [3], [4] is specifically proposed to face the problem of landslides monitoring, with the ability to dynamically change and address the antenna focus beam on a specified sub-area revealing some kind of anomalous displacement. To demonstrate the claimed features of the proposed approach, a X-band radar module prototype is realized and some preliminary indoor and outdoor tests are discussed.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 981–987.
DOI: 10.1007/978-3-642-36981-0_92

2 Functional Principle of Stepped Frequency Continuous Wave Sensors

The front end module adopted in this work is a SFCW radar [5], where the signal frequency is time swept within a set of discrete frequency points. The RF signal is radiated and reflected against different objects and the mixing transmitted and reflected signals give a waveform described by the equation:

$$s(n) = \cos\left(2\pi \frac{2d}{v} \frac{n}{N} B\right) \tag{1}$$

In the above expression, the frequency is proportional to the round-trip electrical distance traveled by the electromagnetic wave [6]. The parameter n is the index of the n-th frequency point, d denotes the distance of the target, v is the wave propagation velocity, B is the signal bandwidth and N is the total number of sweep points. The range resolution ΔR is related to the sweep bandwidth [6] by the expression:

$$\Delta R = \frac{v}{2B} \tag{2}$$

and the frequency spacing between two points is given by:

$$\Delta f = \frac{v}{4d_{max}}. \tag{3}$$

Hence, according to equations (2-3), by properly choosing the bandwidth and the maximum target distance, it is possible to compute the radar's resolving power. Performing a Fourier analysis of the signal (1), according to equation (2) it is possible to explore the signal spectrum coming from the sensor to localize the target distance [7].

3 X-Band Radar Configuration

The proposed X-band radar configuration is illustrated in Figure 1. In particular, the adopted X-band SFCW front end module is connected to an external circulator, in order to separate the transmission and the receiving paths, where proper amplifiers are inserted. A wide angle antenna is adopted to illuminate the scenario under analysis, while a variable scanning antenna array, based on the use of a Butler matrix, is considered in the receiving path. This latter feature can provide significant performances for the specific application of landslide monitoring. As a matter of fact, the idea is to first scan the investigated area with a large beam, but to give the opportunity to focus the antenna beam on a specific sub-area revealing some "anomalies". In order to satisfy this advantageous requirement, a specific synthesis procedure [8] for the design of the antenna array will be applied.

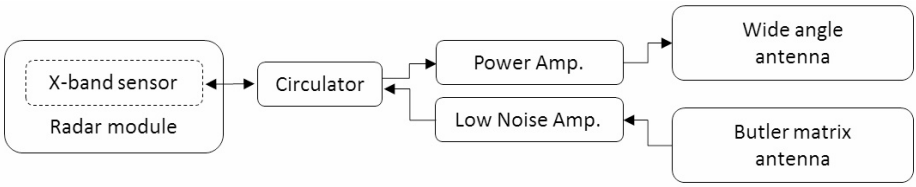


Fig. 1. Final configuration scheme of X-band sensor transceiver chain

The X-band sensor is controlled by adopting an on purpose designed electronic circuit. The control system illustrated in Figure 2 consists of four main sections:

1. The X-band sensor;
2. The microcontroller endowed with auxiliary protection circuitry;
3. The active filter;
4. The Analog to Digital Converter (ADC).

The circuit receives an external 12V supply voltage and produces the required 5V supply voltage for digital devices. A ATMEL ATmega1284P low-power 8-bit microcontroller is used to configure the X-band sensor and synchronize its output with the ADC. Finally, an active high pass filter is used to clean the output signal spectrum of X-band sensor from the low frequency harmonics produced by close reflections.

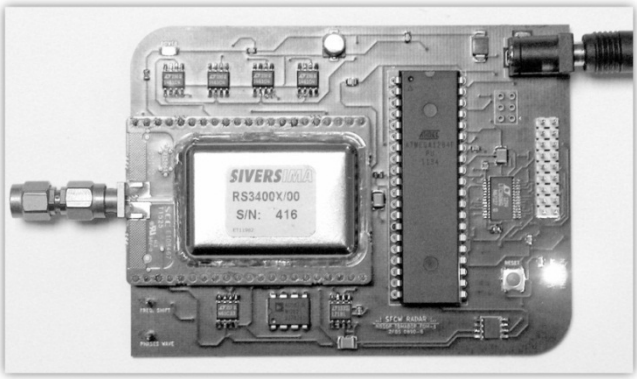


Fig. 2. Photograph of the radar module

However, the main considered specifications of the module are the frequency change in the signal generation and the maximum number of acquirable samples, where a minimum step of 11718.75 Hz and a maximum value of 1501 samples are respectively allowed by the demo board of the X-band sensor provided by the

manufacturer. The frequency is changed by tuning the integrated PLL through its external digital interface. Each frequency value is calculated and coded in a 32 bit digital Qword by the algorithm stored in the microcontroller, transferred by using the SPI and confirmed with a latch pulse over the LE pin of the sensor.

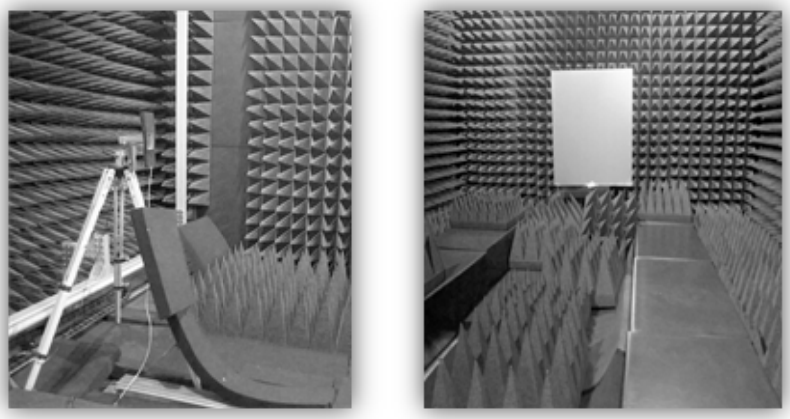
After a time equal to 50 μs, the microcontroller sends another pulse over the R/Cn pin of the ADC, by starting the analog-to-digital conversion of the signal coming from the high pass filter placed at the output of the sensor. At present, the module provides to the sensor a frequency step of 23437,5 Hz in order to sweep a bandwidth of 1.5 GHz. According to equations (2-3) and the Shannon' theorem, the capabilities of such configuration makes possible to extend the maximum detectable distance from 75 m (1500 samples) to 3.2 km (64000 samples).

4 Preliminary Tests on X-Band Radar

To highlight the resolution features of the adopted radar sensors, various preliminary tests are performed by adopting the standard demo board provided by the manufacturer of the radar transceiver [5]. Firstly, indoor tests are performed into the anechoic chamber of Microwave Laboratory at University of Calabria, and outdoor measurements are successfully made to estimate the ability of the radar sensor to detect multiple distant targets. For both tests, a 15 dB X-band horn antenna is adopted, without any external amplifying system, as the assumed sensor equips an integrated circulator that allows the transmission and the reception of the signal through a single antenna.

4.1 Indoor Test: Anechoic Chamber

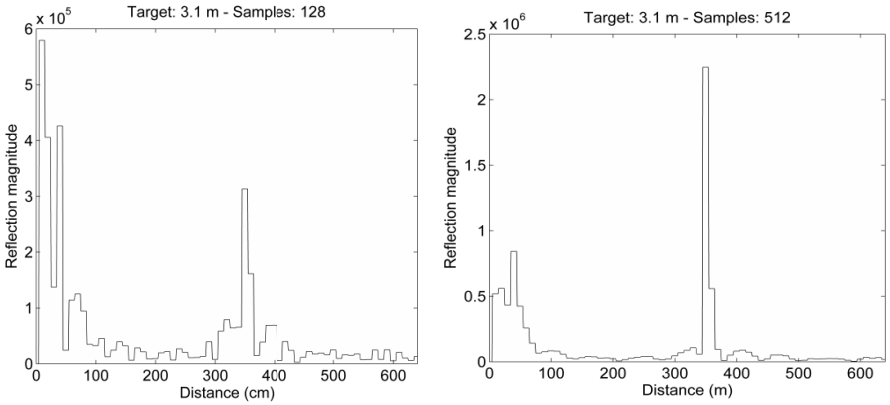
The first tests are performed into the anechoic chamber of Microwave Laboratory at University of Calabria, a shielded environment free of reflective obstructions and high external noise. The test setup (Fig. 3) consists of a ridged horn antenna with an offset angle of 15°, placed at an height of 1.35 m, a X-band sensor connected to the antenna and a metal screen adopted as known reference target. The sensor is mounted on the demo board provided by the manufacturer, thus limiting the maximum number of acquisition samples to 1501. The adopted sensor can sweep from 9.25 GHz to 10.75 GHz, producing a low power signal that ranges from -5 dBm to 5 dBm. The screen is placed at different distances, namely 3.1 m, 4.1 m and 5.6 m. For each distance, different numbers of samples are acquired, namely 128, 512, 1024 and 1500, by fixing the resolution at 10 cm/Hz, according to equation (2), on the basis of the bandwidth B imposed by the sensor.



(a) (b)

Fig. 3. Measurement setup: (a) Radar module; (b) Reflective screen

The Fourier analysis [7] performed on the acquired data has revealed that the number of samples needed to correctly measure a target range must be greater than 128, as illustrated in Fig. 4, in order to accurately detect the target peak. It can be easily observed, from the same Figure 4, that each measurement is 40 cm shifted from the real distance, due to the fixed delay phase introduced by the instrumentation (circuitry, cables and horn antenna) . This can be obviously removed by a proper calibration procedure. Furthermore, in the final radar prototype a high pass filter will be applied to remove the noise effect on the lower part of the spectrum (Fig. 4).

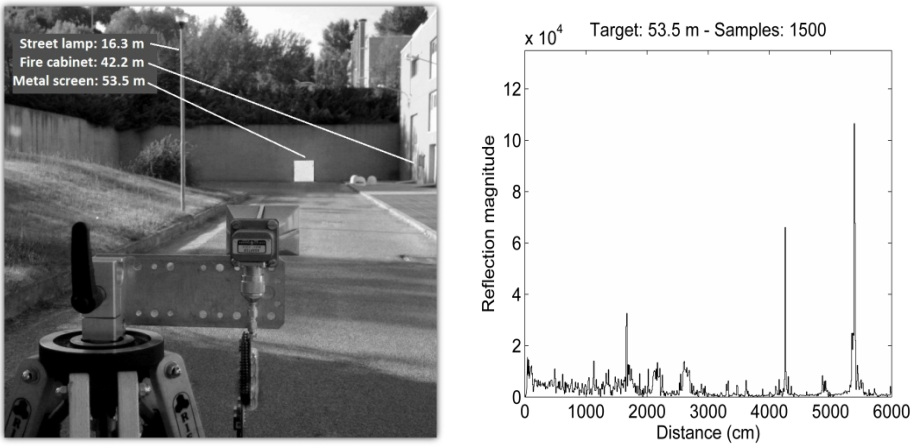


(a) (b)

Fig. 4. Measurements at 3.1 m distance: (a) FFT performed on 128 samples: noise is higher than distance; (b) FFT performed on 512 samples: distance measurement is higher than noise

4.2 Outdoor Test: Noisy Environment with Multiple Obstacles

The outdoor test is performed into the parking area of our Department at University of Calabria, a wide area full of reflective obstacles and external electromagnetic noise (WiFi transmitters and other sources are present all around).



(a) (b)

Fig. 5. (a) Outdoor test setup and (b) Multiple detection: street lamp at 16.3 m, fire cabinet at 42.2 m, metal screen at 53.5 m

Setup and configuration parameters are the same of the previous test, except for the number of samples, fixed to 1500 for all measurements. The screen is placed respectively at distances equal to 2 m, 5 m, 20 m and 53.5 m.

The first two measurements are assumed as calibration, by revealing a constant shift of 40 cm due to the instrumentation. The other two measurements are used to check the noise level and the signal power loss for a long distance measurement (more than 5 m). The last measurement, with the target placed at a distance equal to 53.5 m, shows the ability of the sensor to detect multiple objects (street lamp, fire cabinet, metal screen). As it can be seen from Figure 5, the power received from the object placed at 53.5 m (metal screen) is the highest, but the remaining peaks associated to the other two reflective objects can be also distinguished.

Acknowledgments. This work has been carried out under the framework of PON 01_01503 National Italian Project “Landslides Early Warning”, financed by the Italian Ministry of University and Research.

References

1. Venneri, F., Costanzo, S., Di Massa, G., Borgia, A., Corsonello, P., Salzano, M.: Design of a reconfigurable reflectarray based on a varactor tuned element. In: Proc. of 6th European Conference on Antennas and Propagation, EuCAP 2012, pp. 2628–2631 (2012)
2. Venneri, F., Costanzo, S., Di Massa, G.: Reconfigurable aperture-coupled reflectarray element tuned by single varactor diode. Electronics Letters, 68–69 (2012)
3. Venneri, F., Costanzo, S., Di Massa, G. et al.: Beam-Scanning Reflectarray Based on a Single Varactor-Tuned Element. Int. Journal of Antennas and Propagation 2012, Article ID 290285, 5 pages (2012), doi:10.1155/2012/290285
4. Venneri, F., Costanzo, S., Di Massa, G.: Design and validation of a reconfigurable single varactor-tuned reflectarray. IEEE Transactions on Antennas and Propagation (2013), doi: 10.1109/TAP.2012.2226229
5. Frequency Modulated Continuous Wave Radar, `http://www.siversima.com`
6. Nicolaescu, I., van Genderen, P., Van Dongen, K.W., van Heijenoort, J., Hakkaart, P.: Stepped Frequency Continuous Wave - Data Preprocessing. In: 2nd International Workshop on Advanced GPR, Delft (2003)
7. Brandwood, D.: Fourier Transforms in Radar and Signal Processing. Artech House, Boston (2003)
8. Venneri, F., Costanzo, S., Di Massa, G., Angiulli, G.: An improved synthesis algorithm for reflectarrays design. IEEE Antennas and Wireless Propagation Letters 4, 258–261 (2005)

Compact Slotted Antenna for Wideband Radar Applications

Sandra Costanzo and Antonio Costanzo

DIMES – University of Calabria
87036 Rende (CS), Italy
costanzo@deis.unical.it

Abstract. A new type of slotted microstrip antenna, suitable for wideband radar applications, is proposed in this work. The standard U-slot shape, usually adopted to enlarge the operating bandwidth of a microstrip patch, is properly modified to reduce the undesired cross-polar effects, while providing also a significant size reduction. Validations on X-band and P-band, useful for radar applications, are successfully discussed. In particular, a P-band prototype has been manufactured and tested, while the numerical simulation results of an X-band antenna, performing a 45% bandwidth, are successfully reported.

Keywords: Slotted antennas, broadband antennas, compact antennas, X-band antennas, P-band antennas.

1 Introduction

Most of actual radar systems operates in the microwave region of the electromagnetic spectrum. The range between 300 MHz and 1 GHz, for example, is used for early warning and target acquisition of objects over a long range, while frequencies in the X-band (8-12 GHz) are mainly adopted in those applications requiring small antennas but not so long detecting range. Microstrip patch antennas are often used in radar systems due to their low profile, lightweight, good strength, low costs and easy manufacturing, especially in the case of single layer designs. However, the use of standard microstrip patch antennas is strongly limited by the typical narrow bandwidth behavior and the large size required for low frequencies radar applications. Among the various existing configurations, the U-slot antenna is often adopted to improve the bandwidth of a microstrip patch antenna and/or to reduce its dimensions. In [1], an asymmetric compact U-slot patch antenna, with low probe diameter, performing a 30% bandwidth is presented. Even if the position changing of the U-slot performs good results in terms of bandwidth and compactness, this adopted solution is not recommended in those radar applications which use linear polarization, because of their high level of cross-polarization due to the asymmetry of the structure. Recently, an interesting attempt to provide optimum parameters for a classical U-slot antenna, through a genetic algorithm, has been performed in [2], but in this case the optimization has not been focused on the level of cross-polar field components. Most

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 989–996.
DOI: 10.1007/978-3-642-36981-0_93

of existing works in literature [3], [4], [5] for the U-slot configuration adopt empirical designs essentially based on numerical simulations, with poor theoretical elements able to give a general synthesis procedure. The most significant attempt to establish design criteria for the standard U-slot patch configuration is provided in [6], where empirical formulas are presented to determine the antenna dimensions once fixed the desired operating bandwidth. Even if the narrow band behavior can be avoided by using U-slot configurations, the cuts in the patch usually cause cross-polarization problems, thus motivating the work conducted in this paper. Starting from the study of the U-slot antenna behavior described in [6], the original U-slot geometry is properly modified in this contribution to accomplish for a wideband response, but reducing cross-polarization effects due to the complex combination of antenna and slot modes. Furthermore, a significant reduction of the total surface area of the radiating element is successfully obtained. The proposed approach, valid for wideband radar applications, is tested by numerical simulations on an X-band modified U-slot prototype. As a further validation test, experimental results on a P-band prototype are discussed to demonstrate the antenna compactness and the wideband behavior.

2 Antenna Configuration

Different types of slot shapes may be adopted to enhancing the bandwidth and/or reducing the size of the radiating element [7], [8]. The modified U-slot antenna configuration proposed in this contribution (Fig. 1), allows about a 50% width reduction (compared to the standard U-slot antenna [6]) while maintaining a large operating bandwidth and a good behavior in terms of cross-polar radiation.

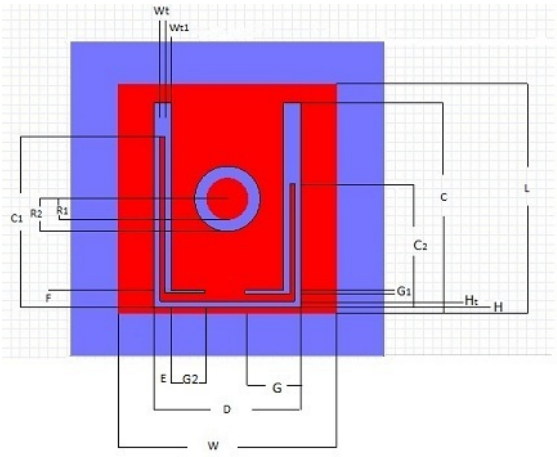


Fig. 1. Layout of modified U-slot antenna. The patch (*dark colored*) surface area is about $\lambda^2/9$ (λ free space wavelength). The ground plane (*light colored*) surface area is about $\lambda^2/4$.

A thick substrate is adopted to satisfy the bandwidth requirements, as in usual U-slotted patch, but significant changes are introduced with respect to the standard U-slot configuration. As discussed in [6], antenna and slot modes combinations give rise to undesired x-directed currents which produce a high cross-polar component unless the ratio between the vertical arm C and the horizontal arm D does not satisfy a specific criterion. This means that bandwidth requirements cannot always be satisfied while guaranteeing, at the same time, a low cross-polar effect. In order to solve this problem, the combination of the cuts in this design allows a good trade-off between bandwidth improvement and the paths minimization of unwanted x-directed modes. To have a better tuning of the operating bandwidth, unequal lengths are chosen for the strips inside the two U-slot vertical arms. As a further variation with respect to the original U-Slot design [6], a ring slot is introduced around the feeding point to compensate the large probe inductance due to the presence of thick substrates, which could worsen the impedance matching. Accurate design formulas, taking into account all relevant geometrical parameters, have been developed and a full description of the design rules will be provided in a future extended version of this work. According to this design, a simulated X-band prototype and a realized P-band prototype are fully described in the next sections.

3 Compact P-band U-slotted Antenna

Starting from the configuration proposed in Section 2, a P-band prototype working at the central frequency of 450 MHz, useful for many radar applications, is realized and tested. A foam substrate ($\varepsilon_r = 1.07$) having thickness equal to 60 mm is assumed in this case as main support substrate, while an upper thin layer (0.762 mm) of Arlon DiClad870 ($\varepsilon_r = 2.33$) is uniquely adopted for robustness reasons. The final dimensions of P-band prototype, according to the layout of Fig. 1, are reported in Table 1, while the realized structure is shown in Fig. 2.

Table 1. Dimensions of P-Band Modified U-Slotted Antenna

Parameter	Dimension [cm]	Parameter	Dimension [cm]
L	24.34	R_1	2.20
W	23.11	C_2	23.11
D	15.60	F	1.80
E	1.80	W_t	0.50
C	21.80	G	5.92
C_1	18.15	R_2	3.50
H	0.6	G_2	0.00

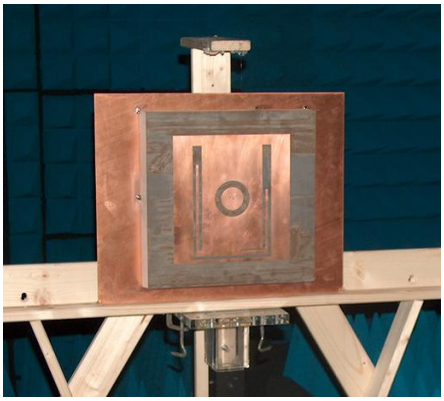


Fig. 2. Photograph of P-band prototype into the anechoic chamber of Microwave Laboratory at University of Calabria

At the central design frequency, the width of the radiating element of a basic U-Slot patch antenna, according to the literature [6], is equal to 49.26cm. In the proposed design, a 50% curtailment of the patch width is obtained, thus achieving a significant compactness. Numerical simulations are performed using the commercial software Ansoft Designer V 3.5, based on the Method of Moment technique. The P-band antenna is tested into the Microwave Laboratory at University of Calabria, equipped for both far-field and near-field tests [9], [10], [11]. The experimental characterization is performed in terms of both return loss (Fig. 3) and radiation pattern (Fig. 4) measurements.

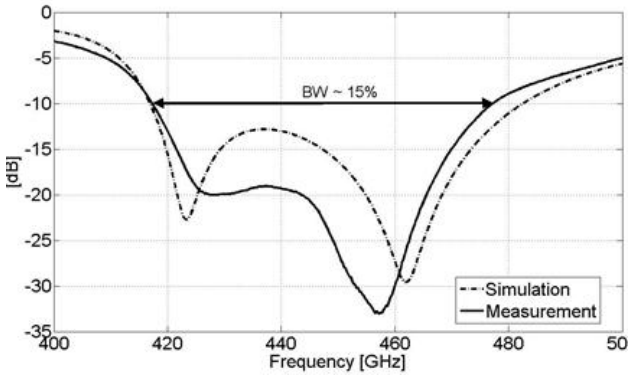


Fig. 3. Return loss of P-band modified U-slot antenna: comparison between simulations and measurements

In particular, a bandwidth of about 15% is successfully obtained, as reported in Fig. 3, where the return loss is maintained properly under the threshold of -10dB in the whole operative bandwidth for both the results obtained by numerical simulations and measurements; even if slotted antennas behavior is usually significantly affected

by small variations of the geometrical parameters, in this case possible errors introduced by the manufacturing and by the planar simulation don't waste the goodness of the final measured results.

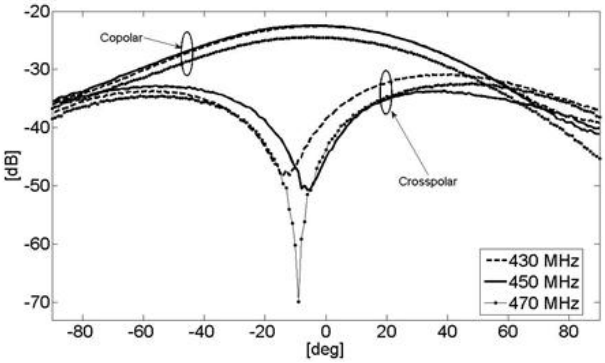


Fig. 4. Measured radiation patterns of P-band modified U-slot antenna

The wideband response of the proposed antenna is also confirmed by the radiation patterns illustrated in Fig. 4, revealing a very similar behavior at different frequencies within the operating frequency range, yet obtained from the return loss characterization, with a cross-polar component properly below the co-polar field within the entire frequency band. This wideband behavior is achieved by properly minimizing the x-directed currents on the radiating element, which depend in particular on the correct sizing of the geometrical parameters D, C, G_1 and G_2.

4 Wideband X-band U-slotted Antenna

Several modern X-band radars require wideband or ultra-wideband antennas, thus motivating the application of the proposed antenna configuration to achieve a large bandwidth of about 45%. In order to prevent cross-polarization effects within the entire frequency band, a main electrically thick substrate of Arlon DiClad 870 (thickness equal to 0.36 cm) is adopted to support the antenna. As highlighted in the previous section, all design equations will be accurately discussed in a future extended version of this contribution. A rectangular copper ground plane having dimensions 20 mm x 15 mm is placed as bottom layer of the structure. The values of all relevant geometrical parameters for the proposed X-band antenna are reported in Table 2. The simulated return loss of the X-band antenna, performed on Ansoft Designer software, is illustrated in Fig. 5.

Table 2. Dimensions of X-band modified U-slot antenna

Parameter	Dimension [mm]	Parameter	Dimension [mm]
L	7.25	R_1	1.65
W	14.18	C_2	2.57
D	11.55	F	0.80
E	0.80	W_t	0.27
C	6.26	G	5.77
C_1	5.74	R_2	1.68
G_2	1.97	H	0.26

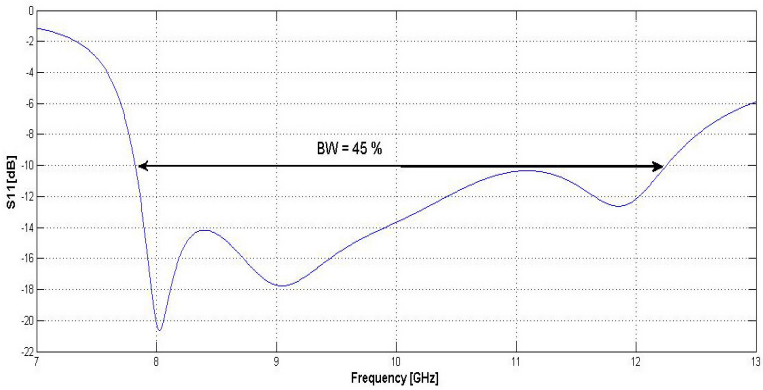


Fig. 5. Simulated return loss of X-band modified U-slot antenna

According to the design rules of the proposed antenna configuration, a bandwidth of about 45% is obtained, thus covering the whole X-band. Furthermore, the numerical simulations on the radiation patterns confirm that the cross-polar component properly remains below the co-polar component, thus avoiding undesired spurious radiations. This advantageous feature is illustrated in Fig. 6, where co-polar and cross-polar patterns on both E-plane and H-plane are reported for the central design frequency of 10 GHz. Further discussions on the X-band antenna, including the photo of the realized prototype and the experimental results, will be illustrated during the conference presentation.

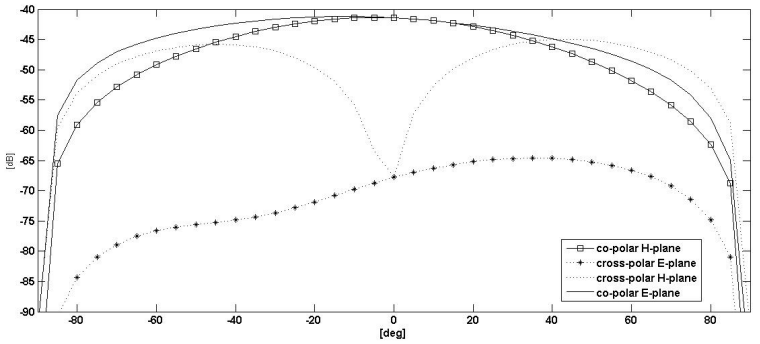


Fig. 6. Simulated radiation patterns of X-band modified U-slot antenna (@ 10 GHz)

5 Conclusions

A new type of slotted microstrip patch antenna for radar applications, based on the modification of the standard U shape slot, has been proposed in this work. The main improvements with respect to the standard U-slot configuration are relative to the cross-polar effects lowering and the significant size reduction of the radiating element. Numerical and experimental results on different frequency bands, namely X-band and P-band, both largely adopted in radar applications, have been discussed to show the advantageous features of the proposed design configuration, in terms of wideband behavior, size compactness and significant cross-polar effects reduction.

Acknowledgments. This work has been carried out under the framework of PON 01_01503 National Italian Project "Landslides Early Warning", financed by the Italian Ministry of University and Research.

References

1. Khodaei, G.F., Nourinia, J., Ghobadi, C.: A practical miniaturized U-slot patch antenna with enhanced bandwidth. Progress In Electromagnetics Research B 3, 47–62 (2008)
2. Rani, A., Dawre, R.K.: Design and Analysis of Rectangular and U Slotted Microstrip Patch using Optimization Program in Java for UHF Applications. International Journal of Computer Applications (0975 – 8887) 3, 26–30 (2010)
3. Tong, K.F., Luk, K.M., Lee, K.-F., Lee, R.Q.: A broadband U-slot rectangular patch antenna on a microwave substrate. IEEE Trans. Antennas Propag. 48, 954–960 (2000)
4. Lee, K.-F., Yang, S.L.S., Kishk, A.A., Luk, K.M.: The versatile U-slot patch antenna. IEEE Antennas Propag. Magazine 52, 71–88 (2010)
5. Lee, K.-F., Luk, K.M., Mak, K.M., Yang, S.L.S.: On the use of U-slots in the design of dual- and triple-band patch antennas. IEEE Antennas Propag. Magazine 53, 60–73 (2011)
6. Weigand, S., Huff, G.H., Pan, K.H., Bernhard, J.T.: Analysis and design of broad-band single-layer rectangular U-slot microstrip patch antennas. IEEE Trans. Antennas Propag. 51, 457–468 (2003)

7. Jolani, F., Dadgarpour, A.M.: Compact M-slot folded antenna for WLAN. Progress In Electromagnetics Research Letters 3, 35–42 (2008)
8. Ansari, J.A., Singh, P., Dubey, S.K., Khan, R.U., Vishvakarma, B.R.: H-shaped stacked patch for dual-band operation. Progress in Electromagnetics Research 5, 291–302 (2008)
9. Costanzo, S., Di Massa, G.: Far-field reconstruction from phaseless near-field data on a cylindrical helix. Journal of Electromagnetic Waves and Applications 18(8), 1057–1071 (2004)
10. Costanzo, S., Di Massa, G.: Direct far-field computation from bi-polar near-field samples. Journal of Electromagnetic Waves and Applications 20(9), 1137–1148 (2006)
11. Costanzo, S., Di Massa, G.: Near-field to far-field transformation with planar spiral scanning. Progress In Electromagnetics Research 73, 49–59 (2007)

High Resolution Software Defined Radar System for Target Detection

Sandra Costanzo, Francesco Spadafora, Antonio Borgia, Oswaldo Hugo Moreno, Antonio Costanzo, and Giuseppe Di Massa

DIMES – University of Calabria
87036 Rende (CS), Italy
costanzo@deis.unical.it

Abstract. A high resolution Software Defined Radar system is implemented in this work by adopting the new generation Universal Software Radio Peripheral USRP NI2920, a software defined transceiver. The enhanced available bandwidth due to the Gigabit Ethernet interface is exploited to achieve the high range resolution features. At this purpose, a specific Labview application implementing the radar operations is developed. The realized SDRadar system is successfully validated by preliminary outdoor tests accurately retrieving the distance of a reference target.

Keywords: Software Defined Radio, Radar, Slant Range Resolution.

1 Introduction

The flexibility of software based systems and their easy adaptability make them useful for many different applications. The Software Defined Radar (SDRadar) system is a special type of versatile radar in which operations and components, typically realized by specific hardware (*i.e.*, mixers, filters, modulators and demodulators), are implemented in terms of software modules [1]. To implement a SDRadar, some recent researches and studies [2], [3] were conducted through the use of FPGA and/or DSP.

The Universal Software Radio Peripheral (USRP) transceiver can be used to develop Software Defined Radio applications like SDRadar, thus leading to obtain a low cost radar sensor. A first attempt to adopt USRP for radar applications was performed by the authors in [4], where a SDRadar system was implemented through the adoption of first generation USRP. Due to the bandwidth limitations imposed by the available USB connection, the solution presented in [4] gives a limited slant-range resolution equal to 75 m, so alternative solutions have been investigated to enhance the radar performance. Other excellent results have been conducted in [5], [6], [7], [8], where the characterization of the USRP N200 e N210 in radar field has been considered. In particular, National Instruments (NI) has recently manufactured a new generation of USRP for wireless communications teaching and research. It successfully combines the NI LabVIEW software and the USRP hardware to deliver an affordable and easy-to-use software-reconfigurable RF platform that works well

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 997–1005.
DOI: 10.1007/978-3-642-36981-0_94

for communications, education, experimentation, research, and rapid prototyping [9]. In this paper, the potentiality of the NI new generation USRP is exploited to enhance the radar resolution of the first SDRadar prototype proposed in [4]. A specific LabVIEW code is developed to control the SDRadar system, with the implementation of a signal processing compression-based technique to achieve a strongly enhanced slant-range resolution equal to 6 m. In the following sections, a complete description of the hardware and the relative control algorithm is provided. Furthermore, experimental results obtained by outdoor tests are discussed to prove the enhanced radar resolution.

2 USRP NI2920 Hardware Description

The first USRP motherboard was designed by Matt Ettus at the "National Science Foundation" in 2006. Nowadays, four versions are available, namely USRP, USRP2, USRP N200 and USRP N210. In the last year, the National Instruments has realized three new boards, namely USRP 2920, 2921, 2922, interfacing with the PC through Labview software. The USRP 292X main features are as follows:

- 2 channels ADC, 400MS/s;
- 2 channels, 100MS/s;
- 1 GIGABIT ETHERNET for PC interface;
- Xilinx Spartan-6;
- 25MHz of operating bandwidth.

The block scheme is shown in Figure 1.

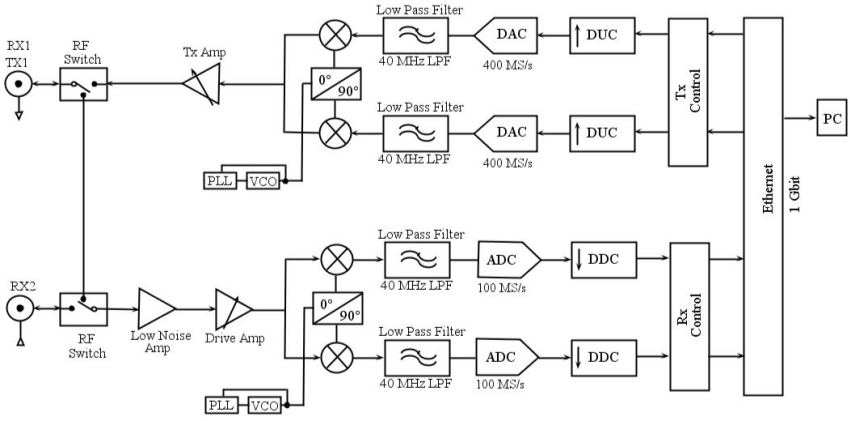


Fig. 1. USRP 2920 block diagram

Incoming signals attached to the standard SMA connector are mixed down from RF using a direct-conversion receiver (DCR) to baseband I/Q components, which are sampled by a 2-channel, 100 MS/s, 14-bit analog-to-digital converter (ADC). The digitized I/Q data follows parallel paths through a digital down-conversion (DDC) process that mixes, filters, and decimates the input 100 MS/s signal to a user-specified rate. The down-converted samples are passed to the host computer up to 20 MS/s over a standard Gigabit Ethernet connection. For transmission, baseband I/Q signal samples are synthesized by the host computer and fed to a USRP-292x up to 20 MS/s over Gigabit Ethernet. The USRP hardware interpolates the incoming signal to 100 MS/s using a digital up-conversion (DUC) process and then converts the signal to analog with a dual-channel, 16-bit digital-to-analog converter (DAC). The resulting analog signal is then mixed up to the specified RF frequency [9].

The main limitation of the SDRadar technology is due to the interface with the PC, that reduces the radar performance in terms of slant range resolution [10]. The first generation USRP, by Matt Ettus, uses a USB 2.0 interface to connect to the PC, thus imposing the adoption of the low USB band for data transmission, which leads to very low slant range resolutions. The behavior and the analysis of the first generation USRP in radar field was conducted by the authors in a recent work [4], where a SDRadar system was implemented with a slant range resolution equal to 75 m. However, radar applications typically require more refined precisions for target detection, so alternative solutions are investigated to enhance the system bandwidth and thus the SDRadar resolution. In particular, the adoption of the new USRP NI 2920 is considered in this work to exploit the associated Gigabit Ethernet interface in order to improve the SDRadar capabilities.

2.1 Signal Processing Algorithm

In order to demonstrate the range resolution improvement using the USRP 2920, a signal processing technique, called Stretch Processor [10], is implemented in Labview code. This processing is a particular pulse compression technique which consists of four distinct steps. First, the radar returns are mixed with a replica (reference signal) of the transmitted waveform. This is followed by Low Pass Filtering (LPF) and coherent detection in order to avoid the high frequency response achieved at the output of the Mixer (see figure 2). Next, Analog to Digital (A/D) conversion is performed, and finally a bank of Narrow Band Filters (NBFs) is used to extract the tones proportional to the target range, since stretch processing effectively converts time delay into frequency. A block diagram for a stretch processing receiver [10] is illustrated in Figure 2. The transmitted signal is an Linear Frequency Modulated (LFM) waveform expressed by the following equation:

$$s(t) = cos\left(2\pi\left(f_o t + \frac{\mu}{2}t^2\right)\right), \qquad 0 < t < \tau' \tag{1}$$

where $\mu = B/\tau'$ is the LFM coefficient, B gives the bandwidth, f_0 is the chirp start frequency and τ' is the chirp duration. The slant range resolution ΔR is given by:

$$\Delta R = \frac{c}{2B} \tag{2}$$

On the basis of the above equation, the use of the USRP 2920 NI, giving a maximum available bandwidth B = 25 MHz, leads to have a slant range resolution equal to 6 m, which is significantly enhanced with respect to the value of 75 m achieved with the first generation USRP where B is equal to 2 MHz [9].

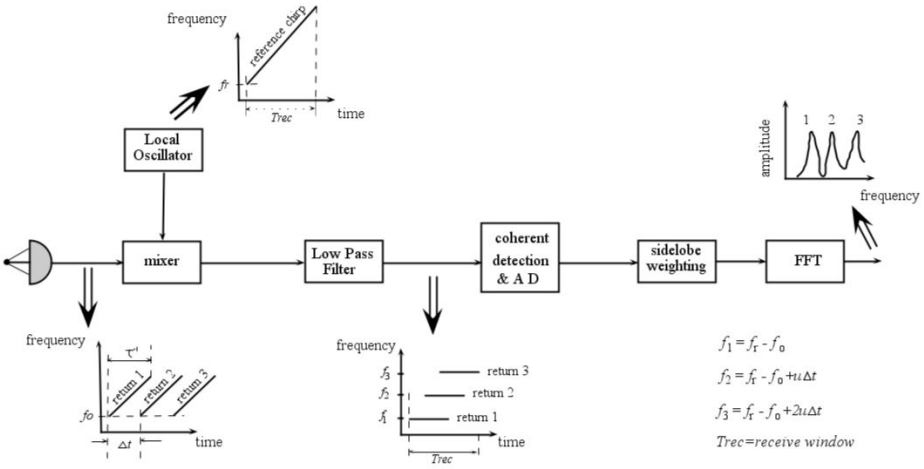


Fig. 2. Stretch processing block diagram

3 3 SDRadar System

The idea is to implement a SDRadar system able to scan a complete area under analysis and to locate, through N different radar scannings in different horizontal positions, the surface topology. In Figure 3 is reported the system block diagram through which the USRP 2920 is used to transmit and receive data by two linear array antennas, that are rotated by a controlled motor. The system is interfaced by a Labview window running on a Single Board Computer (SBC) which processes all the transmitted and received data to determine the topology of the area under analysis. This interface is able to control the motion motor too. A Power Amplifier (AMP) and a Low Noise Amplifier (LNA) are connected to the transmitting (TX) and the receiving (RX) antenna to increase the power along both the transmission and the receiving paths.

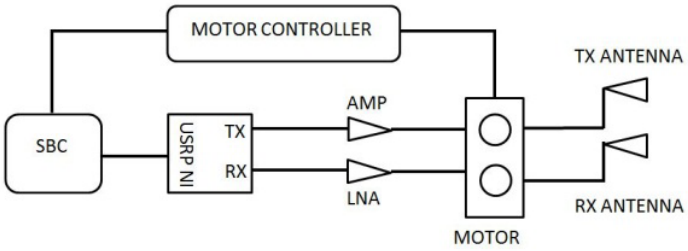


Fig. 3. SDRadar block diagram

3.1 SDRadar Algorithm for Complete Horizontal Scanning

The proposed SDRadar leads to scan, N times, different frames of the area under analysis (eg. Mountain, Landslide, Topography surfaces, Glaciers …) through a horizontal movement of the radar antenna controlled by a driver motor. Figure 4 shows the proposed algorithm, summarized in the following step:

1. Parameter Definition

   - *Footprint* (antenna illuminating area) of each scan defined by the distance between the radar antenna and the analyzed area, the azimuth and the elevation antenna beam widths, the grazing angle and the operating frequency.
   - The *receiving window,* that ensures the correct recognition of any type of topology of the surface under analysis, defined by R_{min} and R_{max} (minimum and maximum target range required).
   - *Total Area size*: which gives the exact number N of radar scanning necessary to retrieve the total topology.

2. An N scan matrix, defined by the parameter of the previous step.
   This Matrix, made up by N rows, that correspond at N scan produced, and M columns, that depend of the receiving window.

3. A For loop is started for each N scan. The scans are obtained by the motor that rotate the antenna by an angle θ_{scan} N times. For each $n<N$ the Matrix is filled with the results from the Stretch Processor described in the previous section and retrieved from the USRP.

4. When the matrix is completed, a colors assignment is performed like in a radar-gram [11]. The colors are helpful for the remote view of the topology.

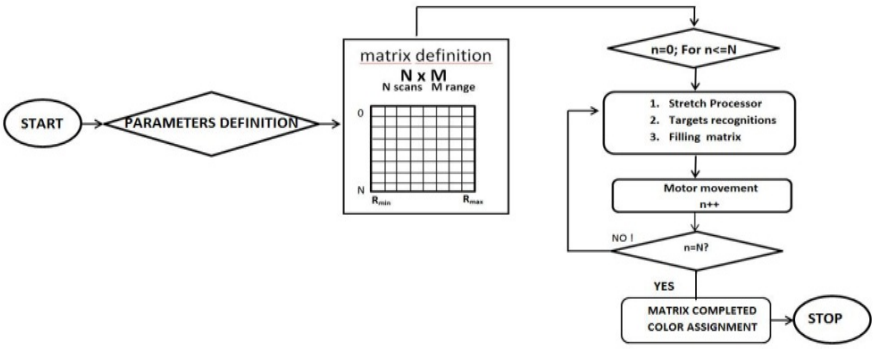


Fig. 4. SDRadar algorithm

4 Outdoor SDRadar Tests

Specific tests are performed on the USRP NI2920 in order to identify key features of the device in radar field. At this purpose, the USRP is connected to an host PC through Gigabit Ethernet and it is controlled by an own developed Labview application illustrated in Figure 5.

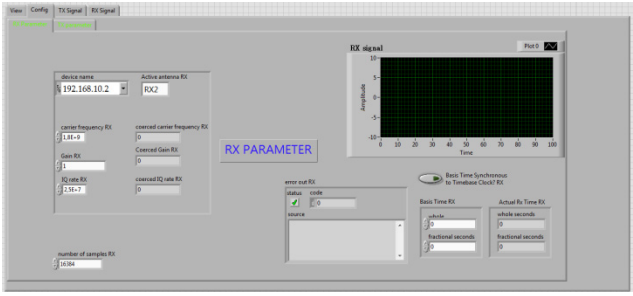


Fig. 5. Labview SDRadar application window

To demonstrate the enhanced Slant Range Resolution, an outdoor experimental setup is assessed (Fig. 6), with a broadband ridged horn antenna employed for the transmission and a broadband logarithmic antenna adopted for the reception. A metal plate, positioned at different distances in line of sight direction from the transmitting/receiving platform is assumed as target under test. This preliminary test is performed without motors so the algorithm test described in the previously section was considered for only one scan with θ_{scan} equal to 0 degree. The real and software retrieved target positions are successfully compared in Table 1 for various target distances. The relative signal peaks, properly retrieved by the implemented Stretch Processor technique, are illustrated in Figure 7. As a further validation, the SDRadar map for a single scan is illustrated in Figure 8, where three different targets at 6 -12-18 m are displayed with different colors.

Table 1. Real and retrieved target positions with the USRP 2920

Real target position [m]	Retrieved software position [m]
0 ÷ 6	6
6 ÷ 12	12
12 ÷ 18	18

Fig. 6. Software Defined Radar Test

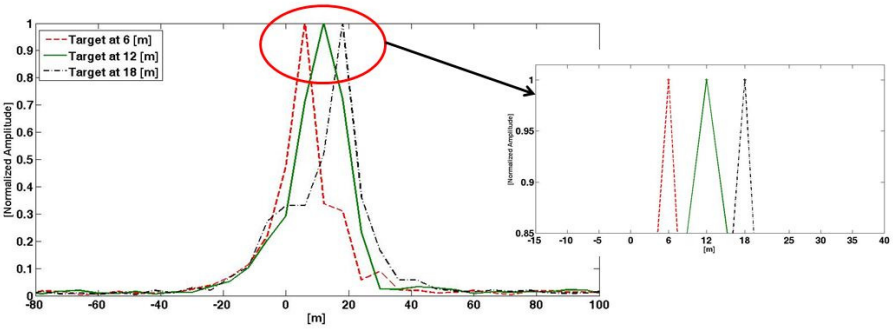


Fig. 7. Retrieved signal peaks for different target positions

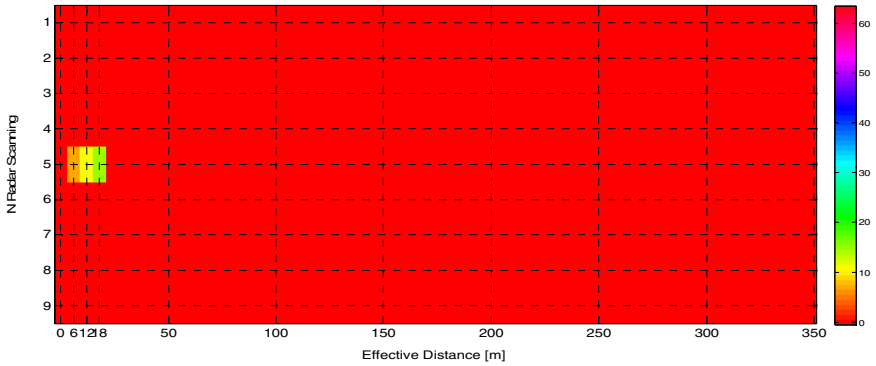


Fig. 8. SDRadar map for a single scanning

5 Conclusions

A low cost, flexible, versatile and small dimensioned solution to create a high performance radar system has been proposed in this work. The USRP NI2920 has been adopted to realized a SDRadar system giving a 6 m Slant Range Resolution, significantly enhanced with respect to that achieved in the existing SDRadar solutions. A specific Labview application has been developed to implement the high resolution radar processing algorithm and outdoor experimental validations are performed to demonstrate the theoretical features.

Acknowledgments. This work has been carried out under the framework of PON 01_01503 National Italian Project "Landslides Early Warning", financed by the Italian Ministry of University and Research.

References

1. Debatty, T.: Software Defined RADAR a state of the art. In: Second International Workshop on Cognitive Information Processing, Brussels, Belgium, pp. 253–257 (2010)
2. Zhang, H., Li, L., Wu, K.: 24ghz software-defined radar system for automotive applications. In: European Conference on Wireless Technologies, Munich, Germany, pp. 138–141 (2007)
3. Kauffman, K., Garmatyuk, D., Schuerger, J.: Multifunctional software-defined radar sensor and data communication system. IEEE Sensors Journal 11, 99–106 (2011)
4. Aloi, G., Borgia, A., Costanzo, S., Di Massa, G., Loscrì, V., Natalizio, E., Pace, P., Spadafora, F.: Software Defined Radar: synchronization issues and practical implementation. In: COGART, International Conference on Cognitive Radio and Advanced Spectrum Management, Barcelona (2011)
5. Prathyusha, C., Sowmiya, S.N., Ramanathan, S., Soman, R., Amrita, K.P., Deepthi, V.V., Chinnam, M., Nandhini, J.: Implementation of a low cost synthetic aperture radar using software defined radio. In: International Conference on Computing Communication and Networking Technologies (ICCCNT), Karur, TamilNadu, India, pp. 1–7 (2010)

6. Manuel, F., Martin, B., Christian, S., Lars, R., Jondral Friedrich, K.: An SDR-based Experimental Setup for OFDM-based RadaR. In: 7th Karlsruhe Workshop on Software Radio Karlsruhe, Germany (March 2012)
7. Marcus, M., Martin, B., Manuel, F., Jondral Friedrich, K.: A USRP-based Testbed for OFDM-based Radar and Communication Systems. In: 22nd Virginia Tech Symposium on Wireless Communications, Blacksburg (June 2012)
8. Fernandes, V.: Implementation of a RADAR System using MATLAB and the USRP, CSUN ScholarWorks (2012)
9. NI USRP-2920, NI USRP-2921, National Instruments Data-sheet, http://sine.ni.com
10. Mahafza, B.R., Elsherbeni, A.Z.: Simulations for Radar Systems Design. Chapman & Hall /CRC (1999)
11. Skolnik, M.: Radar handbook, 3rd edn., pp. 21.1—21.41. Mc Graw Hill, San Francisco (2008)

Design of a Reconfigurable Reflectarray Unit Cell for Wide Angle Beam-Steering Radar Applications

Francesca Venneri, Sandra Costanzo, and Giuseppe Di Massa

DIMES – University of Calabria
87036 Rende (CS), Italy
costanzo@deis.unical.it

Abstract. A reconfigurable aperture-coupled reflectarray element is proposed for the realization of beam steering antennas, suitable for radar applications. Each reflectarray element is coupled to a microstrip line, which is loaded by a single varactor diode acting as a phase shifter element, thus providing a continuously variable reflection phase. A reduced size reflectarray unit cell is properly designed in order to extend the antenna beam scanning capabilities within a wide angular region, but avoiding the occurrence of undesired grating lobes. The radiating structure is properly optimized to obtain a full phase tuning range at the frequency of 11.5 GHz, thus assuring a good agility and accuracy in the reconfiguration of the reflectarray radiation pattern.

Keywords: Reflectarrays, phased arrays, radar antennas.

1 Introduction

Phased array antennas are usually adopted in the design of modern radar systems, offering many advantages such as agile beams, low profile and scalability.

Unlike mechanically scanned antennas, phased arrays integrate the actual radiating structures with phase shifter components and/or more complex T/R modules [1] which control the input signal of each radiating element, thus allowing the radiated main beam to be electronically steered.

By electronic scanning, the radar beams can be positioned almost instantaneously, without time delays and vibration of mechanical systems.

Furthermore, electronically scanned antennas offer increased data rates, instantaneous positioning of the radar beam anywhere, avoiding also mechanical errors and failures associated with mechanically scanned antennas.

A very attractive alternative to traditional phased array antennas is offered by the fairly new reflectarray concept [2]. As a matter of fact, this antenna type may be specifically designed for those applications requiring beam scanning capabilities or pattern reconfigurability. Furthermore, reconfigurable reflectarrays may offer many advantages over conventional phased arrays, such as reduced costs and volume, a simpler architecture due to the absence of complicated beam-forming networks, and increased efficiencies thanks to the use of spatial feeding.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1007–1013.
DOI: 10.1007/978-3-642-36981-0_95

They consist of a printed array illuminated by a feed antenna. Each radiating element is properly designed to give a phase response able to create a total reradiated field with some desired features, such as prescribed beam directions and/or shapes.

Many different reflectarray configurations have been proposed in literature and, recently, many efforts have been spent in the design of reconfigurable reflectarray elements, which are usually based on the use of microstrip patches integrated with one or more electronically controllable components, such as MEMs and varactor diodes [3-5].

Recently, the authors have proposed a novel reconfigurable reflectarray element based on the use of an aperture-coupled patch electronically driven by a single varactor diode [6-9]. The radiating patch is coupled to a microstrip line printed onto a different substrate and loaded by a varactor (Fig. 1). By changing the bias voltage across the diode, a variable phase shift is added to the reradiated field, thus providing a dynamic control of the element backscattering features. A detailed description of the proposed phase tuning mechanism is reported in [9], where the procedure giving a proper phasing line optimization is also illustrated.

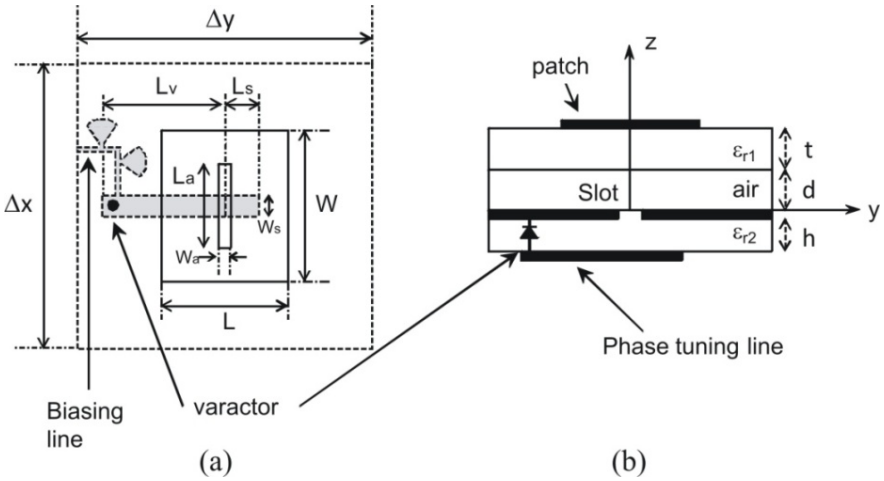


Fig. 1. Single element geometry: (a) top view; (b) side view

The phase tuning capabilities of this configuration have been already demonstrated in [6], while in [7-9] the proposed element has been successful adopted for the design of a reconfigurable reflectarray prototype composed by 3×15 elements and characterized by a unit cell with size equal to $\Delta x \times \Delta y = 0.7\lambda_0 \times 0.7\lambda_0$ at 11.5 GHz. The reconfigurability of the proposed antenna has been tested in [7-9], through several measurements of its radiated pattern, for different configurations of the controlling varactors voltages, which are properly computed through the implementation of the synthesis algorithm described in [10]. In [7-9] the antenna beam steering capabilities have been demonstrated within an angular region, which extends from -25° to 25°, in the principal radiation plane containing the largest number of array elements. This limited scanning range is principally imposed by the unit cell size.

In this work, the reflectarray element proposed in [6-9] is redesigned in order to offer the possibility to obtain a reflectarray with an increased scanning region. Pointing out that a large scan angle requires a close element spacing, less than or equal to half wavelength at the operating frequency [11], a reflectarray unit cell with a reduced size equal to $0.46\lambda_0 \times 0.46\lambda_0$ (f_0=11.5 GHz) is proposed. In order to accommodate the phasing circuitries inside the reduced available area embedded in the unit cell, the antenna stratification layers are properly modified. According to the considerations reported in [9], the varactor loaded line is accurately resized in order to maximize the phase agility of the antenna. A phase tuning range of about 330° is numerically demonstrated, by varying the capacitance of the varactor diode within the values ranging from 0.2 pF up to 2pF.

2 Performance Limitations of Beam Steering Arrays

The angular displacement of an electronically scanned radar beam is practically limited by two principal factors, namely the element pattern and the array elements spacing. As a matter of fact, the radiation pattern of an array of identical radiators is given by the product of the array factor and the element pattern.

If the single array radiator is isotropic, i.e. the array elements radiate an electric field quite uniform along those directions belonging to the scanning plane, only the array factor will affect the total radiation pattern.

However, practical array element patterns are not omnidirectional, showing an amplitude that decays moving away from the broadside direction. In these cases, the single element will significantly reduce the amplitude of the scanned beam, except in the zone where it is nearly isotropic [11].

The second limitation, namely the array elements spacing, is more relevant. As a matter of fact, it is well-known that a large scan angle requires close element spacing, in order to avoid grating lobes appearance. The maximum scan angle, that a linear phased array can achieve, may be derived from the well-known relation [11]:

$$\theta_{s\max} = \sin^{-1}\left(\frac{\lambda}{d} - 1\right) \tag{1}$$

where $\theta_{s\max}$ is the maximum scan angle from broadside direction, d is the spacing between two adjacent elements and λ is the operating wavelength.

Equation (1) is derived from the array factor expression of a linear array placed along the x or y-axis and its validity can be extended to the principal cuts of a planar array placed in the x-y plane [11]. If the array scan angle exceeds the value imposed by (1), grating lobes will appear along other directions. Equation (1) also states that half wavelength spaced arrays will have a complete theoretical scan range of ±90°. On the contrary, when the spacing between elements increases beyond a half wavelength, the scan range of the array is significantly reduced, due to the appearance of grating lobes having the same amplitude of the main beam.

The maximum scan angle achievable by a phased array is also a function of the array length and the desired half-power beam-width [11], however the condition imposed by (1) is necessary for the design of an array with prescribed beam-steering capabilities.

3 Design of a Reconfigurable Reflectarray Element Embedded into a Reduced Size Unit Cell

In order to improve the scanning capabilities of the reconfigurable reflectarray configuration proposed in [6-9], the single reflectarray element is properly redesigned by reducing the unit cell size. In fact, as discussed in the previous paragraph, a closer array elements spacing assures a larger scanning region.

The unit cell dimension is fixed to a value less than a half-wavelength at the operating frequency f_0 =11.5 GHz. In particular, the array grid size $\Delta x \times \Delta y$ is set to a value equal to $0.46\lambda_0 \times 0.46\lambda_0$. Furthermore, as demonstrate in [12-14], a reduced unit cell size allows to improve the bandwidth performances of reflectarray antennas. This last aspect is not considered in the present paper, however it could be analyzed in a future work.

In order to allow the accommodation of the tuning circuitries in the smaller area embedded inside the unit cell, the phasing line substrate adopted in [6] is properly substituted with a layer of Arlon AR600, with ε_r=6 and thickness h=0.762mm (see table 1). The use of a substrate with a higher permittivity allows to reduce the wavelength inside the printed lines, thus providing the possibility to design a shorter phase tuning line. As reported in Table 1, the other layers composing the antenna stratification are equal to those adopted in [6].

Table 1. Element stratification

	Element designed in [6] $\Delta x \times \Delta y = 0.7\lambda_0 \times 0.7\lambda_0$		**Element designed in this work** $\Delta x \times \Delta y = 0.46\lambda_0 \times 0.46\lambda_0$	
Layer	***Material***	***Thickness***	***Material***	***Thickness***
Patch	Copper	35 μm	Copper	35 μm
Antenna substrate	Diclad870 (ε_{r1}= 2.33)	t= 0.762 mm	Diclad870 (ε_{r1}= 2.33)	t= 0.762 mm
	Air	d= 0.762 mm	Air	d= 0.762 mm
Ground plane with slot	Copper	35 μm	Copper	35 μm
Phasing line substrate	Diclad870 (ε_{r2}= 2.33)	h= 0.762 mm	AR600 (ε_{r2}=6)	h= 0.762 mm
Phasing line	Copper	35 μm	Copper	35 μm
-	Air	s= 3.7 mm	Air	s= 3.7 mm
Ground plane reducing back radiation	Copper	35 μm	Copper	35 μm

The following dimensions are obtained for the different parts composing the antenna: W= 7.75 mm, L= 7.75 mm, W_a= 0.5 mm, L_a= 5.7 mm, W_s= 1.6 mm (Fig. 1(a)). The assumed line width W_s corresponds to a characteristic impedance of value equal to 37 Ω. A varactor diode, with a tunable capacitance ranging from 0.2pF to 2pF, is integrated to the microstrip line in order to obtain the required reconfiguration capabilities.

As described in [9], the two line sections L_v and L_s (see Fig. 1) are optimized in order to maximize the phase agility of the element for the assigned varactor capacitance range. At this purpose, the length L_v is set to a value of 4.2 mm, while the stub length L_s is varied from 3.5 mm to 5.4 mm. Fig. 2 shows the reflection phase curves versus varactor capacitance computed for each considered value of the stub length. It can be observed that by increasing L_s a higher phase tuning range of about 330° is obtained. As accurately demonstrated in [9], this last result is due to the introduction of a proper inductive effect, which is directly related to the stub length.

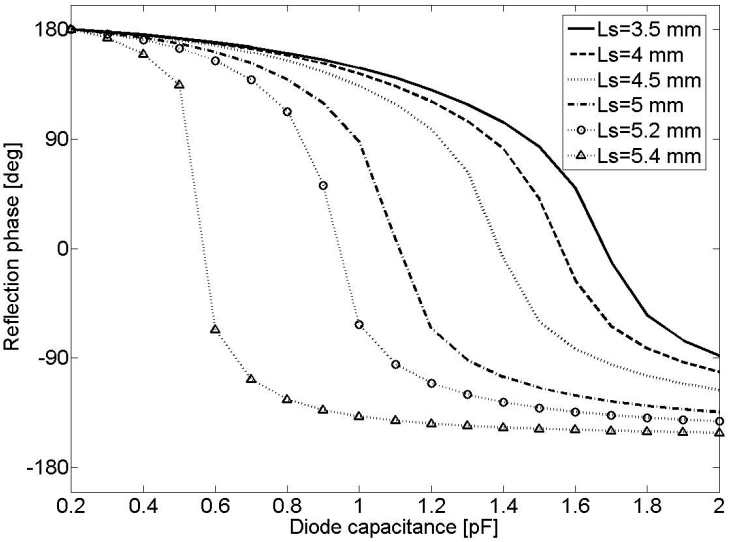


Fig. 2. Phase curve versus diode capacitance for different stub length

The element pattern of the designed unit cell is reported under Fig. 3. The depicted diagrams refers to the reflectarray element with a phasing line having the following dimensions: L_v= 4.2 mm and L_s=5.2 mm. The radiation patterns computed in the two principal planes show a nearly isotropic behavior within the range from -45° to 45°, as in the case of a typical *cos(θ)* source.

In conclusion, the proposed unit cell could be suitable for the design of reflectarray antennas with improved beam steering capabilities. As a matter of fact the main beam could be scanned within an angular region greater than about 40° with respect to the one demonstrated in [7-9], without occurring in the grating lobes phenomena.

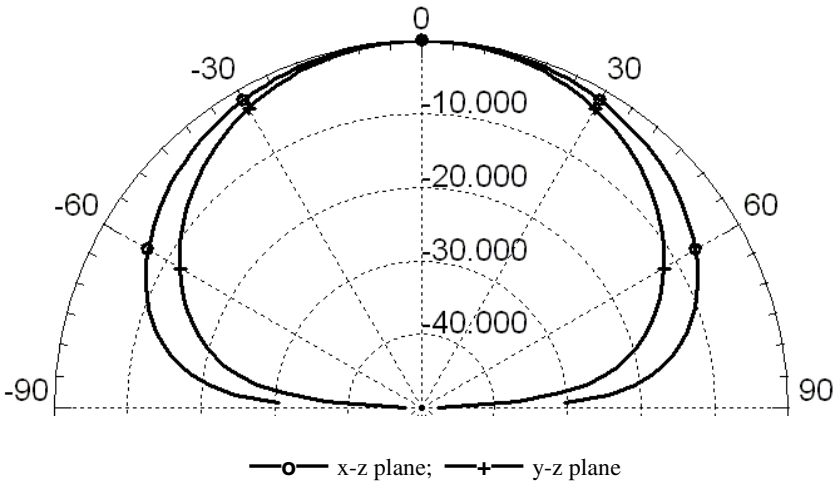


Fig. 3. Element pattern of designed reflectarray element

4 Conclusions

The reflectarray concept has been applied in this work to the design of beam steering antennas suitable for radar applications. A reflectarray unit cell based on the use of a single varactor diode has been proposed and specifically optimized to provide wide angle reconfigurability features. At this purpose, the antenna has been properly designed by reducing the unit cell size, in order to achieve a large angular scanning. As a specific numerical example, a varactor loaded reflectarray element, embedded into a $0.46\lambda_0 \times 0.46\lambda_0$ cell at f_0=11.5 GHz, has been synthesized, obtaining a full phase tuning range of about 330°.

References

1. Fenn, A.J., Temme, D.H., Delaney, W.P., Courtney, W.E.: The Development of Phased-Array Radar Technology. Lincoln Laboratory Journal 12, 321–340 (2000)
2. Huang, J., Encinar, J.: Reflectarray Antennas. Wiley-IEEE Press (2008)
3. Sorrentino, R., Vincenti Gatti, R., Marcaccioli, L., Mencagli, B.: Electronic steerable MEMS antennas. In: 1st EuCAP, France (2006)
4. Hum, S.V., Okoniewski, M., Davies, R.J.: Realizing an electronically tunable reflectarray using varactor diode-tuned elements. IEEE Microw Wirel. Compon. Lett. 15, 422–424 (2005)
5. Riel, M., Laurin, J.J.: Design of an electronically beam scanning reflectarray using aperture-coupled elements. IEEE Trans. Antennas Propag. 55, 1260–1266 (2007)
6. Venneri, F., Costanzo, S., Di Massa, G.: Reconfigurable Aperture-Coupled Reflectarray Element Tuned by a Single Varactor Diode. Electronics Letters 48, 68–69 (2012)

7. Venneri, F., Costanzo, S., Di Massa, G., Borgia, A., Corsonello, P., Salzano, M.: Design of a Reconfigurable Reflectarray Based on a Varactor Tuned Element, EuCAP 2012, Prague (2012)
8. Venneri, F., Costanzo, S., Di Massa, G., et al.: Beam-Scanning Reflectarray Based on a Single Varactor-Tuned Element. Int. Journal of Antennas and Propagation 2012, Article ID 290285, 5 pages (2012), doi:10.1155/2012/290285
9. Venneri, F., Costanzo, S., Di Massa, G.: Design and validation of a reconfigurable single varactor-tuned reflectarray. IEEE Transactions on Antennas and Propagation (2013), doi: 10.1109/TAP.2012.2226229
10. Venneri, F., Costanzo, S., Di Massa, G., Angiulli, G.: An improved synthesis algorithm for reflectarrays design. IEEE Antennas Wirel. Propag. Lett. 4, 258–261 (2005)
11. Balanis, C.A.: Antenna Theory: Analysis and Design, 2nd edn. John Wiley and Sons (1997)
12. Costanzo, S., Venneri, F., Di Massa, G.: Bandwidth enhancement of aperture-coupled reflectarrays. IEE Electronics Letters 42, 1320–1321 (2006)
13. Venneri, F., Costanzo, S., Di Massa, G., Amendola, G.: Aperture-coupled reflectarrays with enhanced bandwidth features. Journal of Electromagnetic Waves and Applications 22, 1527–1537 (2008)
14. Venneri, F., Costanzo, S., Di Massa, G.: Bandwidth Behavior of Closely Spaced Aperture-Coupled Reflectarrays. International Journal of Antennas and Propagation 2012, Article ID 846017, 11 pages (2012), doi:10.1155/2012/846017

Age Differences in Computer Input Device Use: A Comparison of Touchscreen, Trackball, and Mouse

Ho-chuen Ng, Da Tao, and Calvin K.L. Or

Department of Industrial & Manufacturing Systems Engineering,
The University of Hong Kong, Pokfulam, Hong Kong
{hochuen,taodahuoniao,klor}@hku.hk

Abstract. This study examined age-related differences in user performance and preference with three computer input devices (mouse, touchscreen, and trackball) among older, middle-aged, and younger adults. Sixty-six participants were recruited and equally split into the three age groups. The results showed that age and input device had significant effect on task completion time and number of error. There were significant age-related performance differences in task completion time among the input devices. Ratings of users' preference indicated that the older adults preferred trackball to the mouse and touchscreen. Our findings suggest that the touchscreen could moderate only part of the age-related performance differences and conferred limited benefits to older adults. More research efforts are required to examine user characteristics, user perception on the use of the devices, and task requirements before we can determine the benefits of an input device for users in different ages.

Keywords: Age-related difference, input device, user performance.

1 Introduction

It is projected that the number of people aged 65 or older will increase from 524 million to 1.5 billion by the year 2050 worldwide [1]. This aging population is the fastest growing computer user group [2]. As the older adults are increasingly engaged in varied computer systems for social, recreational and healthcare purposes, more attention is being paid to the study of computer use by older adults, which is currently under-investigated.

One important issue facing the aging computer users is that almost any computer systems use a non-keyboard input device (e.g., mouse, trackball, and touchscreen) as one of the primary communication approaches between the users and the systems. Successful human-computer interaction may depend largely on users' ability to manipulate these input devices. However, given the age-related declines in functional and cognitive abilities, the older adults encounter many difficulties in using the devices [3]. Promisingly, the recently popularized touchscreen technology appears to hold the promise of addressing some of human-computer interaction issues related to age-dependent limitations [4]. Despite of its potential benefits, we know little about age-related differences in the use of the touchscreen compared with the use of other

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1015–1024.
DOI: 10.1007/978-3-642-36981-0_96

input devices. This paper presented preliminary results of a study that examined age-related performance differences with three widely used input devices: two indirect input devices (i.e., mouse and trackball) and a direct input device (i.e., touchscreen). The objective is to examine whether the touchscreen can moderate age-related performance differences, and to provide implications to optimize the selection and use of input devices for users with different ages.

1.1 Input Device Comparison

Non-keyboard input devices can be divided into direct input devices (require no movement translation between the user behavior and the device response, e.g., light pen and touchscreen) and indirect input devices (require a translation between the user behavior and the device response, e.g., mouse and trackball) [5]. Numerous studies have been conducted to compare user performance with different input devices. The majority of the empirical studies that compared mouse with other indirect input devices revealed that the mouse generally yielded superior performance in terms of speed, accuracy, and user preference, followed by the trackball [6-8]. Some early concerns also had much of an effort to examine the use of touchscreen [9]. Findings from several following studies suggested that the touchscreen appeared to confer more benefits than indirect input devices in some specific task situations [10-12]. For instance, Sears and Shneiderman found that the touchscreen yielded higher speed and fewer errors for larger size target selection in unimanual tasks compared with mouse, yet the users preferred mouse to touchscreen [10]. Forlines et al. suggested that the touchscreen outweighed the mouse in bimanual tasks performed on tabletops, while the mouse may be more appropriate for single-pointing tasks [12]. However, these generalizations and guidelines were drawn from studies where the input devices were evaluated for tasks in isolation and only by younger adults. The experience of the older adults as computer users may differ from that of the younger users when they interact with the input devices.

1.2 Age-Related Differences on Input Device Performance

Old adults are reported to have poor motion control with their functional and cognitive abilities decreased over times [3]. These changes may have negative impact on their ability to use input devices. A handful of studies examined the effects of age-related differences on mouse input performance. It was quite evident from the limited studies that the older adults experienced more difficulties in a wide-range of mouse control tasks compared with younger adults [13-15]. They committed more errors in target acquisition task [13]; yielded poorer performance in cursor control tasks [14]; and expressed higher ratings of perceived exertion after task performance [15]. A debate has since evolved with respect to the findings that the mouse, as the frequently used input device, might be problematic for the older adults.

Nowadays, touchscreen technology is becoming more available at a relatively low cost. As a direct input device, the touchscreen has many advantages over indirect input devices, such as easier hand-eye coordination, faster target acquisition, and less

cognitive demanding [5]. These advantages are expected to compensate for the limitation of other indirect input devices used by older adults [4]. A previous study suggested that a direct input device, light pen, could minimize the age-related performance differences and have generally more benefits to older adults [16]. However, it remains unknown whether the promising benefits of the light pen for older adults can generalize across other direct input devices, such as touchscreen. In addition, previous studies pointed out that direct input devices do not necessarily yield better performance in all circumstances, as they can have negative impact on computer tasks due to fatigue, inadvertent activation, or a lack of precision [5]. For example, Jastrzembski et al. compared task performance between direct and indirect input devices for different age groups [17]. They argued that the light pen yielded no superior performance than the mouse for older adults, and produced even worse performance at the first several task trials. Another study comparing task performance for a touchscreen and a rotary encoder found little benefits of the use of touchscreen by older adults [5]. Conversely, their results suggested that the task performance may be dependent on both the age of the user and the task requirements. Therefore, research efforts is needed to understand how the degree of directness of input devices and age-related differences interact with each other with regard to task performance; and how this interaction might vary upon different task scenarios.

2 Methods

2.1 Participants

Twenty-two younger (mean age=27 years), 22 middle-aged (mean age=56 years), and 22 older adults (mean age=69 years) were recruited in our study. Eligible participants were identified if they reported having normal or corrected-to-normal vision and healthy upper extremity function in a screening questionnaire. The study protocol was approved by the institutional review board of The University of Hong Kong and informed consent was obtained from all participants.

2.2 Materials and Tasks

The experiment was conducted in a university laboratory. Task scenarios were programmed using Visual Basic 2010 and performed on a Lenovo all-in-one touchscreen computer. The three input devices evaluated were a touchscreen, a trackball, and a mouse.

We designed three types of computer tasks that involved types of actions typically performed using non-keyboard input devices. Figure 1 presents a schematic representation of the three tasks. The point-and-click task asked participants to select a target square back and forth ten times between two squares on the computer screen. To select the target square, the participants needed to point to the target square (by moving the cursor with the mouse and trackball, or by using a finger) and click it (by pressing the left button on the mouse and trackball, or by pressing the target square on the touchscreen with a finger). The drag-and-drop task required participants to drag

six target squares into a fixed box at the center of the computer screen. To drag a target square into the box, the participants needed to select it and hold down the button or maintain pressure with the finger on the touchscreen until dragging the square into the box. The track-and-click task asked participants to track and click the eight squares by following a computer-generated, random sequence. The first target square in each task was indicated by yellow color. Once the participants accomplished all steps with the target square, a new target square with yellow color immediately appeared.

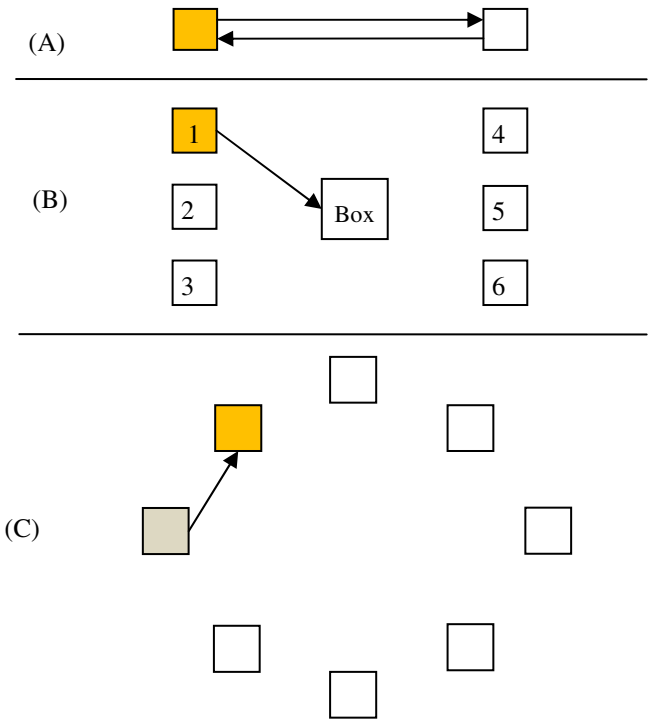


Fig. 1. Schematic representation of computer tasks: (A) point-and-click task; (B) drag-and-drop task; (C) track-and-click task

2.3 Procedures

Each participant started the experiment by signing a consent form and completing the screening questionnaire. The experimenter gave them brief instructions on how to use the input devices and how to complete the tasks. Following a brief practice session in order for them to get familiar with the input devices, the participants were presented with main experimental tasks. They were asked to click a start button on the center of the computer screen to initiate the task performance. The order of input devices and tasks were counterbalanced across participants for each age group. However,

participants completed a session of all three tasks with the same input device before they continue to use another input device. Each session lasted about 10 minutes with 3-minute breaks between the sessions. A total of approximately 40 minutes were required to complete the experiment. Participants were asked to respond as quickly and accurately as possible. After the completion of all tasks, an overall preference rating scale was administered. This scale asked participants to rank their input device preference on a 3-point scale (1 for the most preferred, 2 for the second preferred, and 3 for the least preferred).

2.4 Experimental Design

This study utilized a 2×3 mixed factorial design, with age as the between-participant variable, and input device as the within-participant variable. Dependent variables were task completion time, frequency of error, and users' preference. The task completion time referred to the total time to complete a task. Frequency of error was calculated by the number of errors made by participants in a task. Error in the point-and-click task and track-and-click task was defined as any target selection outside the actual target square, while error in the drag-and-drop task occurred when the participants failed to drag a target square into the box or drop the target during the dragging process.

Repeated measures analyses of variance (ANOVAs) were used to analyze the effects of age and input device on user performance and preference. Post hoc analyses were performed using Tukey-HSD procedure with p-value=0.05.

2.5 Hypotheses

The hypotheses shown below aimed to examine the effects of age and input device on the user performance and preference.

Hypothesis 1. The older adults will have longer task completion time and commit more errors than that of other age groups.

Hypothesis 2. There will be differences in task completion time and frequency of error among different input devices.

Hypothesis 3. The touchscreen will moderate the age-related differences on task completion time and frequency of error.

Hypothesis 4. The older adults will prefer touchscreen to the mouse and trackball.

Hypothesis 1 was based on previous evidence that the older adults encountered more difficulty in computer tasks; Hypothesis 2 was based on the nature of the difference among the input devices; and Hypothesis 3 and 4 were based on anticipated benefits of the touchscreen for the older adults.

3 Results

A total of 594 dataset (66 subjects×3 input devices×3 tasks) were collected in this experiment. Table 1 and 2 show the effects of age and input device on mean task completion time and mean frequency of error, respectively. The median for task

performance time is also presented in Table 1. The results are also illustrated graphically to facilitate interpretation of the tables. The sphericity assumption for the within-participant variable was violated. Thus, adjustment values of the degree of freedoms were applied.

Table 1. Mean task performance time (seconds) by age and input device in different tasks

	Older			Middle-aged			Young		
	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD
Point-and-click task									
Mouse	18.3	18.9	5.3	11.6	13.0	4.3	7.9	8.3	1.4
Touchscreen	8.2	9.0	2.8	7.7	8.5	2.9	5.3	5.7	1.3
Trackball	25.0	27.1	7.4	16.6	18.3	6.0	12.7	12.5	3.6
Drag-and-drop task									
Mouse	16.6	26.1	24.8	12.0	13.2	5.3	7.4	7.7	1.5
Touchscreen	16.0	18.9	7.9	10.0	13.6	9.6	8.2	9.8	4.8
Trackball	28.2	48.8	65.3	19.1	25.7	19.4	13.4	15.6	8.9
Track-and-click task									
Mouse	12.0	14.0	5.3	8.6	9.8	2.5	6.3	6.6	0.9
Touchscreen	7.4	7.3	1.8	6.4	6.7	2.0	4.3	4.7	1.2
Trackball	18.8	21.2	7.0	12.3	14.6	4.8	9.7	10.0	1.6

Table 2. Mean frequency of error by age and input device in different tasks

	Older		Middle		Young	
	Mean	SD	Mean	SD	Mean	SD
Point-and-click task						
Mouse	0.4	1.3	0.2	0.4	0.2	0.2
Touchscreen	0.3	0.5	0.1	0.2	0.2	0.4
Trackball	1.0	1.9	0.3	0.4	0.2	0.3
Drag-and-drop task						
Mouse	2.5	3.0	1.0	1.2	0.5	0.7
Touchscreen	3.1	3.0	1.8	2.9	1.6	3.6
Trackball	5.0	5.7	3.9	7.4	2.3	2.8
Track-and-click task						
Mouse	0.3	0.5	0.1	0.2	0.1	0.2
Touchscreen	0.0	0.1	0.1	0.5	0.3	0.5
Trackball	1.0	1.6	0.2	0.2	0.1	0.1

3.1 Effect of Age on User Performance

The ANOVA results indicated that there was a significant main effect of age on task completion time, $F(2, 63) = 13.931$, $p<0.001$, and on frequency of error, $F(2, 63) = 4.785$, $p=0.012$, respectively (see figure 2). Post hoc Tukey HSD multiple comparisons tests showed that the older adults were slower than the middle-aged ($p<0.01$) and younger adults ($p<0.01$), and yielded more errors than the younger adults ($p<0.5$).

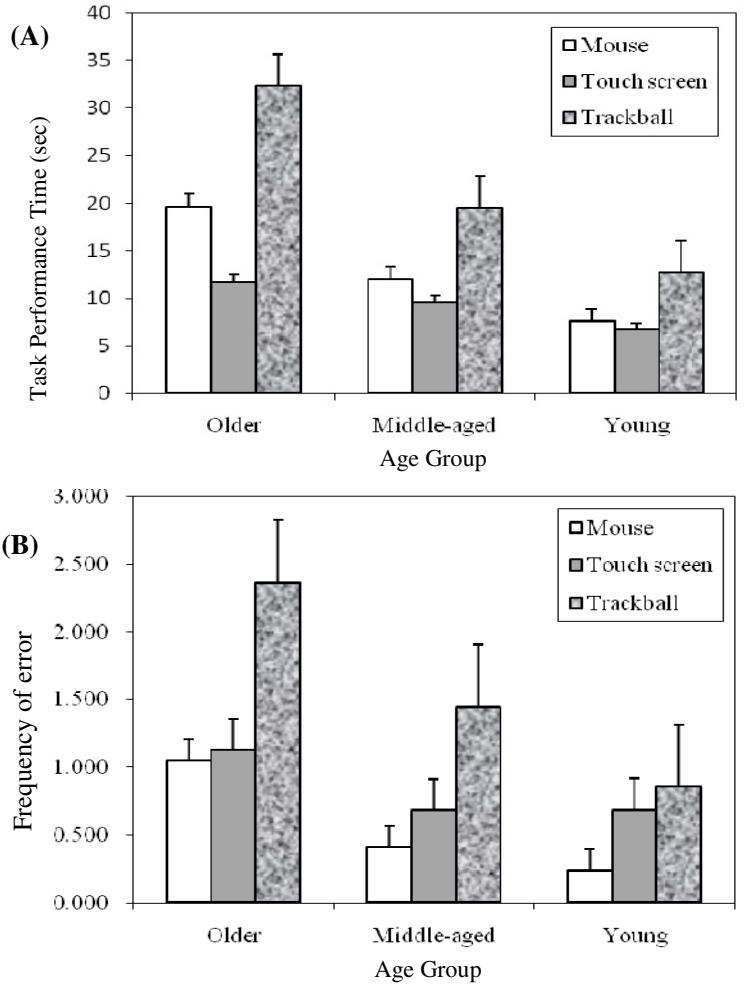


Fig. 2. Age-related differences in performance: (A) Device by age interaction for mean task completion time; (B) Device by age interaction for mean frequency of error

3.2 Effect of Input Device on User Performance

There was a significant main effect of input device on task completion time, $F(1.119, 70.482) = 43.766$, $p<0.001$, and frequency of error, $F(1.462, 92.091) = 10.478$, $p<0.001$. The touchscreen was the slowest input device, followed by the mouse and trackball (p's<0.001). Using the trackball yielded the highest frequency of error compared with other input devices (p's<0.05). The results also indicated that the interaction between age and input device on task completion time was significant, $F(2.238, 70.482) = 5.332$, $p=0.005$. The older adults had comparable speed with other age groups using the touchscreen, while they were much slower than other age groups

using the mouse and trackball (Figure 2A). The frequency of error, however, had no differences across input devices for the age groups (Figure 2B).

3.3 User Preference

Table 3 shows the distribution of the users' device preference. There were significant differences in the users' device preference among the three age groups, $F(2, 63) = 4.99$, $p=0.001$. With regard to the older adults, the trackball was most preferred (45%), followed by the touchscreen (32%) and mouse (23%). Sixty-four percentages of the middle-aged adults preferred using the touchscreen to using the mouse or trackball, while 27% of them rated mouse as the most preferred input device. The younger adults had comparably favorable preference for the mouse (45%) and trackball (45%). They also had least preference on the trackball (86%).

Table 3. Distribution of user preference by age (indicated by percentage of the participants)

	Most Preferred			Second Preferred			Least Preferred		
	Mouse	Touchscreen	Trackball	Mouse	Touchscreen	Trackball	Mouse	Touchscreen	Trackball
Older	23%	32%	45%	41%	36%	23%	36%	32%	32%
Middle-aged	27%	64%	9%	41%	23%	36%	32%	14%	55%
Young	45%	45%	10%	50%	45%	5%	5%	9%	86%

4 Discussion

This study examined age-related performance differences on the use of three input devices. This proposition was largely driven by the rapid diffusion of computer systems designed for improving older adults' well-beings in their workplace, healthcare, and personal life [3, 18]. The results from this experiment provided evidence that age-related differences and the nature of input devices accounted for the observed differences in task performance. The touchscreen could moderate part of the age-related performance differences.

Hypothesis 1, which tested the effect of age, was supported. The older adults were slower and had less accurate performance than younger participants regardless of input devices and tasks. This finding was intuitive and consistent with previous studies [13-16]. As commonly acknowledged, the reason could be largely attributed to the older adults' deteriorative abilities in spatial perception, memory, and motor skills [3]. Another reason may lie in the fact that the older adults were less educated and less experienced in computer tasks. Lack of computer knowledge and unfamiliarity with the input devices may therefore result in poor computer task performance.

Hypothesis 2, which examined the effect of input device, was supported. Our results indicated that using the touchscreen was significant faster than that of the

mouse and trackball. The touchscreen and mouse yielded fewer errors than the trackball. This findings was partly incongruent with previous studies [9, 11], reporting that the touchscreen was the fastest but the least accurate input devices. This discrepancy may be accounted by the differences in task design between previous studies and our study. For example, our study only evaluated computer tasks in a stationary environment, while Yau et al. study involved tasks in both motion and stationary situations by simulating workplace environment on a ship [11].

With regard to Hypothesis 3, the significant interaction between age and input device indicated that although the older adults were slower using the mouse or trackball, they had comparably fast speed with other age groups using the touchscreen. The result was similar to that in Charness et al. study, which suggested that the direct input device, light pen, could minimize age-related differences in task completion time [16]. The reason could be attributed to the aforementioned advantages of the direct input device, which make the devices less cognitively demanding in task performance [5]. This merit is especially beneficial to older adults who were susceptible to the age-dependent declines in cognitive ability. However, this interaction was not significant on frequency of error. Therefore, hypothesis 3 was not fully supported. It meant that the touchscreen could moderate only part of the age-related differences in task performance. One possible explanation is that more physical efforts and unnecessary body movement may be required, and more fatigue may occur when the older adults use the touchscreen. This could increase the likelihood of error commitment.

Hypothesis 4, which examined the effect of age on users' device preference, was not supported. Interestingly, although the older adults were faster and made fewer errors with the touchscreen, they preferred the trackball to other input devices. As mentioned above, the differences in users' perception on using different input devices may serve as one possible reason of this counter-intuitive result. It is true that using a trackball requires less physical efforts through rolling a ball held in a socket than that of holding and shuttling a mouse across the working area or stretching the arm to point and press the touchscreen with a finger. Moreover, the trackball is less likely to be affected by the limited range of wrist movement than other input devices. Therefore, using the trackball may reduce strains on the hand and wrist, leading to better user experience. This explanation may be confirmed by observing users' task behaviors and by assessing users' subjective perception, such as perceived exertion, fatigue, mental efforts, and ease of use in using the input devices.

5 Conclusion

This study provided new insight into the evaluation of age-related performance differences in computer tasks by examining three widely used input devices among older, middle-aged, and younger adults. The experimental results showed that older adults were likely to be affected by their age-related declines in functional and cognitive abilities, resulting in poor task performance. The results also confirmed that the touchscreen could moderate part of the age-related performance differences and

confer benefits, though limited, to older adults. However, the findings also indicated that no input device could yield universally better performance in both of user performance and user preference. The selection of input device may be largely optimized by taking the user's characteristics, their subjective perception, and the task requirements into account.

References

1. WHO and US National Institute of Aging: Global health and ageing (2011)
2. Czaja, S., Lee, C.: The Impact of Aging on Access to Technology. Universal Access in the Information Society 5, 341–349 (2007)
3. Czaja, S., Lee, C.: Information Technology and Older Adults. In: Sears, A., Jacko, J. (eds.) Handbook of Human-Computer Interaction, pp. 777–792. Lawrence Erlbaum and Asso, Mahwah (2007)
4. Caprani, N., O'Connor, N.E., Gurrin, C.: Touch Screens for the Older User. In: Auat Cheein, F. (ed.) Assistive Technologies, pp. 95–118 (2012)
5. Rogers, W.A., Fisk, A.D., McLaughlin, A.C.: Touch a Screen or Turn a Knob: Choosing the Best Device for the Job. Human Factors 47(2), 271–288 (2005)
6. Card, S.K., English, W.K., Burr, B.J.: Evaluation of Mouse, Rate-controlled Isometric Joystick, Keys, and Text Keys for Text Selection on a CRT. Ergonomics 21, 601–613 (1978)
7. Epps, B.W.: A Comparison of Cursor Control Devices on a Graphics Editing Task. In: The 31st Human Factors and Ergonomics Society Annual Meeting, Santa Monica, CA (1987)
8. Mackenzie, I.S., Sellen, A., Buxton, W.: A Comparison of Input Device in Elemental Pointing and Dragging Tasks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM Press, New York (1991)
9. Albert, A.: The Effect of Graphic Input Devices on Performance in a Cursor Position Task. In: The 26th Annual Meeting of the Human Factors Society, Santa Monica, CA (1982)
10. Sears, A., Shneiderman, B.: High Precision TouchScreens: Design Strategies and Comparisons With a Mouse. Int. J. Man. Mach. Stud. 34, 593–613 (1991)
11. Yau, Y., Hwang, S., Chao, C.: Effects of Input Device and Motion Type on a Cursor-positioning Task. Percept. Mot. Skills 106, 76–90 (2008)
12. Forlines, C., et al.: Direct-touch vs. Mouse Input for Tabletop Displays. In: The SIGCHI Conference on Human Factors in Computing Systems. ACM Press, New York (2007)
13. Walker, N., Philbin, D.A., Fisk, A.D.: Age-Related Differences in Movement Control: Adjusting Submovement Structure to Optimize Performance. J. Gerontol. B. Psychol. Sci. Soc. Sci. 52, 40–52 (1997)
14. Smith, M., Sharit, J., Czaja, S.: Aging, Motor Control, and the Performance of Computer Mouse Tasks. Human Factors 41, 389–396 (1999)
15. Chaparro, A., et al.: The Impact of Age on Computer Input Device Use: Psychophysical and Physiological Measures. Int. J. Ind. Ergonom. 24, 503–513 (1999)
16. Charness, N., et al.: Light Pen Use and Practice Minimize Age and Hand Performance Differences in Pointing Tasks. Human Factors 46, 373–384 (2004)
17. Jastrzembski, T., et al.: Input Devices for Web Browsing: Age and Hand Effects. Universal Access in the Information Society 4, 39–45 (2005)
18. Melenhorst, A., Rogers, W., Bouwhuis, D.: Older Adults' Motivated Choice for Technological Innovation: Evidence for Benefit-driven Selectivity. Psychol. Aging 21, 190–195 (2006)

Color-Concept Associations among Chinese Steel Workers and Managerial Staff

Heller H.L. Wang and Calvin K.L. Or*

Department of Industrial and Manufacturing Systems Engineering,
The University of Hong Kong,
Pokfulam, Hong Kong
{whliang,klor}@hku.hk

Abstract. Previous studies have tested population stereotypes for color without considering the specific environments under which color associations are critical for human safety, or whether color associations vary on the basis of various occupations. In this study, a questionnaire survey was administered to 139 steel workers and 74 managerial staff in two industrial companies in China to examine the color associations among occupational groups and to compare population stereotypes for color. Participants were presented with 16 concepts and were asked to select one of 10 colors to represent each concept. Chi-square tests were conducted on each of the sixteen concepts and the results indicated that each concept was significantly associated with at least one color in the two occupational groups. The study also revealed the differences in the color associations produced by participants from different occupations and cultures (Chinese steel workers, managerial staff, Hong Kong Chinese, Yunnan Chinese, and Americans).

Keywords: Color association, population stereotype, human factors design.

1 Introduction

Color plays an important role in the visual search process and recognition [1], especially when it concerns human safety, such as on warning labels and control panels. Although color should not be used as the primary or the only coding scheme for communicating hazards [2], color displays attract attention more effectively than achromatic displays [3]. A number of studies have focused on the associations between various colors and the perception of hazards and risks [4-6]. The findings of such studies can inform the design of manufacturing processes and workstations to promote workers' safety. However, as can be seen in three prior studies, color-concept associations show diversity among population groups [7, 8]: Bergum and Bergum's study [9] that tested the color associations with U.S. participants,

* Corresponding author.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1025–1031.
DOI: 10.1007/978-3-642-36981-0_97

Courtney's study [10] that examined the color associations with Yunnan Chinese participants, and Chan and Courtney's study [7] that studied Hong Kong Chinese participants.

While these previous studies focused on the diversity of population stereotypes for color, there are two ideas that should be noted when applying these color association results to specific industries. First, some of these research data were obtained several years ago and may not reflect developments in society. Color associations may alter as a result of changing circumstances, such as globalization, which cause color associations to become more similar across different culture environments. Second, the study conducted by Courtney [10] included participants from so many occupational backgrounds that it failed to represent the relationships between color association and occupation. The participants in the study conducted by Bergum and Bergum [9] and that by Chan and Courtney [7] were university students, thus their findings cannot be applied to specific occupations. With the variability and diversity of the population stereotype for color, it is necessary to test the population stereotype within specific occupational groups who are exposed to warnings.

This study was designed to find out how colors and concepts were associated with each other in industrial workshops and clerical offices and to provide suggestions for manufacturers to design warning labels and control panels for workshops. Two experiments were carried out in mainland China with two occupational groups, steel workers and managerial staff, as participants. Steel workers were selected because they work in highly complex environments with rigorous safety requirements, thus fully demonstrating the importance of color-concept associations. In addition, improving the understanding of color-concept associations among managerial staff will help to increase efficiency in offices and reduce potential risks in workshops.

By including the data from three previous color-association studies conducted with Hong Kong Chinese [7], Yunnan Chinese [10], and U.S. participants [9], together with the data collected from steel workers and managerial staff in Hebei Province, this study provides a comparison of the most frequent color associations among these five population groups.

2 Design and Procedure

A questionnaire survey was conducted to examine the associations between 10 colors (black, blue, gray, green, orange, pink, purple, red, white, and yellow) and 16 concepts (caution, cold, danger, go, hard, hot, normal, off, on, potential hazard, radiation hazard, safe, soft, stop, strong, and weak) adapted from Chan and Courtney's study [7]. The questionnaire was administered as a paper-and-pencil test, in which participants were asked to select their preferred one from the 10 colors to match each of the 16 concepts. If no color was considered appropriate, then participants were told to select the additional option, "None of the ten colors is suitable."

Two experiments were conducted in two industrial companies in Hebei Province of China to examine population stereotypes for color within two groups: steel workers from a steel company and managerial staff from a soybean processing company. Color blindness tests were not conducted on participants because the annual health

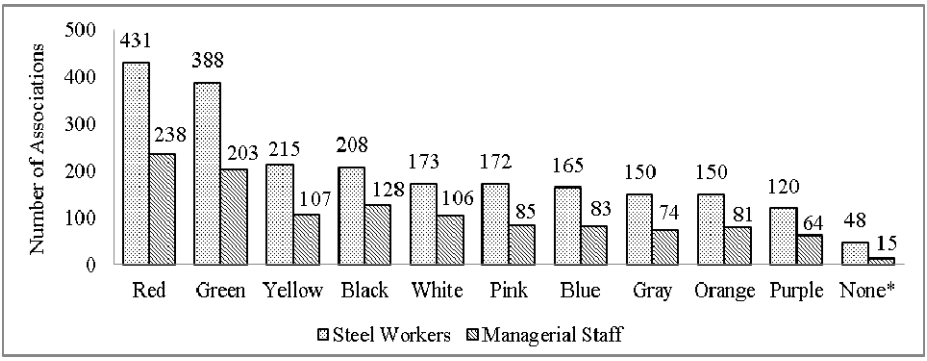
reports of the two companies confirmed that none of the staff members were color blind. The chi-square test was applied to determine whether there were significant color associations for each concept.

3 Results

In *Experiment 1*, 160 questionnaires were distributed via the human resources department to steel workers, and 139 valid responses were received (response rate: 86.9%). The participants' ages ranged from 21 to 30 (55.4%), 31 to 40 (36.7%), 41 to 50 (7.2%), and above 50 (0.7%). All of the participants worked in the steel workshops with duties such as machine maintenance, equipment testing, steel rolling, and die designing.

In *Experiment 2*, 80 questionnaires were sent to managerial staff in the soybean processing company, and 74 valid responses were received (response rate: 92.5%). The participants' ages ranged from 21 to 30 (55.4%), 31 to 40 (43.2%), and 41 to 50 (1.4%). The participants were employed in office-related jobs including management, secretarial, human resources, finance, accounting, and training.

Steel workers and managerial staff produced 2,220 and 1,184 associations, respectively. Figure 1 shows the number of associations for each color made by the two occupational groups. To verify the associations between colors and concepts, chi-square tests for each of the sixteen concepts were performed, and the results indicated that each concept was associated with at least one color at the 0.001 level of significance ($df = 10$) (e.g., for the concept "go" to the managerial staff, the calculated statistic value at the 0.001 level was $\chi^2 = 448.35$, which was much greater than the χ^2 table value 29.59, showing colors were strongly associated with the concept "go").



*Note: "None" represents "None of the ten colors is suitable".

Fig. 1. Total number of associations for each color by steel workers and managerial staff

Table 1 shows how many participants associated each color with each concept, with the percentages for steel workers without brackets and the percentages for managerial staff inside the brackets.

Table 1. Percentages of associations between each color and each concept

Concept	Color										
	Red	Green	Yellow	Black	White	Pink	Blue	Gray	Orange	Purple	None*
Go	2.2 (1.4)	74.6 (79.7)	1.4 (0.0)	0.7 (4.1)	3.6 (1.4)	2.9 (4.1)	5.8 (2.7)	0.7 (1.4)	3.6 (2.7)	4.3 (2.7)	0.0 (0.0)
Stop	59.0 (64.9)	2.9 (5.4)	7.9 (6.8)	10.1 (9.5)	1.4 (1.4)	2.9 (2.7)	1.4 (2.7)	7.2 (5.4)	3.6 (0.0)	0.7 (1.4)	2.9 (0.0)
Hot	50.4 (59.5)	2.9 (1.4)	3.6 (5.4)	2.9 (1.4)	4.3 (0.0)	6.5 (9.5)	3.6 (1.4)	0.7 (0.0)	23.0 (18.9)	2.2 (2.7)	0.0 (0.0)
Cold	0.7 (0.0)	2.2 (1.4)	5.0 (0.0)	5.0 (1.4)	31.7 (29.7)	1.4 (2.7)	39.6 (52.7)	8.6 (9.5)	0.7 (0.0)	3.6 (1.4)	1.4 (1.4)
On	12.2 (10.8)	51.8 (46.0)	2.2 (1.4)	3.6 (6.8)	10.1 (24.3)	3.6 (2.7)	6.5 (4.1)	2.2 (1.4)	1.4 (1.4)	4.3 (0.0)	2.2 (1.4)
Off	46.8 (39.2)	6.5 (6.8)	0.7 (4.1)	15.1 (25.7)	3.6 (8.1)	0.7 (1.4)	8.6 (1.4)	11.5 (6.8)	2.2 (1.4)	2.2 (2.7)	2.2 (2.7)
Normal	1.4 (0.0)	50.4 (46.0)	5.0 (4.1)	0.7 (1.4)	12.9 (18.9)	3.6 (4.1)	12.9 (16.2)	5.0 (4.1)	0.7 (1.4)	4.3 (2.7)	2.9 (1.4)
Danger	60.4 (55.4)	3.6 (0.0)	18.7 (17.6)	4.3 (8.1)	0.0 (0.0)	0.7 (0.0)	0.7 (1.4)	3.6 (2.7)	5.0 (9.5)	2.9 (4.1)	0.0 (1.4)
Safe	1.4 (1.4)	60.1 (74.3)	2.2 (0.0)	3.6 (2.7)	9.4 (6.8)	2.9 (5.4)	7.2 (5.4)	6.5 (0.0)	0.7 (1.4)	5.1 (2.7)	0.7 (0.0)
Caution	33.8 (44.6)	0.0 (0.0)	51.1 (41.9)	6.5 (4.1)	0.0 (0.0)	1.4 (0.0)	2.2 (2.7)	0.7 (1.4)	3.6 (2.7)	0.7 (1.4)	0.0 (1.4)
Soft	2.9 (0.0)	2.9 (2.7)	4.3 (5.4)	2.9 (0.0)	14.5 (16.2)	50.0 (54.1)	4.3 (2.7)	3.6 (8.1)	6.5 (6.8)	6.5 (1.4)	1.4 (2.7)
Hard	3.6 (2.7)	2.2 (0.0)	3.6 (0.0)	48.6 (58.1)	3.6 (6.8)	0.7 (0.0)	7.2 (2.7)	20.3 (16.2)	1.4 (2.7)	4.3 (8.1)	4.3 (2.7)
Weak	0.7 (1.4)	5.0 (1.4)	9.4 (20.3)	6.5 (0.0)	20.9 (24.3)	23.0 (12.2)	5.0 (6.8)	12.9 (21.6)	8.6 (4.1)	1.4 (5.4)	6.5 (2.7)
Strong	14.4 (23.0)	5.0 (6.8)	2.9 (1.4)	24.5 (29.7)	1.4 (0.0)	2.9 (2.7)	6.5 (8.1)	2.2 (1.4)	13.7 (12.2)	20.9 (12.2)	5.8 (2.7)
Potential Hazard	4.3 (5.4)	5.0 (1.4)	18.0 (16.2)	2.9 (8.1)	1.4 (1.4)	15.1 (10.8)	2.9 (1.4)	14.4 (12.2)	26.6 (27.0)	6.5 (16.2)	2.9 (0.0)
Radiation Hazard	15.8 (12.2)	5.0 (1.4)	18.7 (20.3)	12.2 (12.2)	5.8 (4.1)	5.8 (2.7)	4.3 (0.0)	7.9 (8.1)	6.5 (17.6)	16.5 (21.6)	1.4 (0.0)

*Note: "None" represents "None of the ten colors is suitable"; the numbers without brackets are for steel workers; the numbers inside the brackets are for managerial staff.

Nine of the 16 concepts tested in our study (i.e., go, danger, safe, stop, on, caution, hot, off, and cold) were also examined in prior studies with Hong Kong Chinese [7], Yunnan Chinese [10], and U.S. participants [9]. The colors most-frequently used to

represent each of these nine concepts are shown in Table 2, indicating differences in the population stereotypes for color. Table 2 also presents the average percentage of color associations for these nine concepts, thus reflecting the strength of the color associations among the five groups.

Table 2. Percentages of the most frequent color associations for Chinese and U.S. participants

Concept	Steel workers		Managerial staff		Hong Kong Chinese*		Yunnan Chinese		U.S. participants	
	Color	%	Color	%	Color	%	Color	%	Color	%
Go	Green	74.6	Green	79.7	Green	62.6	Green	44.7	Green	99.2
Danger	Red	60.4	Red	55.4	Red	63.0	Red	64.7	Red	89.8
Safe	Green	60.1	Green	74.3	Green	38.2	Green	62.2	Green	61.4
Stop	Red	59.0	Red	64.9	Red	66.4	Red	48.5	Red	100
On	Green	51.8	Green	46.0	Green	23.8	Green	22.3	Red	50.4
Caution	Yellow	51.1	Red	44.6	Red	40.2	Yellow	44.8	Yellow	81.1
Hot	Red	50.4	Red	59.5	Orange	28.2	Red	31.1	Red	94.5
Off	Red	46.8	Red	39.2	White	23.3	Black	53.5	Blue	31.5
Cold	Blue	39.6	Blue	52.7	Blue	22.5	White	71.5	Blue	96.1
Average		54.9		57.4		40.9		49.3		78.2

*Note: Hong Kong Chinese data retrieved from Chan and Courtney [7], U.S. data retrieved from Bergum and Bergum [9], and Yunnan Chinese data retrieved from Courtney [10].

4 Discussion

The results revealed that red and green yielded 431 (19.4%) and 388 (17.5%) associations for steel workers, respectively, and 238 (20.1%) and 203 (17.1%) associations for managerial staff, respectively, thus supporting the concept that primary colors yield the largest number of associations [9]. The percentage of associations with the color green was 11.3% for Hong Kong Chinese, which is less than that for steel workers (17.5%) and managerial staff (17.1%). This difference probably arises because steel workers and managerial staff work in industrial companies and thus pay more attention to green as a symbol representing "go" (74.6%, 79.7%), "safe" (60.1%, 74.3%), and "on" (51.8%, 46.0%). These percentages are all larger than the 62.6%, 38.2%, and 23.8% reported for Hong Kong Chinese university students. In this study, there is no color-concept association with the percentage score equal to or larger than 85%, which is considered the minimum level for determining a population stereotype [9]. Purple had the lowest frequency of association with any concept among both steel workers (5.4%) and managerial staff (5.0%). Ergonomists and engineering designers are suggested not to use purple in equipment and warning-label designs for safety recognition.

As illustrated in Table 1, none of the steel workers supported the associations between white-danger, white-caution, and green-caution. For managerial staff, 25 pairs received a score of zero, indicating that no managerial staff supported the following color associations: red-cold, red-normal, red-soft, yellow-go, yellow-cold, yellow-safe, yellow-hard, orange-stop, orange-cold, blue-radiation hazard, green-danger, green-caution, green-hard, pink-danger, pink-caution, pink-hard, purple-on, gray-hot, gray-safe, black-soft, black-weak, white-hot, white-danger, white-caution, and white-strong. This diversity may be due to the difference between the work environments of the two groups: for managerial staff, the associations between colors and concepts are limited to those relevant to the office environment and daily equipment (such as computers, printers, scanners, and water dispensers), whereas steel workers need to master many different machines and remember the exact meanings of controls, displays, and other entities with various colors.

The comparison of color associations for the nine concepts examined in previous studies (Table 2) showed that the color-concept pairs identified by the two occupational groups in this study were identical except for the concept "caution", which steel workers associated with yellow and managerial staff with red. The four strongest associations for the five population groups were green-go, red-danger, green-safe, and red-stop, with more than half of the participants in each study choosing at least two of these four pairs. The strong associations with red and green arise because they concur with people's daily experiences [11].

The four Chinese samples chose green to represent "on", where the U.S. group chose red (see Table 2). Red was the third most-frequent color chosen to represent "on" for managerial staff and the second most-frequent for the other three Chinese groups. This result further demonstrates that green and red do not have completely opposite meanings for different population groups. Ergonomists and production engineers should be aware of this finding when designing equipment control panels or warning labels for international customers. For Chinese users, it appears that green-on association is primarily perceived.

Courtney [10] found that 71.5% of Yunnan Chinese participants selected white to represent "cold". However, steel workers and managerial staff, similar to the Hong Kong Chinese and U.S. participants, associated "cold" with blue. One possible explanation is that American culture has influenced Chinese people's perception over the past three decades, as a result of increasing globalization. Another discovery was that the five groups chose four different colors to represent "off": steel workers and managerial staff (red-off, 46.8% and 39.2%), Hong Kong Chinese (white-off, 23.3%), Yunnan Chinese (black-off, 53.5%), and U.S. participants (blue-off, 31.5%). Equipment manufacturers and engineers need to pay more attention to which color best represents the common concept "off" for their customers.

5 Conclusions

The comparison between Chinese and U.S. participants revealed a consensus for four common concepts: green for "go", green for "safe", red for "danger", and red for "stop". It appears that these color associations can be used universally, although more

research is needed to confirm their use in different environments. Purple was the least-frequently used color, thus purple should be avoided in machine control design and information recognition. Because Chinese groups generally use green to represent "on", it is sensible for workshops to use the opposite color, red, to represent "off" in China. However, the use of color coding alone has limitations in information communication, especially in complex workshop environments. It is suggested that color coding should be used together with other coding schemes, such as location, shape, size, and auditory signals to convey information in specific environments. Using suitable colors for computer controls and warning labels helps to achieve safe and efficient production processes in workshops. After the optimal colors are confirmed to represent certain concepts, training among all the staff is necessary to strengthen the color-concept associations.

References

1. Christ, R.E.: Review and Analysis of Color Coding Research for Visual Displays (in Aircraft). Human Factors 17, 542–570 (1975)
2. Wogalter, M.S., Laughery, K.R.: Warnings and Hazard Communications. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, pp. 889–911. John Wiley & Sons, New York (2006)
3. Laughery, K.R., Young, S.L., Vaubel, K.P., Brelsford, J.W.: The Noticeability of Warnings on Alcoholic Beverage Containers. Journal of Public Policy & Marketing 12, 38–56 (1993)
4. Braun, C.C., Silver, N.C.: Interaction of Signal Word and Colour on Warning Labels: Differences in Perceived Hazard and Behavioural Compliance. Ergonomics 38, 2207–2220 (1995)
5. Leonard, S.D.: Does Color of Warnings Affect Risk Perception? International Journal of Industrial Ergonomics 23, 499–504 (1999)
6. Or, K.L., Chan, H.S.: The Effects of Background Color of Safety Symbols on Perception of the Symbols. In: The 3rd International Conference on Applied Human Factors and Ergonomics, Miami, FI (2010)
7. Chan, H.S., Courtney, A.J.: Color Associations for Hong Kong Chinese. International Journal of Industrial Ergonomics 28, 165–170 (2001)
8. Jacobs, L., Keown, C., Worthley, R., Ghymn, K.I.: Cross-Cultural Colour Comparisons: Global Marketers Beware! International Marketing Review 8, 21–30 (1991)
9. Bergum, B.O., Bergum, J.E.: Population Stereotypes: an Attempt to Measure and Define. In: Proceedings of the 25th Human Factors and Ergonomics Society Annual Meeting, Santa Monica, CA, pp. 662–665 (1981)
10. Courtney, A.J.: Chinese Population Stereotypes: Color Associations. Human Factors 28, 97–99 (1986)
11. McCormick, E.J.: Human Factors Engineering. McGraw-Hill, New York (1970)

Cycle of Information Retrieval through Color in e-Commerce: Store Choice

M. Rosario Fernández Falero, Libertad Sánchez Gil, and Luis V. Gordillo Tapia

Facultad de Ciencias de la Información y Documentación de la Universidad de Extremadura
Plazuela de Ibn Marwan s/n, 06001 Badajoz, Spain
rferfal@alcazaba.unex.es, libersanchez@gmail.com,
luisvi1976@hotmail.com

Abstract. This paper addresses a forgotten aspect in electronic commerce: Social Information Retrieval and Culture.

Researches on this topic reference marketing and user studies (person to person), specifically relating to Human Behavior Information, which is the study of information retrieval focused on user needs, particularly in e-commerce. The key to selling is that the client finds the information that indicates that the product being sold is the one they want, and consequently, they buy it when they need it.

The main conclusion is that color is an essential mechanism in the process of choosing an online store and must be determined through a holistic process.

Keywords: color, electronic commerce, information retrieval, choice store.

1 Introduction

This article addresses a forgotten aspect in electronic commerce: Social Information Retrieval and Culture.

Researches on this topic reference marketing and user studies (person to person), specifically relating to Human Behavior Information [15], which is the study of the information retrieval focused on user needs. E-commerce is based on information retrieval of the online store to sell a product [4]. The key to selling is that the client finds the information that indicates the product being sold is the one they want, and consequently, they buy it when they need it.

The aim of this article is to show a different way to research / do research, adding to user studies the experience obtained or apprehended socially, through common-knowledge, about a particular color / particular colors. How non verbal communication between buyer and seller on online selling is being influenced by this cultural experience.

Colors [10] are the elements that faster and longer attract the attention of whoever looks at them. Color can be perceived through vision and conversely to words, which are located in a particular place and need the eye's focus to be understood. In fact, color can contribute to the understanding of the words, if it is properly used.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1033–1040.
DOI: 10.1007/978-3-642-36981-0_98

Besides, one must keep in mind that color is the element that most affects the emotional memory —it can transmit sensations, memories, etc., without people perceiving it consciously, and can serve as a factor that guides the buyer in the desired direction.

Multidisciplinary techniques must be used to choose the most appropriate color:

Web Design (colors combination or how to display these on a computer monitor).
Psychology, Marketing, Anthropology, Communication and Information Science.

Therefore, the choice of colors that will make up the webpage will be a key issue at the time of its design, marking a turning point in the process of deciding if the site is most suitable for the purchase of the product sought. It would establish a positive or negative sale of the product.

1.1 Theoretical Model

Society and Culture contain a Global Knowledge, which includes messages transmitted by color, and consequently, the information they contain. Color being an exponent of this society, for this reason this information will be understood and comprehended by the community to which you want to sell.

Figure 1 shows the recovery cycle of information through color in both directions, because the process is cyclical.

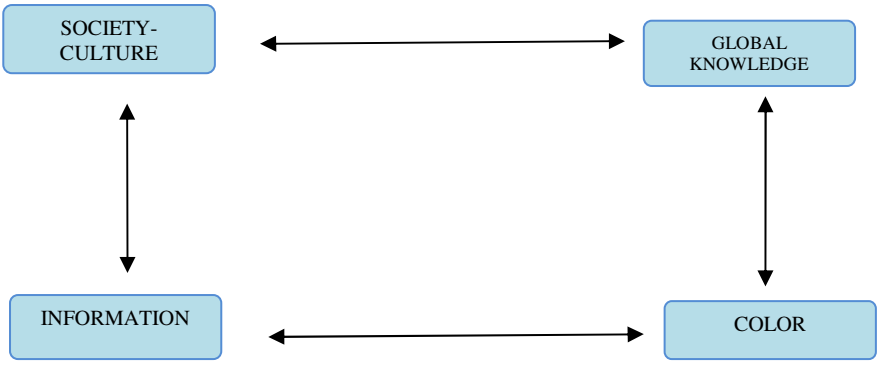


Fig. 1. Cycle of Information Retrieval through Color

1.2 Hypotheses

Hypothesis 1: The study of Society and Culture takes us to the Global Knowledge.

Hypothesis 2: The suitable color for selling a product is determined by the global knowledge of a Society.

Hypothesis 3: Color is a transmitter of information between vendor and client.

2 Methods

We set up a comparison between different colors, thus instituting the existing common knowledge about them in society through the work of Paola Fraticola [5], Fernando Parodi Gastañeta [12] and Javier Nó [10].

RED promotes emotions and instincts, it symbolizes passion. It creates a bad work environment, as it facilitates discussion and debate. It represents danger on traffic signals or alarms.

YELLOW AND ORANGE are colors with very similar characteristics. They express happiness and euphoria. They have a great visual power, and if they are exposed to the human eye for a long time may cause fatigue. They stimulate the appetite and incite to action and movement. They are aggressive colors.

WHITE gives a feeling of cleanliness, purity. It offers a sense of emptiness, of dispersion, a situation which is not conducive to the concentration necessary for buying. Although in the Japanese Buddhists' opinion white is associated with eternal knowledge, this is a minority view in the global sense, specifically in Europe and the West. From an anthropological point of view, while as far as the Catholic Church is concerned, the color white symbolizes the sky, the truth unveiled, Buddhism considers it represents the transcendent wisdom.

BLACK. This case is interesting because on the one hand color black is identified with elegance, luxury and quality, which obviously would be a good premise for any commercial website. However, this color carries certain connotations that are not very positive for e-commerce. This color transmits darkness, opacity; it is the symbol of silence, mystery, and may cause a negative impact on the human eye: rejection, restlessness, among other things.

GREEN is the color of indifference calm: it does not convey happiness, sadness or passion. According to thework of Carolina Vera entitled "Generation of outside advertising impact through the use of the principles of visual neuromarketing" [14], this color stimulates the development of new ideas, encourages creativity and innovation, giving rest and toning up.

BLUE evokes authority, effectiveness, consistency and safety. Also, following the work of Carolina Vera, color blue facilitates evaluation and control. And it also arranges the development of methodologies.

Omar Lara Álvarez [2] considers that the color blue is very appropriate because it evokes relaxation. With the color blue there is a sense of tranquility and regeneration of physical vitality and mental health is achieved. In addition (14), blue light decreases the muscular and nervous tension, while the client feels calm and concentrated.

From different color-related studies and e-commerce, we note the article by Simon Lee and V. Srinivasan (Chino) Rao [9] and the study of e-learning [6]. Through different methodologies (user studies and Benchmarking) they concluded that the color blue is the most used and more socially acceptable for electronic selling.

The review of previous works has shown that the study methodology, determining blue as the most used color, as well as the answer of users to blue, has been through individualized counting.

The method used in this work starts from the theory of knowledge [1] as a source for the development of market research. It is/was based on relativism, a philosophy that encompasses the common knowledge of a group of people depending on the experience of the individuals who make up the community.

There are three worlds of experience: nature, culture-society and subjectivity.

Nature would be the object of study, commercial websites (Benchmarking).

Subjectivity analyzes the individual (user studies).

Culture and society would include people who form a group or community, and establish the possibility of specialization, global knowledge, as well as their mechanisms of communication and interaction among its members.

3 Results

In fig. 1 we show that these three sections: Benchmarking (nature), Theory of Knowledge (culture and society) and User Studies (one's own subjectivity).

They have been analyzed independently; however, the section of culture and society has been associated actively with the other two blocks, and as a function of this, we analyze the object of study, if the environment is modified, society and culture will be changed, then behavior of the individual will be affected in front of an object. It is, consequently, Culture and Society who determines our response to a particular item, in this case the object of purchase. As Li-Chen, M. Ronnier Lou and others (11) show in their job to assess color emotion, they study people from eight countries (United Kingdom, Taiwan, France, Germany, Spain, Sweden, Argentina, Iran) and they show for like/dislike response, the results show strong effect of culture.

People from a community have the same answer when they see certain iconography; then, behind every message, must be some knowledge and therefore accords and rules that everyone must know.

The knowledge of Culture and Society, on which our product is focused, makes for two key variables in the business world - Commercial are permitted:

1st-. a larger aperture of scope and range of our sales, reaching more people.

2nd-. greater specialization, showing a particular market, with common characteristics to which our products respond.

BLUE. The works of Simon Lee and V. Srinivasan (Chino) Rao [9], user studies and Luis Gordillo Tapia [6], web analytics, show two ways of reaching the same conclusion, that blue is the favorite color for purchase. However, both studies are limited to specific groups, therefore, to sell on the Web, consider the All, for this the Internet is a tool that offers the possibility of global information society, culture, and therefore of all its members, because even if the user does not have Web access, information about the individual is still provided.

4 Discussion

Normally, works on web design have based their findings in the study of users such as potential customers for understanding their needs and interests.

However, in the case of color choice, it is possible to dispense with the opinions or individual tastes of each group of users and make your choice based on socially acquired knowledge within each country or region, and even within each age group [10]. That is, the meaning that the colors have for the general public, taking into account the culture and traditions of a particular sector. According to Fernando Parodi Gastañeta [12], color "... is related to communication: our fellow transmitted by certain cultural or social code sharing, news about our character and purposes while allowing us to know to other." Javier Nó [10] sometimes says color transmits to us, it is related to something innate to color, but also that sometimes it is learned, in either case colors are carriers of meaning and information.

Relativism, Theory of Knowledge [13], argues that certain realities and knowledge are common to a social group or universal. In order to understand this common knowledge, we would have to rely on social sources apprehended along history, regardless of individual knowledge. This is the particular case about colors communicating to us, because it can establish a correlation between the color used and the perception of the society without entering into discussions of personal preference.

It is therefore important to talk about the professional development of these forms of electronic sales, with the increasing apogee of the custom pages by the vendors themselves without consideration of an expert because they would take currents of opinion that establish the study of users, as well as how we keep the wisdom of structures such as the Theory of Knowledge.

The website represents an online store that should provide a sense of professionalism and expertise. It must also convey confidence to the buyer, especially in the time of payment of the product, since a time of distrust can cause loss of the sale. Therefore, we must understand that colors communicate to us and they achieve a non-verbal communication between seller and buyer, thus establishing a connection between them and intend to continue with the same “conversation”. From this point of view, we must also ask ourselves if, in the case that colors change on the page, language will also change in the same way, and therefore the buyer will receive and will perceive a different message, thus breaking the rules of communication so far established.

As John says Carlos Alútiz Colorado [1] in his dissertation entitled “The legal sources of the modern public morality: The contributions of Durkheim, Habermas and Rawls in his study of the theory of knowledge”, the two sides of the communication process (transmitter and receiver), the "actors" agree in advance to the "rules" necessary to understand each other and thereby create communication. These rules are based on socially established standards, so if you break these adopted conventions, communication will be broken.

Simon Lee and V. Srinivasan (Chino) Rao [9] provides, through a user study, color blue inspires confidence at the time of purchase of products, as demonstrated

[6], [4], most Spanish companies engaged in e-learning, using the color blue for their website, which demonstrates a positive response from users; however, having done some research about what colors mean in society, and specifically blue, Spanish respond to the socially established canons of what colors communicate to us and the experience of a particular sector of society. It is not answer a personal taste but common knowledge.

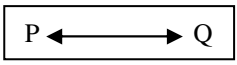
Marketing Research and Information Human Behavior were based on the user study, i.e. are directed to the majority group, but statistical studies are showing a greater democratization of the access to new technologies of various social groups (Mobile Internet, public access to the Internet, e-literacy, ...), which shows that social groups which have access to the Internet are increasing, and therefore the challenge for the on-line shops is to attract not only the groups already settled but new users too.

In this sense, we think that the study of the parties would give us all, giving us a sampling of the Society as a whole, but does not consider minority groups, or potential customers.

According to Logic [8], we can see how the sum of the parts does not equal the whole, and therefore, is based on false premises that provide us with wrong knowledge.

P: are parts

Q: is everything.



Q will be everything when be P give the same answer as Q.

We start from the Premise "P", which, it is a user study, where he is established a statistical study in which the majority response will be accepted as true, we demonstrate that in no case the sum of the parts would give us the whole answer.

Therefore, user studies we get skew data, they do not consider the minority groups, or potential customers, therefore they close us to a future market because the lack us of global knowledge of the users, when all this information is in the online and is visible through the recovery work performed by information specialists in the field, traditional area of library and documentation (Information Sciences).

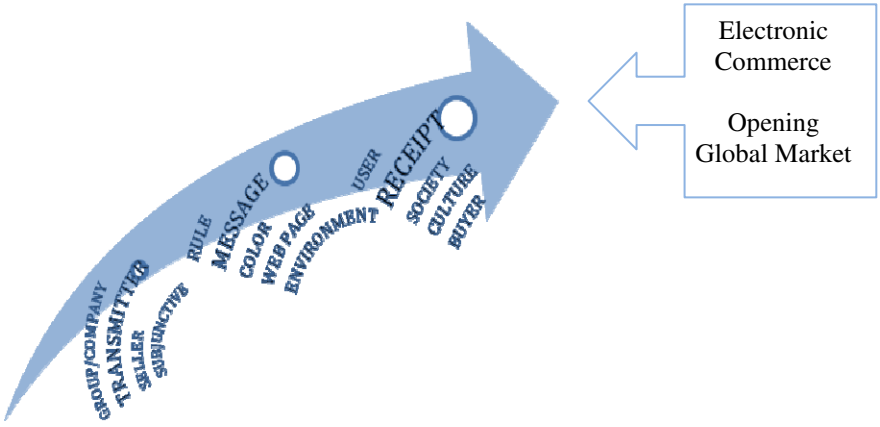


Fig. 2. Holistic analysis of color in electronic commerce

5 Conclusion

It is shown that the theory of knowledge comes from different sources and different spheres of life. From this point of view, and always understanding that the use of color is a form of nonverbal communication, or extra-verbal, we have taken as reference the Communicative Action of Jürgen Habermas [7]; then, for the application of color and therefore the management of this language, we provide a social information valuable to the relationship between actor and audience is correct, and meets the established roles in society, through the application of "rules", understood by the subjects and therefore, with explicit content on the message about society as a whole, which is expected by members of the same, will be accepted and respected.

The choice of color of a web page would not be established through a user study, but rather analyzing the meanings of colors and what they evoke in different sectors of the population, thus enabling specialization, as well as an easier way of conveying an idea or message through the colors that make up the web site, possibly with different colors associated with the product for sale, brand, logo, ... From an anthropological perspective, colors have different meanings, depending on the social or geographic area in which they are. Therefore, we must pay special attention to the country or countries to which the product is intended for sale on the website, then the color appropriate to sell the product in an area can be harmful in another.

In conclusion, this article advocates the use of color in business websites, since it is a form of "global communication" between seller and buyer. This is achieved through a series of "social norms" which are based on rules for the proper understanding between both parties. They have their origin in social skills seized on what colors evoke in every sector of society and whose break or change, cause them to miss the communication.

For these reasons we can say that in the commercial websites, the color chosen to illustrate, is a form of "Communication of Group" between virtual enterprise and buyers.

References

1. Alútiz Colorado, J.C.: Las fuentes normativas de la moralidad pública moderna. Las contribuciones de Durkheim, Habernas y Rawls. Servicio de Publicaciones de la Universidad Pública de Navarra, Pamplona (2002)
2. Álvarez Lara, O.: Influencia del color en las preferencias de los consumidores. Revista Observatorio Calasanz 2(4), 228–246 (2011)
3. Fernández Falero, M.R., Meléndez González-Haba, G., Peral Pacheco, D.: Usabilidad de sitios Web comerciales dedicados a la teleformación. Revista de Estudios Extremeños 65(1), 517–533 (2009)
4. Fernández Falero, M.R., Peral Pacheco, D.F.: Comportamiento humano de la información comercial: teleformación en España. Information Research: and International Electronic Journal 16(1), 1–16 (2011)

5. Fraticola, P.L.: La comunicación y el simbolismo del color. Actas de Diseño. IV Encuentro Latinoamericano de Diseño 2009 Diseño en Palermo Comunicaciones Académicas (8), 229–230 (2010)
6. Gordillo Tapia, L.V.: Comportamiento Humano de la Información en los sitios web de e-learning españoles. Badajoz: Universidad de Extremadura (2011) (unpublished)
7. Habermas, J.: Teoría de la acción comunicativa: Racionalidad de la acción y racionalización social. Grupo Santillana de Ediciones, Madrid (1999)
8. Hortalá González, M.T., Leacha Albert, J., Rodríguez Artalejo, M.: "Matemática Discreta y Lógica Matemática", Editorial Complutense, Madrid (2000)
9. Lee, S., Rao, V.S.(Chino): Color and store choice in electronic commerce: the explantory role of trust. Journal of Electronic Commerce Research 11(2) (2010)
10. Nó, J.: Color y comunicación. Salamanca: Departamento de Ediciones y Publicaciones de la Universidad Pontificia de Salamanca (1996)
11. Ou, L.-C., Lou, M.R., et al.: A Cross-Cultural Comparison of Colour Emotion for Two-Colour Combinations. Color Research & Application 37(1), 23–43 (2012)
12. Parodi Gastañeta, F.: La Cromosemiótica, el significado del color en la comunicación visual. Comunicación 2(3), 46–58 (2002)
13. Vargas-Mendoza, J. E.: Teoría del conocimiento. México: Asociación Oaxaqueña de Psicología A.C. En (2006), http://www.conductitlan.net/conocimiento.ppt
14. Vera, C.: Generación de impacto en la publicidad exterior a través del uso de los principios del neuromarketing visual. Telos 12(2), 155–174 (2010)
15. Wilson, T.D.: Human Information Behavior. Informing Science 13(2), 49–55 (2000)

Visualization and Manipulation of Information in 3D Immersive Environments

Filipe Costa[1], João Paulo Pereira[2], and António Castro[1]

[1] GILT, Departamento de Engenharia Informática (DEI),
Instituto Superior de Engenharia do Porto (ISEP),
Instituto Politécnico do Porto (IPP), Porto, Portugal
[2] GECAD, Departamento de Engenharia Informática (DEI),
Instituto Superior de Engenharia do Porto (ISEP),
Instituto Politécnico do Porto (IPP), Porto, Portugal
{1920077,JJP,avc}@isep.ipp.pt

Abstract. Developments in recent years has resulted in the creation of computational environments with potential for users to carry an immersive environment. This article aims to describe the development process of 3D immersive environments that require the use of specific tools and techniques, that allow the original environment with a set of associated information that will help to describe the own environment. For this purpose prototypes were created and results were evaluated.

Keywords: 3D Immersive Environments, Panoramic Image, Human-Computer Interaction.

1 Introdution

With the proliferation of Internet access, arisen different navigation environments, called 3D immersive environments. The navigation in these environments is not limited to the query text or following links between pages on related topics. These new visualization tools are based on the exploration of images, allowing you to select a specific image area to access related content, of all kinds, including sound or video [1]. Images that served as environment framing, becomes the basis for navigation. That's why increasingly panoramic pictures are used to this exploration, increasing the realism of the navigation in these environments [2]. Above this image we can place some connection points (known generically as "Hot Spots") that are marked generally by small symbols, which allow access to other environments or contents related to a particular point of interest of the image [3].

These environments are nowadays used with various purposes, creating virtual tours, from museums, monuments, shops and even cities, or even showing scenes of events in 360 °, giving the user the possibility to fill involved with all the given place and time. The major challenge in the development of these environments is to engage the visitor, as much as possible, trying to convey the same sensations as if it were in the real environment. For this, we need to recreate all the surroundings and, at the same time, give the user the ability to interact with the environment.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1041–1051.
DOI: 10.1007/978-3-642-36981-0_99

2 3D Immersive Environments

Nowadays, The most evolved form of portraying a location is the use of a panoramic image. This technique was first used in 1787 by the Irish painter Robert Barker (Fig. 1). The viewer was placed on a central platform, completely involved with the image from where he could look across the landscape, allowing the viewer to feel as if transported to the pictured location. [4]

Fig. 1. Robert Barker –The Panorama (landscape of Edinburgh in 1787)

Over the years grew the desire to give the viewer a greater ability to interact with the presented scene. This capability was definitely made easier with the advent of digital photography. The digital image can be processed and manipulated in real time, which allowed equip display devices with capacities of interaction with the viewer, that becomes the user and have the power to choose, to drive, to act on the environment itself, making a fully interactive but different visualization from individual to individual. Thus, the user can create a unique and personal experience, becoming a co-creator of space [2] [4].

Being an 3D immersive environment based on panoramic image, there are several rules to keep in mind, when you want to build this image. The camera used must be chosen to ensure the necessary quality to be achieved. It is necessary to note that the more the resolution photographs, more details can be presented to the user, but also may require a greater effort in its production or later in the environment navigating. Although there are cameras specific to take panoramic photos, nowadays you can make panoramic photo in any proportion, up to 360 º, using only common cameras and scanners (scanners), or simply digital cameras and of course, a special computer application to combine photos into a single image [7]. In addition, is also recommended to use a tripod, preferably suitably leveled in order to ensure the perfect junction of several photographs. The photographs are to be taken sequentially in order to capture the scene of 360 ° (Fig. 2). They are usually takes only 12 pictures.

Fig. 2. Orientation Control of the camera with rudimentary method

Once obtained, the captured images are compiled into 2D cylindrical panoramic view (Fig. 3) and later processed with specific applications. Overlays are removed by image manipulation and color filters are applied as needed.

Fig. 3. Twelve photographs of the panoramic image for mounting the main lobby of ISEP

In a virtual environment, the user has the freedom to navigate through this environment like moving a virtual camera, focusing it in any direction or changing the zoom factor of the image. This identifies two components that make up these environments:

- *The contents of the environment*: a panoramic surface, where it will be mapped / projected the panoramic image of the environment;
- *The viewer environment*: a screen of the virtual camera, where is projected this surface overview, reflecting the viewing angle chosen, at each moment by the user.

3 Visualization and Navigation in 3D Immersive Environment

Today, the most common way to see a panoramic image is using a computer with a specific application or a web browser, endowed with its own tools. Only with one screen it's possíble to feel the 360 º image. The image printed and placed around the spectator would not have the charm or the effect of early panoramic photography because the viewer of our times is born in the world of image. The internet is therefore the natural place for the presentation and dissemination of these images and also allows an intuitive handling of them. The computer is therefore the ultimate device for this operation.

After ensuring access and availability to the user's environment, we need to add some lack of interaction capabilities to create more complete immersion of the user. To do this, connections are created, pointing interactive objects in the 3D environment, so when we navigate through space, we have the notion that you can select a specific object / connection and thus have access to other content [3]. This transition is performed between content via 3D links, which can be signaled by an icon placed over an image area. The very form of the icon can elucidate the user about the content that is accessible. The 3D link, also called HotSpot, can be used to [3]:

- Make the transitions between different views of the environment, so that the user scroll through the various scenes of the environment.
- Show a content associated with a particular detail, as an enlarged image, text or sound.
- Link to external documents or other internet sites related to a particular element of the scene.
- Open a particular application or document.

Instead of using an icon, we may signal an image areas, which react in passing the mouse over this area, changing color, giving the indication that there is a possibility of interacting with this content. Instead of this it can be used color tones, or other special effects. The possibilities are endless by combining many effects, if desired.

The interaction with a virtual environment can be implemented using VRML (Virtual Reality Modeling Language) or a newer version called X3D (eXtensible 3D graphics) that allow you to specify dynamic 3D scenes in which the user can navigate using a browser with appropriate tools installed [5]. So we can achieve different levels of interaction with the virtual environment [6]:

- *Animation of scenes:* Using interpolators and timers that control the display of animation continuously;
- *Spatialized audio and video*: audiovisual resources mapped on the geometry of the scene;
- *User interaction:* Mouse Function "pick and drag" and other inputs from the keyboard;
- *Navigation:* Cameras, user movement in the 3D scene, collision detection, visibility and proximity;
- *Networking:* Installation of 3D scenes with content from links to objects on the internet;
- *User Defined Objects*: Ability to create user-created objects borne by the browser;
- *Scripting:* Ability to dynamically change the scene through programming;
- *Physical Simulation:* Humanoid animation, geospatial data, integration with protocols DIS (Distributed Interactive Simulation)

All these possibilities of interaction must be mapped into the environment in order to allow access to the contents expeditiously. Although the information is related to each other and to the environment, it must be accessible without the user having to go through several steps.

4 Software Imaging Panoramic 360 º

Once described the concepts and implementation needs, is important to know what forms and effort is needed to develop an 3D immersive environment. The most basic way to construct the panoramic image is using an image processing tool, such as Adobe PhotoShop CS4. This process is very time consuming and requires much skill and experience of the composer.

Thus, we investigated several tools available in the market to be able to perform a comparison of the performance and results obtained with each one. Just six building tools panoramic images were selected and were submit to an initial composition test of a panoramic image, of the multimedia laboratory room of ISEP, from 12 photographs taken sequentially with differences of 30 degrees.

These tools were *Microsoft ICE*, *Panorama Plus (Serif), Panorama Composer 3 (FirmTools), VR Worx (VRToolbox), Ulead Cool 360 (Ulead)* and *Autopano Giga 2.5.2 (Kolor).*

With the survey and analysis of these tools, it was possible to identify the various types of applications available in the market to build panoramic images. Comparing the benefits of each of these tools turned possible to obtain a score covering aspects of quality of panoramic image; possible export formats loss; capacity / ease of direct export to the Web, or a more comprehensive assessment to own working environment and the features available in each one. Once, the tool elected was *Panorama Composer 3* as the best satisfied the above criteria and that allowed a better quality / usability.

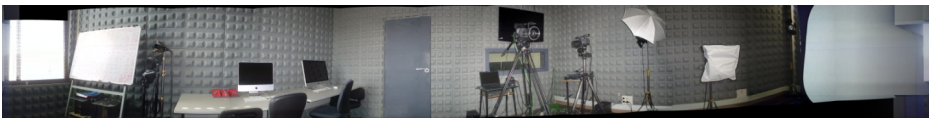


Fig. 4. Best panoramic images obtained in the tests using *Panorama Composer 3*

5 Tools for Creating 3D Immersive Environments

The creation of the environment itself also requires a tool capable of providing the user the ability to interact with the space. So we selected the following four tools: *Photosynth (Microsoft)*, *Dermandar.com, VR Works v2.6.1 (VRToolbox)* and *Panotour Pro v1.5.3 (Kolor).*

With these tools is possible to load multiple photos of a certain place or object and thus recreate a three-dimensional view of the environment. It can be used an amount of pictures without any mandatory limit. Images can have non-sequential numbering and even photographs can be repeated. None of these factors influence the final outcome, that only is based on the combination of textures of different images. The scene can be covered with the help of a mouse or a command line.

In the scene it can be added some contents that can be of the following types:

- *Still Image* - Place the pan that will be shown entirely and without animation;
- *Generic Node* - To add only written information;
- *URL* - To add a link to an Internet site;
- *QTVR Object* - To put a movie (QuickTime VR), defining the initial appearance and the animation speed.

For each additional content we can also add more connection points (hot spots), allowing to create a real network of interconnections, where each content can lead to other related content to reach the desired level of knowledge.

To produce the interactivity in an 3D immersive environment, we need to identify places of interest in the image (hot spots) which are associated with the desired objects. These sights are geometrically defined (Fig. 5) to better adapt to the shape of interactive objects and to provide different types of interactions, such as:

- A simple jump to an image, a flash animation, a movie or even an audio file.
- A link to an Internet site through a URL;
- Run HTML specific code;

Fig. 5. Definition of interaction areas in each scene

On these sights is still possible to define the field of view or the limits to the zoom or also set a sound environment customized for each panoramic image comprising the environment. We can also define an auxiliary navigation menu to jump directly for each panoramic scene of the 3D immersive environment without needing to respect existing connections.

6 Prototype Built

In order to illustrate and demonstrate how to create an interactive 3D environment, two prototypes were built using two different tools. The intention was to show the different results obtained from the same sources.

The prototype built is a virtual tour of the facility LAMU (Multimedia Lab) of Department of Computer Engineering (DEI) of the Instituto Superior de Engenharia do Porto (ISEP). On this tour you can start the virtual tour, from outside ISEP, enter to the main hall and then reach the halls of access corridor and subsequently their rooms. This approach gives virtual tour a more realistic sense. However, since it is a virtual tour, you can, for example, access the room by clicking on the window outside the building or came back to the car parking, also through the window and down to the main lobby, clicking on the elevator.

To develop this prototype photographs were taken at ten sites in ISEP. At each site were taken twelve successive shots with guidance of 30 to 30 º, until reaching 360º to build the panoramic picture (Fig. 3). It were also selected three objects from the LAMU room (two video cameras and a beam of light) to mount films rotation objects (.MOV), to which were also required twelve photographs each (Fig. 6). In this case, however, the machine was fixed and the object had to be rotated 30° successively until a total of one complete rotation. To demonstrate another technique for panoramic viewing of large objects due to its size, can not be rotated about themselves, a vehicle was photographed in car parking of ISEP (Fig. 7) where it was necessary to move the camera around the vehicle, taking again twelve photographs, keeping the distance to the center of the object and describing angles of 30 ° between each shot.

Fig. 6. Twelve photographs for assembling film of rotation of small objects

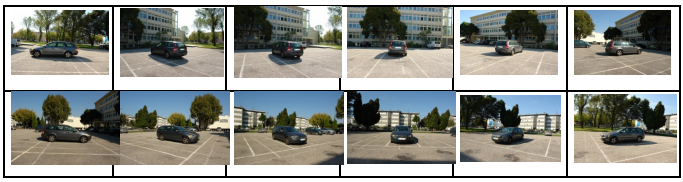


Fig. 7. Twelve photographs for mounting the rotation of large objects

To mount the panoramic images we choose the *Panorama Composer 3*, for the reasons given above. This process proved to be quite affordable and fast, with the exception of a few scenes with very specific characteristics that made the job more difficult to obtain a satisfactory outcome and therefore required a greater adjustment of parameterizations.

The objective of this phase of the study was to obtain panoramic images with the fewest possible defects, using only automatic generation tools.

To create both 3D immersive environments, it was used *VR Worx v2.6.1* and *Panotour Pro v1.5.3* to obtain a basis for comparison of results. With the construction of these prototypes of 3D immersive environment, it proved clear that is necessary a well defined initial plan for the desired end result. In fact, when shooting the real environment we should know, since this stage, for what purpose we must highlight certain areas and which links can be included and in what areas.

Finally, the assembly of the 3D immersive environment is the culmination of work and thus, it must be taken into account the consistency of sequences and links (Fig. 8) to provide a satisfactory effect for the end user. It was therefore necessary to arrange the sequence and the connections between the environment and the interactive objects (Fig. 9), in each one of the tools.

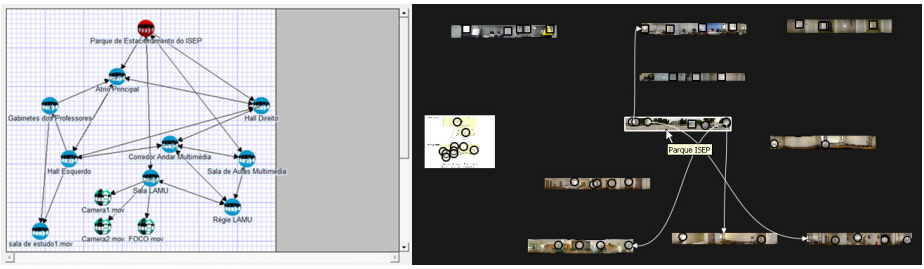


Fig. 8. Schema of links between objects to implement the 3D immersive environment with VR Works and Panotour Pro

The virtual tour with *VR Worx* appeares to be only the simple image of the scene with only the cursor marking the locations of connections by changing its appearance. In the construction or viewing the scene, it is not possible to change the type of scene transition, whereas, when clicking on a link, it appears immediately the new scene without any transitional effect.

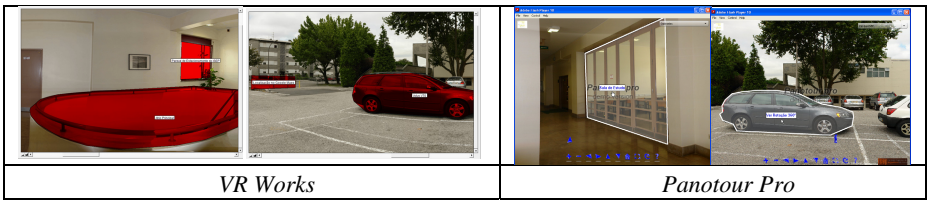


Fig. 9. Creation of links to other panoramas or links to Internet sites

The virtual tour built with *Panotour Pro* contains more visual information than the corresponding version produced with *VR Worx*, because of the enhanced possibilities for interaction it provides. As differentiating effect we choose to put different sound environments depending on the scenes being viewed. Also a map of the virtual tour, ISEP's logo and yet the list of locations visit to allow direct access.

7 Obtained Results

The prototypes of 3D immersive environments created were subjected to an evaluation by a group of students from two classes of the discipline of design and interface of the master program of graphics and multimedia systems from ISEP. The survey was done so in a universe composed of 86% of the male and 14% female. The ages of the respondents were mostly below 25 years (48%), and there is a significant group between 26 and 35 years (38%) and only 14% were older than 36 years. The evaluation was carried out with the help of a questionnaire divided in two parts. The first serves for the perception of the information contained in the images and verify the advantages of the use of panoramic images. The second part evaluates the 3D environments built as well as the interaction options implemented, so as to evaluate the suitability and usability achieved.

One of the objectives were to confirm that without any single reference known, the image itself does not give us complete information about the site. The images where the respondent could recognize some benchmark, achieved the highest number of right answers. It was then confirmed that increasing information about the same location can help to better identify the location (Fig. 10). The identification of the site with a enlarged image view (by air) only works when the viewer is aware of some references.

Fig. 10. Question about which photography provides more information about the location

By presenting a panoramic picture, it could be proved that the vast majority of respondents (86%) were able to recognize a 360º view, which confirms that is a perception usually easy to achieve by the viewer.

In the second part of the questionnaire, we asked the respondent to use both developed prototypes (VR Worx and Panotour Pro) to make a virtual visit to the multimedia facilities of ISEP. The user is presented to a sequence of steps to be carried out in order to discover the various scenes and possible interactions with various objects and locations in both prototypes. This phase of the investigation aimed to understand user satisfaction when performing these virtual visits in each of the tools. Since the two tools have different ways of displaying scenes and menus, this only intended to realize, which solutions reached the higher degree of acceptance by the users, including immersiveness, navigability, interaction with the scenes objects, the transition between scenes or the sound environment of each scene (Fig. 11).

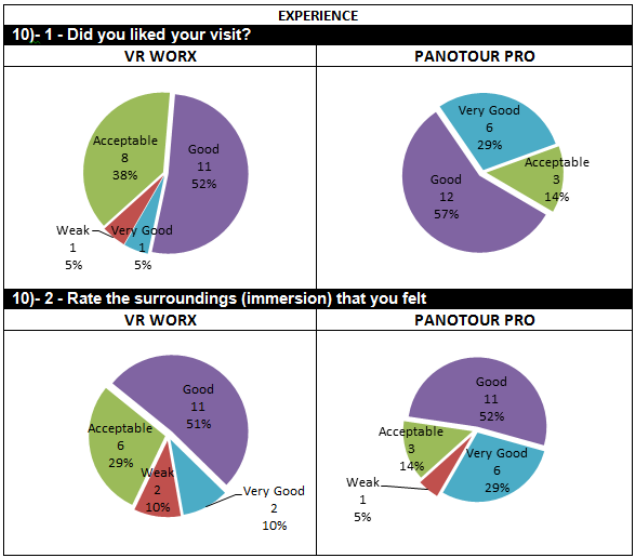


Fig. 11. Comments about the immersive experience comparing the two prototypes

Overall immersive experiences were much appreciated and users also identified their preferences about the possibilities of manipulating objects, and it has been the the car rotating to experience most voted (Fig. 12).

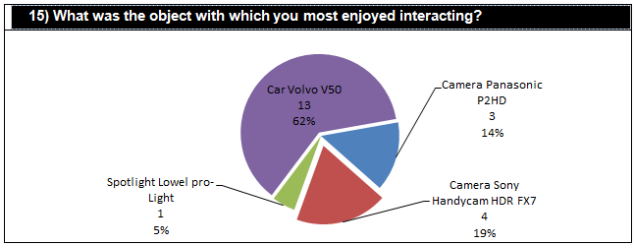


Fig. 12. Vote on experience of interacting objects

8 Conclusions

As it was evident throughout this work, the technological area of production, visualization and manipulation of information in 3D immersive environment is constantly evolving

In terms of techniques for generating panoramic images, as it was demonstrated in this study, there is a multitude of tools in the market, which then become the basis for creating 3D immersive environments. These tools have been improved over time, which was evident in the comparison of the two tools *VR Worx* and *Pro Panotour*

The concept of 360-degree panoramic image is to extend to other concepts such as the spherical image of 360, in which there is no imposition of limits to space exploration, giving more and more possibilities to the user. Another concept is the synthesis of photos provideed by *Microsoft Photosynth* that to explore the space, no longer even need to do the actual junction of different images, it only establishes the relationships between them, in order to present the most correct sequence of photographs, to give the perception of 3D space.

With the prototypes evaluation it was possible to identify that the picture, even panoramic, is not yet enough to give the viewer all the information he need to identify the location pictured. In the absence of recognized benchmarks is necessary to "feed" the picture with more information as clearly and directly as possible. Thus, it was possible to identify users preferences in terms of navigation environment and the types of interactions. Due to the immense potential demonstrated by these 3D immersive environments, we can certainly say that in a few years, there may be information systems based on multimedia technology like this, allowing these systems provide the user with new ways to interact with information. Assuming that any content can be associated with any 3D immersive environment, we can forecast the revolution that is coming.

References

1. de Souza e Silva A., Ferreira L.: CIBERIDEA 3D — Por uma construção do espaço digital [Conferência] // Anais do Sigradi 2001 (V Congresso Ibero-Americano de Gráfica Digital). [s.n.], Chile (2001); Z ou como estar imerso no espaço digital [Conferência] // Anais da Compós 2001 (X Encontro Anual da Associação Nacional dos Programas de Pós Graduação em Comunicação). [s.n.], Rio de Janeiro (2001)
2. André, P.: André Parente.net - Panoramas - História dos dispositivos (1991), http://andreparente.net/figurasnapaisagem/#/panoramas/
3. Ribeiro, N.R.F.: Artdoc3D: Representação Digital de Instalações Artísticas. Universidade Nova de Lisboa - Faculdade de Ciências e Tecnologia - Departamento de Informática (2010), http://run.unl.pt/handle/10362/4246
4. Ferreira Luciana O Espaço Digital Imersivo [Conferência] // Anais da Compós 2000 (IX Encontro Anual da Associação Nacional dos Programas de Pós Graduação em Comunicação) - Rio de Janeiro: Escola de Comunicação da Universidade Federal do Rio de Janeiro (ECO/UFRJ) (2000)

5. Schneider Daniel, K. Sylvere, M.-M.: VRML Primer and Tutorial [Relatório] - University of Geneva : TECFA, Faculté de Psychologie et des sciences de l'education (1998)
6. Roberto, R.: Introduction to X3D [Relatório]. HCI Lab, University of Udine, Italy
7. United Photo Press UPP Foto panoramica. United Photo Press (August 10, 2007), http://www.unitedphotopress.net/2006/05/foto-panoramica.html
8. Google ©2011 Google Earth Pro (2011), http://earth.google.com/enterprise/earth_pro.html#comparison

Construction Processes Using Mobile Augmented Reality: A Study Case in Building Engineering Degree

Albert Sánchez Riera[1], Ernesrt Redondo[2], David Fonseca[3], and Isidro Navarro[2]

[1] EPSEB, Universitat Politècnica de Catalunya Barcelona Tech,
Avd. Gregorio Marañon 44, 08028, Barcelona, Spain
albert.sanchez.riera@upc.edu
[2] ETSAB, Universitat Politècnica de Catalunya Barcelona Tech,
Avda. Diagonal, 649, 08028, Barcelona, Spain
{ernesto.redondo,isidro.navarro}@upc.edu
[3]Arquitectura La Salle, Universitat Ramon Llull,
C/. Quatre Camins 2. 08031, Barcelona, Spain
fonsi@salle.url.edu

Abstract. In this paper we describe the implementation and evaluation of an augmented reality (AR) application, on mobile devices. It's based on optical images recognition from real environment, and it has been particularly intended to be used in educational environments. Some improvements have been implemented compared to existing commercial ones, such as the ability to compare and move different models consecutively with a single marker. The objective was to assess the feasibility of using AR on mobile devices in educational environments. In addition we evaluated academic performance improvement using this technology. Validation was done through a case of study where students were able to see a virtual construction process overlapped to real environments. It was carried out by Building Engineering students of the EPSEB (UPC-Barcelona-Tech). Results obtained by student's PRE and POST test, and by questionnaires responses, demonstrated the application suitability as a new tool to be used in learning processes.

Keywords: Augmented Reality, Construction processes, educational research, Human-Computer Interaction, mobile learning.

1 Introduction

Beyond the superimposition of virtual objects on real environments, AR features are usable in a wide range of applications in the field of engineering and construction, offering potential advantages in all stages of a construction process [1], from conceptual design, to building systems management and maintenance throughout its servicetime. Virtual models, once overlapped to real space, can provide additional information for a better understanding of the building, thus contributing to a greater efficiency in construction processes, rehabilitation or building maintenance.

In this case we focused on the implementation of new digital technologies in building construction and maintenance learning processes, within the course of

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1053–1062.
DOI: 10.1007/978-3-642-36981-0_100

Technical Projects II in the School of Building Construction (EPSEB) (UPC-Barcelona-Tech). In general, a synergy between traditional methodologies and AR new technology was proposed, to visually build up hybrid (virtual and real) construction processes. In this case, we tested the process of opening a void in a load-bearing wall. The Study case was held with 146 students divided in a control and an experimental group. Students blend the physical and virtual worlds, so that, real objects (markers) were used to interact with three-dimensional digital content and to increase shared understanding. We used light maps in textures to incorporate lighting conditions from surroundings, and introduced *occluders* for a better integration of the scene in its real location.[2].

The objective was twofold. First, to evaluate the feasibility of using AR technology on mobile devices, in educational environments, and secondly, to assess the student's academic performance improvement. To do that we compared two scenarios: S1 (based on slides and traditional methodologies) and S2 (based on augmented reality technology on mobile devices). The research questions were:

1. What's the student's degree of satisfaction and motivation using this new methodology?
2. Are there any differences in academic results depending on which of the two teaching scenarios proposed are used?

On the first case, some experiences have been done to evaluate Virtual Environments (VE) usability [3, 4]. In our case we were based on ISO 9241-11 which provides usability guidelines: Effectiveness, defined as the user's ability to complete tasks during the course, in relation to the "accuracy and integrity" that it had been made; Efficiency, on the assigned resources, they asked questions related to the expenditure of time and effort for solving the proposed exercise; Satisfaction, understood as subjective reactions of users about the course.

On the second case, to evaluate academic performance improvement, we compared final results between control and experimental groups. Results obtained by student's PRE and POST test, and by questionnaires responses, demonstrated that combining an attractive technology, and by the user-machine interaction that involves the AR, students feel more motivated, their graphic competences and space skills are increased in shorter learning periods, and their academic performance is highly improved. The experiment can be seen in: http://youtu.be/8UEs8T6vSPI

2 Theoretical, Pedagogical and Didactic Foundations

Learning, by definition, is the process by which memories are built, while memory is the result of learning [5]. In recent years, the desire of learning process improvement has led a transition to a technologically enhanced classroom, where computers, media players, interactive whiteboards, internet, web 2.0 tools, and games have been incorporated. E-mail and mobile phones have transformed the way we communicate, and the list of technologies that can be useful in learning processes is huge and constantly growing, not being simple to define which may be suitable for learning and which are not [6].

More recently, immersive technologies in virtual and augmented reality worlds have been used. Its usefulness has been assessed by numerous international projects

[7], [8], [9], [10], [11], [12], [13], [14]. These experiences, that used augmented reality in the area of entertainment and education, demonstrated the great potential of this technology. But in education, it may still be considered as a new tool, and further investigation is necessary, paying special attention to user experience and learning processes [15]. Because, despite the ongoing effort to technology implementation, there is also the need to immerse students in new learning environments, these in turn, continuously changing [16]. Teachers, meanwhile, face the challenge of constantly be updated to provide new forms of teaching, focusing on the acquisition of generic skills in which students must construct their own knowledge through constructivism, proposed by Piaget [17] and meaningful learning proposed by Ausubel [18]. In contrast, we can frequently find situations in the classroom where educational contents are simply exposed and presented without any interaction by the student, that receive passively new concepts and content to be memorized, getting bored easily and consequently minimizing their learning. Student motivation is essential to reverse this situation [14].

There is where AR can helps to improve the learning process performance [14], [19]. AR and virtual reality (VR) share some common features such as immersion, navigation and interaction [20]. However, AR has two main advantages over VR: Allow collaborative experiences in a real scene. So users can work with computer-generated objects as if they were real objects in a real environment, in real time; and Tangible Interaction. By superimposing virtual objects in a real environment through markers, user can modify and manipulate the scale, position and location of virtual objects. So we could say that AR technology, by providing new interaction possibilities, promote active student participation in its own knowledge construction. Thus, it becomes a suitable medium to be used in schools [21].

In our field of study, construction processes, AR technology features would allow to show a "completed" reality, superimposed to the real. It could create an impossible image of what does not exist, as a result of the analysis of existing building systems (structural, facilities, envelope ...). It would facilitate rehabilitation and maintenance tasks, as well as systems verification, update and interactivity in the same place, and in real time, promoting more efficient management and control processes of building constructive elements. During the case of Study, described below, we implemented AR technology applying these principles to future construction engineers, with the purpose of increasing their learning, getting greater efficiency performing these tasks, traditionally handcrafted and little technified.

3 Case of Study

Technical Projects II is a subject which is intended mainly to provide students with technical capacity enough to tackle construction and execution issues of a technical project. They acquire the ability to apply advanced tools needed to solve all problems involved in technical project management, as well as, the ability to write technical projects and construction works and to draft documents which are part of them. In summary, at the end of the whole course they should be able to analyze technically project performance and its translation to the execution of works, and get the capacity to integrate into a work team.

The course is divided in three blocks: Structural falsework, Building Facilities, and building envelope. As we said, we worked on the first one, focused on the process of opening a void in a load-bearing wall. At the end of this practice, the student should be able to: Identify different ways to implement this construction process; analyze the structure of a building and quantify their loads to replace a structural element; and calculate and design all the elements in a practical case.

3.1 Procedure

In order to evaluate usability of AR technology on mobile devices and to assess the student's academic performance improvement, the experience was performed in three stages. PRE-Test, Lectures, and POST-Test. The total number of students enrolled in the course was 183. At the end of this first block, after students evaluation, we excluded those students who had not performed any of the tasks required for assessment (PRE-Test, practical exercises, or Final test), so the final number of students who participated in the experiment was 146. They were divided into 4 groups. 3 control groups (1M, 2M, and 3T) and 1 experimental group (4T). Control groups followed the traditional course based on slides (Scenario 1), and the experimental group was involved in AR specific training (Scenario 2). They used a self-developed application (U-AR) under Android platform.

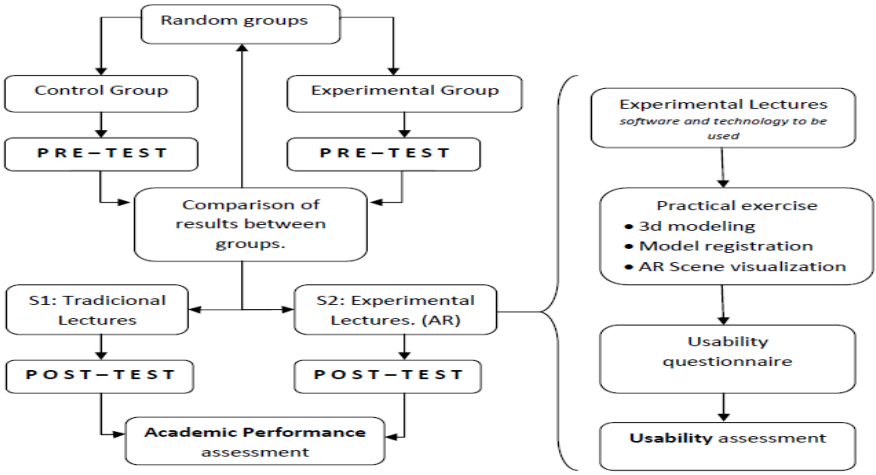


Fig. 1. General Scheme of the methodological process

So, the first day of the course, all students answered a test (PRE) that was used to determine prior knowledge on the subject. It is based on previous years used tests. It was useful to verify that all students groups were similar before start the experience.

During three sessions, they all received a conventional class, based on lectures and practical exercises. Participants, divided into small working groups of 5 or 6 students, consulted and clarified doubts with the teacher. Students from 4T group (experimental) , however, received an additional lecture which taught the application

operation, and how to manage distributed contents to be visualized through augmented reality. In addition, they got detailed instructions of the assay to be performed through their devices, and 3d virtual construction process to be visualized was explained and distributed.

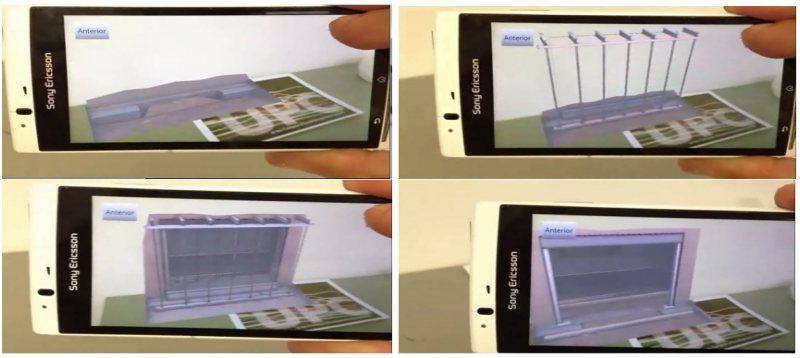


Fig. 2. Images of the construction process to open a void in a load-bearing wall, using AR on mobile devices

In short, they should select a place in the school, and through their mobile devices, they should watch "in situ" the five steps we had divided the construction process: Reinforcement of existing foundations; falsework for shoring and temporary support loads; Demolition of the brick wall; placement of columns and beam to support final loads; and falsework removal and final state.

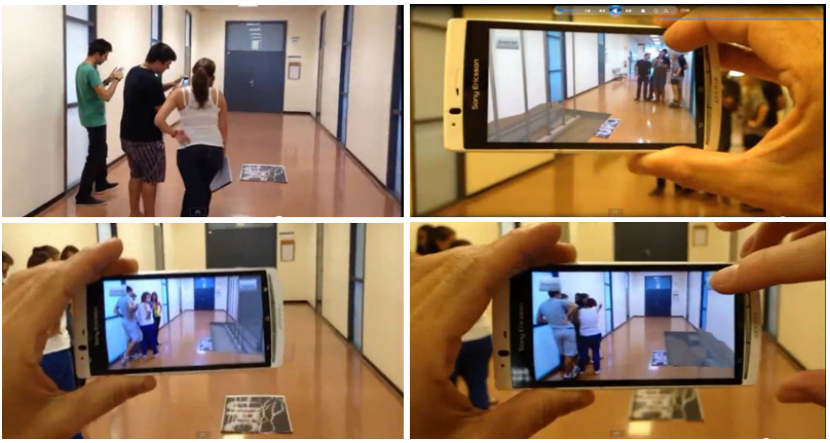


Fig. 3. Images of experimental group students viewing a construction process using their mobile devices

Once they finished, experimental group students were also required to answer a usability questionnaire in order to get their opinion related to efficiency, effectiveness and satisfaction opinion about the experience.

Academic performance was assessed comparing results between Control and experimental Groups.

3.2 Usability Assessment

As we mentioned before, we evaluated user's assessment using questionnaires based on ISO 9241-11. Responses average related to effectiveness, efficiency and satisfaction were very similar, ranged from 3.31 to 3.46, out of 5.

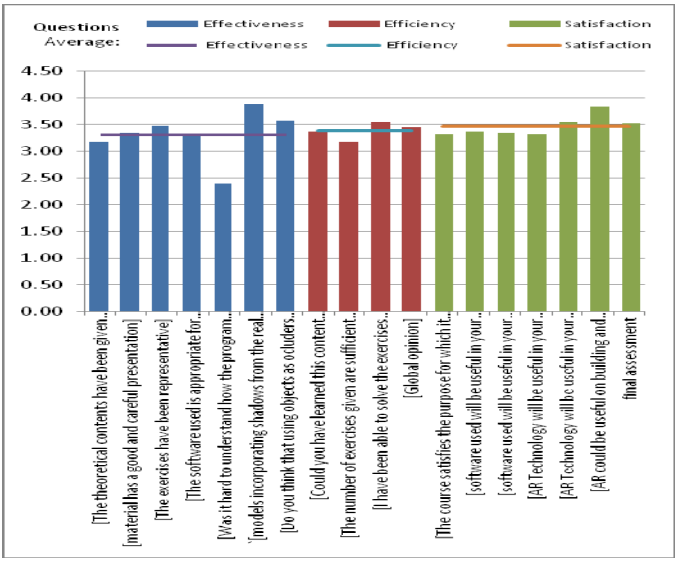


Fig. 4. Student's responses of Usability questionnaire

The overall assessment of the courses was rated 3.51 points out of 5. Similar results were found also in previous studies [22] which confirms the feasibility of using this technology in educational environments.

In a correlation analysis between the course final assessment and the other variables, a high correlation (0.71) was detected with: the representativeness of the exercise and material presentation (0.73). So these variables seem crucial to the success of this kind of teaching experience. On other hand, variables related with the fact of being able to solve the exercises independently (0.09) did not correlate significantly with the course final assessment. No correlations were found between PRE and POST scores, nor with the gain.

Correlations		**PRE_TEST**	**POST_TEST**	**GAIN**	**final**
final	Pearson Correlation	-0.00064	-0.24794	-0.16382	1
assessment	Sig. (2-tailed)	0.99743	0.20333	0.40485	

3.3 Results and Discussion

PRE-TEST mean scores were very similar in all groups. The group that has the highest score was the 2M control group. Experimental group (4T) hovers slightly under the global average of the four groups (Table 1).

Table 1. PRE-Test Results

GROUP	SUBGROUP	N	Mean	Std. Deviation
Control	1M	26	2,52	1,32
	2M	44	3,14	1,45
	3T	38	2,66	1,71
	Total	**108**	**2,82**	**1,53**
Experimental	4T	38	2,62	1,74
	Total	**38**	**2,62**	**1,74**
Total		**146**	**2,77**	**1,58**

To estimate what is the probability that groups are significantly similar, we used Student's t-distribution [23] setting to null hypothesis (H_0) that there are no differences in scores between groups. Statistical significance (2-tailed) was 0.502, higher to 0.05, which means that there is very little chance that the groups are different in their skills, previous training, and therefore the experimental group, who will practice with mobile devices is very similar to the other groups. Null hypothesis is accepted (no significant differences between groups).

Table 2. Independent Samples Test

		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
PRE_TEST	Equal variances assumed	,673	144	**,502**	-,06067	,29882

Once students training in this block is finished, they were scored using the following criteria:

- Task 1. 10%: practical exercises at the beginning of each session.
- Task 2. 30%: Theoretical Test.
- Task 3. 60%: Final presentation, which must provide constructive description and graphic of the entire construction process of felling.

The table 3 shows, for groups and subgroups, the results and the gain on the pre-to post course:

Table 3. Results and gain on the pre-to post course

SUBGROUP/GROUP		PRE_TEST	POST_TEST	Gain
1M	Mean (S.D.)	2,52 (1,32)	4,24 (1,13)	1,72 (-0,19)
	N	26	26	
2M	Mean (S.D.)	3,14 (1,45)	4,36 (1,02)	1,22 (-0,43)
	N	44	44	
3T	Mean (S.D.)	2,66 (1,71)	4,80 (0,95)	2,14 (-0,76)
	N	38	38	
Control	**Mean (S.D.)**	**2,82 (1,53)**	**4,49 (1,04)**	**1,67 (-0,49)**
	N	**108**	**108**	
4T	Mean (S.D.)	2,62 (1,74)	4,81 (0,86)	2,19 (-0,88)
	N	38	38	
Experimental	**Mean**	**2,62 (1,74)**	**4,81 (0,86)**	**2,19 (-0,88)**
	N	**38**	**38**	
Total	**Mean (S.D.)**	**2,77 (1,58)**	**4,57 (1,01)**	**1,80 (-0,57)**
	N	**146**	**146**	**0**

The results show that the experimental group (4T) gets better results after training, 0.24 points above the mean of the control groups. Higher gain in relation to the average of the control groups is achieved by experimental group.

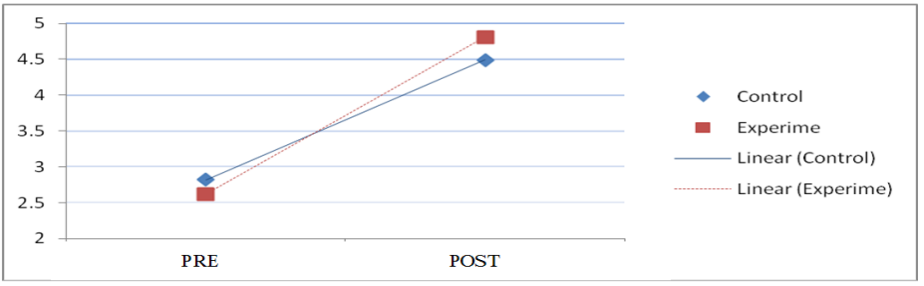


Fig. 5. PRE and POST results evolution

4 Conclusions

In relation to the first question referred as a research question, we may say that outcomes obtained from questionnaires were very positive. The overall assessments about efficiency, effectiveness and satisfaction were all around 3.5 out to 5. So we can affirm that student's felt satisfied and motivated using this new methodology.

In addition, exercises representativeness and material presentation seem to be crucial to the success of this kind of teaching experiences. On other hand, variables related with the fact of being able to solve the exercises independently did not correlate significantly with the course final assessment. There were also no correlations between PRE and POST scores and Final assessment, nor with the gain achieved.

In relation to the second research question, results showed how AR technology, through student's motivation, can help to improve their academic performance. Experimental group achieved best gain and performance results. We tested these strategies in previous case studies [24], and in all of them students felt active and motivated. They engaged, participated, and interacted with 3d virtual content, learning, therefore, was maximized. Despite this, there is a group with similar improvements (3T) and Statistical significance between groups previous training indicated that there are some possibilities that groups were not equal before the course started. Perhaps it was because 3T group was composed of students who traditionally combine study with work. So could be easier for this group to get better results. In next approaches some more AR training and group membership equivalence will be needed to get more significant results.

Finally, we can affirm that AR technology using mobile phones on Building construction area, offers the opportunity of visualizing "in situ" different stages of a constructive process, helping to improve its understanding. This fact allows verifying and comparing different scenarios and virtual proposals, prior to real construction. It could replace, somehow, real interventions. To do that, it's very important the ability of viewing different models with the same marker. In order to show different layers, models, etc… thereby simulating a real constructive process.

Acknowledgements. This research is being carried out through the National Program R+D. Project EDU-2012-37247, Government of Spain. Titled: "*E-learning 3.0 en la docencia de la arquitectura. casos de estudio de investigación educativa para un futuro inmediato*."

References

1. Shin, D.H., Dunston, P.S.: Identification of application areas for Augmented Reality in industrial construction based on technology suitability. Automation in Construction 17, 882–894 (2008)
2. Sanchez Riera, A., Redondo, E., Fonseca, D.: Lighting simulation in augmented reality scenes: Teaching experience in interior design. In: 2012 7th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–6 (2012)
3. Martin Gutierrez, J.: Estudio y evaluación de contenidos didácticos en el desarrollo de las habilidades espaciales en el ámbito de la ingeniería. Doctoral Thesis (2010)
4. Stanney, K.M., Mollaghasemi, M., Reeves, L., Breaux, R., Graeber, D.A.: Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design. International Journal of Human-Computer Studies 58, 447–481 (2003)
5. Squire, L.R.: Memory and Brain. Oxford University Press (1987)
6. Dror, I.E.: Technology enhanced learning: The good, the bad, and the ugly. Pragmatics & Cognition 16, 215–223 (2008)

7. Andersen, T.L., Kristensen, S., Nielsen, B.W., Grønbæk, K.: Designing Augmented Reality Board Games: The BattleBoard 3D experience. In: Proceeding of the 2004 Conference on Interaction Design and Children: Building a Community, pp. 137–138 (2004)
8. Galantay, R., Torpus, J., Engeli, M.: "living-room". Interactive, Space-Oriented Augmented Reality. In: Proceedings of the 12th Annual ACM International Conference on Multimedia, MULTIMEDIA 2004, pp. 64–71. ACM Press, New York (2004)
9. Brederode, B., Markopoulos, P., Gielen, M., Vermeeren, A., de Ridder, H.: pOwerball: The design of a novel mixed-reality game for children with mixed abilities. In: Proceeding of the 2005 Conference on Interaction Design and Children, IDC 2005, pp. 32–39. ACM Press, New York (2005)
10. Pan, Z., Cheok, A.D., Yang, H., Zhu, J., Shi, J.: Virtual reality and mixed reality for virtual learning environments. Computers & Graphics 30, 20–28 (2006)
11. Budhiraja, R., Verma, S., Pandey, A.: Designing interactive presentation systems for classrooms. In: Proceedings of the 28th ACM International Conference on Design of Communication, SIGDOC 2010, p. 259. ACM Press, New York (2010)
12. Alvarez, C., Alarcon, R., Nussbaum, M.: Implementing collaborative learning activities in the classroom supported by one-to-one mobile computing: A design-based process. Journal of Systems and Software 84, 1961–1976 (2011)
13. Chen, C.-M., Tsai, Y.-N.: Interactive augmented reality system for enhancing library instruction in elementary schools. Computers & Education 59, 638–652 (2012)
14. Di Serio, Á., Ibáñez, M.B., Kloos, C.D.: Impact of an Augmented Reality System on Students' Motivation for a Visual Art Course. Computers & Education, 1–11 (2012)
15. Portalés Ricart, C.: Live LEGO House: a Mixed Reality Game for the Edutainment 19 - 28. Schottish Online Journal of E-Learning 1, 19–28 (2007)
16. Arvanitis, T.N., Petrou, A., Knight, J.F., Savas, S., Sotiriou, S., Gargalakos, M., Gialouri, E.: Human factors and qualitative pedagogical evaluation of a mobile augmented reality system for science education used by learners with physical disabilities. Personal and Ubiquitous Computing 13, 243–250 (2007)
17. Piaget, J., Gruber, H.E., Vonèche, J.J.: The essential Piaget. J. Aronson (1995)
18. Ausubel, D.P.: Educational psychology: a cognitive view. Holt, Rinehart and Winston (1968)
19. Chen, R., Wang, X.: An Empirical Study on Tangible Augmented Reality Learning Space for Design Skill Transfer. Tsinghua Science & Technology 13, 13–18 (2008)
20. Dunleavy, M., Dede, C., Mitchell, R.: Affordances and Limitations of Immersive Participatory Augmented Reality Simulations for Teaching and Learning. Journal of Science Education and Technology 18, 7–22 (2008)
21. Portalés, R. C.: Entornos multimedia de realidad aumentada en el campo del arte (2008)
22. Redondo, E., Navarro, I., Sánchez Riera, A., Fonseca, D.: Augmented Reality on architectural and building engineering learning processes. Two Study Cases. Special Issue on Visual Interfaces and User Experience: new approaches, 1269–1279 (2012)
23. Gosset, W.S.: The Probable error of a mean. Biometrika 6, 1–25 (1908)
24. Redondo, E., Navarro, I., Sánchez, A., Fonseca, D.: Visual Interfaces and User Experience: Augmented Reality for Architectural Education: One Study Case and Work in Progress. In: Cherifi, H., Zain, J.M., El-Qawasmeh, E. (eds.) DICTAP 2011, Part I. CCIS, vol. 166, pp. 355–367. Springer, Heidelberg (2011)

A User-Centered Interface for Scheduling Problem Definition

Jesus Piairo, Ana Madureira, João Paulo Pereira, and Ivo Pereira

GECAD - Knowledge Engineering and Decision Support Research Center,
Institute of Engineering – Polytechnic of Porto (ISEP/IPP)
Porto, Portugal
{jeefp,amd,jjp,iaspe}@isep.ipp.pt

Abstract. In this paper we present a user-centered interface for a scheduling system. The purpose of this interface is to provide graphical and interactive ways of defining a scheduling problem. To create such user interface an evaluation-centered user interaction development method was adopted: the star life cycle. The created prototype comprises the Task Module and the Scheduling Problem Module. The first one allows users to define a sequence of operations, i.e., a task. The second one enables a scheduling problem definition, which consists in a set of tasks. Both modules are equipped with a set of real time validations to assure the correct definition of the necessary data input for the scheduling module of the system. The usability evaluation allowed us to measure the ease of interaction and observe the different forms of interaction provided by each participant, namely the reactions to the real time validation mechanism.

Keywords: Scheduling Systems, Human-Computer Interaction, User Interface Development.

1 Introduction

The global competitiveness requires that organizations adopt agile techniques for planning, scheduling and strategic decision making. The scheduling process consists in the affectation of operations to the available resources. In this process is necessary to take in account a number of constraints such as the resource availability, manufacturing capacity, and production goals to be achieved as the delivery dates fulfill, completion time, delays reduction [1]. A scheduling system helps to strengthen the market position of an organization through the optimization operations allocation to scarce resources, i.e., optimization of the scheduling process.

This work arises from the need to provide user-centered interface to an existing manufacturing scheduling system, where the concepts of functionality and usability are considered from the beginning of the development process. The main goal of the interface is to allow the user to define the scheduling problem in a interactive and graphical way.

To fulfill this task it was adopted the star life cycle [2] in order to support the development process of the interface as a user-centered interface. The chosen

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1063–1073.
DOI: 10.1007/978-3-642-36981-0_101

development process comprises the following steps: system/task/functional/user analyses, requirements/usability specifications, design and design representation, rapid prototyping and usability evaluation. After analyzing the scheduling system, the target users, the tasks and the functionalities, and after identifying the usability requirements, the collected information was used in the design phase to create the interaction model, where the interaction objects are identified. The prototype was developed based on this model.

In order to evaluate the usability of the prototype, ten participants were recruited. All the participants have experience in scheduling systems. This experience was acquired by working or studying this king of systems. Each participant had to define the same scheduling problem and answer a written survey on the interaction aspects and available functionalities of the prototype.

After addressing the concepts that provide the proper support to the user-centered interface development process, some considerations about related scheduling systems are presented. Furthermore, it is also presented the structure and functionalities of the prototype, followed by usability evaluation. This paper ends with the conclusions and future work.

2 Concept Review

The concept review that supports this work is based on three research vectors: scheduling systems, human-computer interaction (HCI) and interface development.

The research area HCI [3] is dedicated to the design, development and implementation of interactive computing systems for human use and the study of phenomena surrounding the interaction process [4]. In the development process of interactive computing systems the concepts functionality (set of actions or services available to users) and usability (efficiency degree and appropriateness in achieving certain goals for specific users) should be considered since the first stage [5][6].

An interactive computing system includes two components: the functional component and the communicational component [2]. The first one, also known as computational component, is responsible for implementing the tasks that the system is capable of performing. The second one, known as user interface, is responsible for managing the communication process between the system and the user. From the user perspective, the user interface represents the system.

The user interface represents the boundary between the system and the user. In this sense, it gains a greater importance in representing the system behavior (system status) and providing the user with the means to control or operate the system [7]. The user interface development process is divided in two parts [2]: the interaction development, and the interface software development. The interaction component defines how the user interface works, i.e., its behavior. The interface software component provides the means for implementing the code that instantiates the interaction component. Both components are developed in different domains. The development of the integration component occurs in the behavioral domain, while in the constructional domain the interface software component is developed [2]. Both domains are necessary and equally important, each serving its purpose. The behavioral and constructional domains contribute to define the usability degree.

The chosen method for developing user interaction was the star life cycle [2] illustrated in Figure 1. This method supports top-down and bottom-up development and is centered on the continuous evaluation throughout the development process, i.e., before starting the next development activity the results are evaluated. The designation star life cycle is due to its star shape. The points in the star are not connected or ordered according to any sequence. This means that the development process can be started at any point of the star. The points (activities) are interconnected through the usability evaluation located in the center of the star.

For this work only the steps included in the area of interest on Figure 1 were completed. These five steps allowed us to develop a prototype for evaluation.

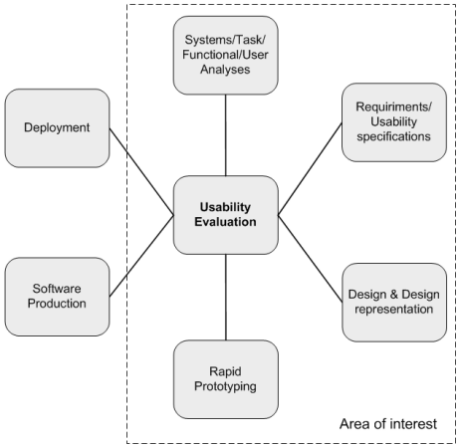


Fig. 1. Star method for user interaction development. Adapted from [2].

3 Problem Analysis

Before starting the analysis stage, a set of scheduling applications were studied in order to identify the kind of objects used in the interaction with the user and to understand the different ways in which the interaction takes place. The studied applications were: Lekin [8], Sistrade Scheduling [9], and JD Edwards EnterpriseOne Production Scheduling [10]. In these applications the scheduling plans are represented as diagrams, where the rectangles represent tasks. The possibility of dragging and dropping tasks is the major interactive feature provided by such applications.

On the other hand, in what data input concerns, the effort to provide the same interactive experience is much reduced, and in some cases the scheduling problem definition is made by filling in a usually complex form. Our work seeks to attenuate the imbalance between the data input and the visualization of results.

The system, user, tasks and functional analyses allowed us to understand and contextualize the problem. This set of analyses enabled us to specify the usability requirements. The gathered information during the analysis and requirements specification phase was used to build the interaction model (Figure 2) in the design phase. In the interaction model are identified the objects with which user will interact. The interaction objects are: operations, precedences and tasks.

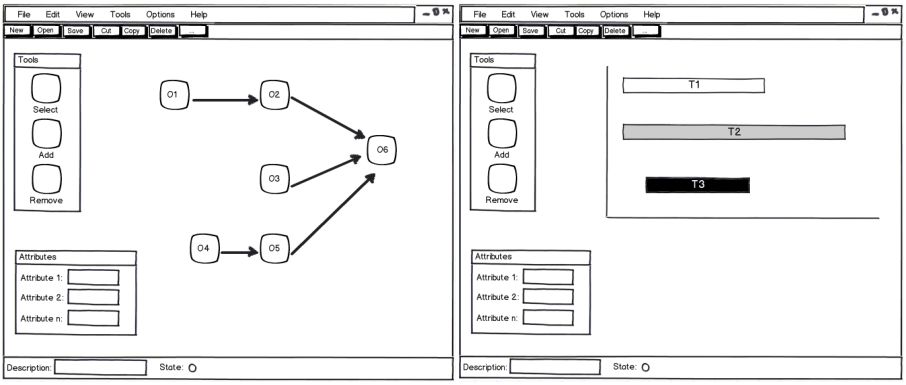


Fig. 2. Interaction model: task definition (left) and scheduling problem definition (right)

The design phase achievement is the starting point for the prototyping phase. The development of a prototype will not only evaluate all analytical work, but also the usability of the system.

4 Prototype

The developed prototype considered at this paper consists of two modules [11]: the Task Module (A) and the Scheduling Problem Module (B). They are responsible for the input data, i.e., for the definition of the scheduling problem. The system's global view is illustrated in Figure 3. The information flow between each component of the system is represented by the arrows.

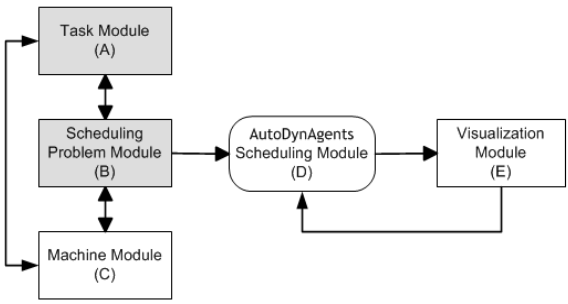


Fig. 3. System global vision

In the Task Module the user can create a new task or edit an existing one. An operation is the basic unit of a task, since a task is defined as a sequence of operations. At Scheduling Problem Module a user can create a new scheduling problem or edit an existing one. A task is the basic unit for the definition of a scheduling problem. Each task and scheduling problem specification is stored in a XML file (Figure 4), and after this the content of the files is sent to the AutoDynAgents Scheduling Module (D) in order to produce a scheduling plan.

```
<?xml version="1.0" encoding="UTF-8"?>
<task description="Task 1"shopFloorFile="C:\hopFloor_1.xml" state="invalid">
    <operations>
        <operation>
        <operation height="35.0" id="O2" value="Operation 2" width="70.0"
            xCoord="312.0" yCoord="161.0">
            <machines>
                <machine id="m2" t_proc="2.0"/>
            </machines>
            <precedences>
                <precedence op="O1"/>
            </precedences>
        </operation>
        <operation>
    </operations>
</task>
```

Fig. 4. Task XML specification

4.1 Task Module

The Task Module (Figure 5) allows the user to define, in a graphic and interactive way, a set of operations to perform a specific task. A task corresponds to a sequence of operations that determinates the order which the operations should be performed. An operation can take as precedence another operation or a set of operations. Each operation is characterized by a description, a machine where is performed, and by the processing time in the respective machine.

This module includes the following components: menu bar (A), tool bar (B), tool window (C), attribute window (D), working area (E), and state bar (F).

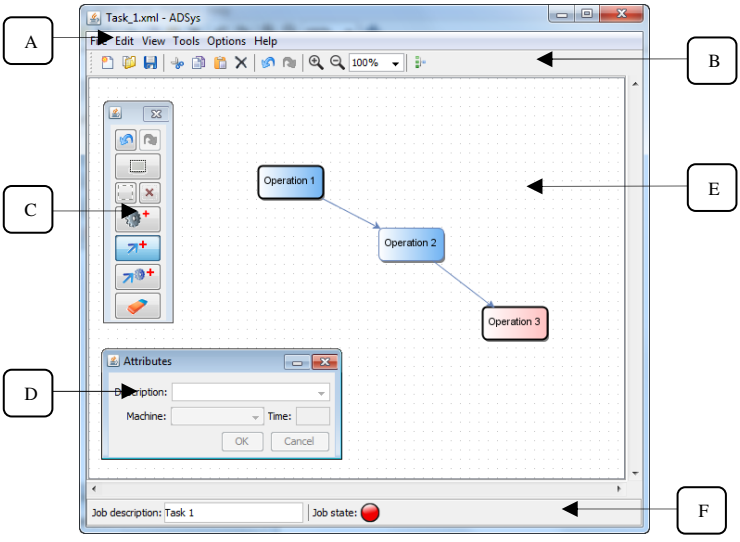
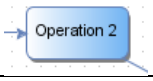
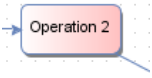
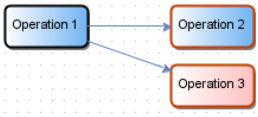
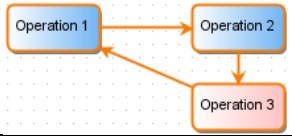


Fig. 5. Task module

The menu bar and the tool bar offer a set of options and tools that allow working the task (e.g. manage files, zoom in and out or auto-layout). The undo and redo options stand out for their usefulness and the auto-layout tool stand out because it allows structuring operations automatically. From the tool window it is possible to insert, delete and select operations and precedences. The working area is used to graphically represent and manipulate (move and resize) the interaction objects (operations and precedences). The attributes window is used to visualize, insert and edit the attributes of each operation. In the state bar the user can know the global state of the task.

To ensure the correct task definition a set of real time validations (Table 1) were developed. This means that any change in the task made by the user is immediately validated. In case of validation failure, the color of the object that caused the failure changes and a tooltip message is presented so that the user can understand the failure cause. The performed validations are: isolated operations (T1.2) - a task with more than one operation cannot have isolated operations; precedences (T1.3) - a precedence must be linked to a origin operation and a target operation; between the same origin and target operation only one precedence is allowed, i.e., repeated precedences are not allowed ; attributes (T1.4) - all attributes of an operation must be defined (description, proceeding time and machine); final operations (T1.5) - a task can only have one final operation; cycle (T1.6) - the existence of a cycle (a set of operation that forms a closed net) or cycles is not allowed in a task; task description: every task must have a description.

Table 1. Some validations made in the Task Module

Correct operation (T1.1)	**Isolated operations (T1.2)**	**Precedences (T1.3)**
		
Tooltip: • Operation 2	Tooltip: • Missing relationship	Tooltip: • Missing terminal connection • Missing origin connection • Repeated precedence
Attributes (T1.4)	**Final operations (T1.5)**	**Cycle (T1.6)**
		
Tooltip: • Missing operation description • Missing machine ID • The processing time must be greater than zero	Tooltip: • There is more than one final operation	Tooltip: • This operation is part of a cycle

When all validations are performed successfully (similarly to the particular case illustrated in T1.1 where the interaction object have blue edges and fill), the user is notified by a green color signal located on the status bar. Otherwise the signal will present a red color (Figure 5).

4.2 Scheduling Problem Module

The Scheduling Problem Module (Figure 6) enables the scheduling problem definition. A scheduling problem corresponds to a set of tasks to be performed in a given period of time. Each task is characterized by a release date, a due date, a priority, and by the amount to be produced.

In order to preserve the look and feel of the system and thereby minimize the learning effort/cognitive load on the user, this module's structure is similar to that of the Task Module.

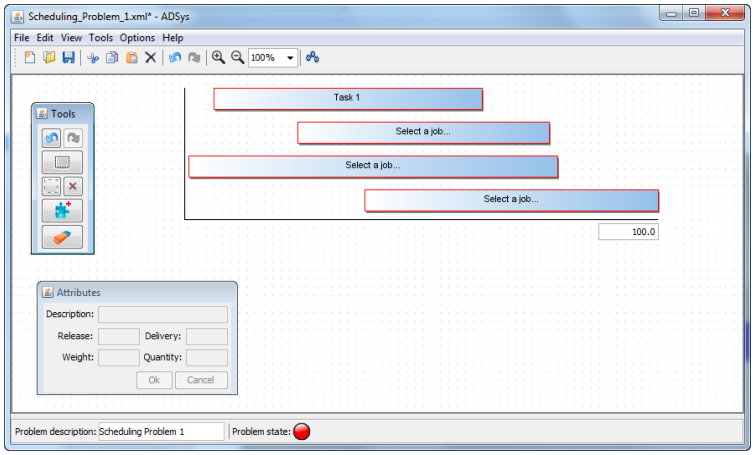


Fig. 6. Problem Module

From the tool window it is possible to select, insert, and delete tasks. The working area is used to graphically represent and manipulate the interaction objects (tasks). When the user horizontally moves a task, the duration is maintained while the release and due date are redefined. When the user resizes a task, one of the dates (release date or due date) is maintained while the duration and the other date (due date or release date) is redefined.

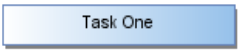
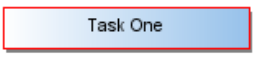
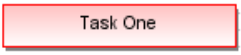
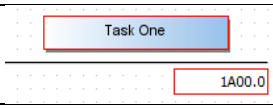
The maximum value of the time scale can be edited by the user by entering the desired value directly in the text box. The objects are automatically adapted to the new scale.

Now let's discuss the available vertical movements. The user can switch one or multiple tasks. The positional exchange of tasks includes selecting multiple tasks that may or may not be adjacent. However, the tasks to be handled are adjacently placed at the target position. Some actions can change the position of the tasks in the working

area: deleting a task (when a task is deleted the free space is also deleted so that there is no unoccupied spaces); cutting, copying and pasting a task (when a task is cut it is removed from the working area; when a task is copied a clone is made; when a task is pasted it is inserted below the last task); inserting a new task (a new task is inserted below the last task in the working area).

As in the Task Module, a set of validations (Table 2) is performed: set of operations - a task must have an associated set of operations; state of the operation set - the set of operations associated with a task can be valid or invalid; attributes - all attributes of a task must be defined (release date, due date, priority, and quantity); maximum time value - must be an integer number greater than zero; scheduling problem description: every scheduling problem must have a description.

Table 2. Some validations made in the Scheduling Problem Module

Correct task (T2.1)	**Set of operations (T2.2)**	**State of the operation set (T2.3)**
		
Tooltip: • Task One	Tooltip: • Missing job association	Tooltip: • Invalid job associated
Attributes (T2.4)	**Maximum time value (T2.5)**	**Scheduling problem description (T2.6)**
		
Tooltip: • Due date must be smaller or equal to the max value • The job quantity must be greater than zero • Release date must be positive	Tooltip: • The max value date must be an integer	Tooltip: • Missing problem description

When all validations are performed successfully (similarly to the particular case illustrated in T2.1 where the interaction object have blue edges and fill), the user is notified by a green color signal located on the status bar. Otherwise the signal will present a red color (Figure 6).

5 Usability Evaluation

In order to evaluate the prototype usability a script was prepared. This script has the following elements: an introduction to contextualize the participants in order to understand the purpose of the evaluation session, the tasks to be performed and the

type of questions presented; a initial survey with the objective of drawing the participants profile and evaluating their attitude towards a scheduling system; a prototype test where the participants had to define a specific scheduling problem; and final survey that is answered after the prototype test.

The scheduling problem used in the prototype test was based on [12] and consisted of a set of seven tasks with different release and due dates, priority and quantities. To each task, a sequence of eight to eleven operations had to be associated. In order to do that the participants had to create five different operation sequences. For each participant the time to scheduling problem completion was recorded (Figure 7).

The average time spent by each participant in creating the scheduling problem was approximately 33 minutes with a standard deviation of 5 minutes. As expected the participants that spent less time defining the scheduling problem where the ones with greater experience in using scheduling problems.

As can be seen in Figure 7 it was recruited 10 participants for evaluation session. This number, according with [13], allows to identify approximately 90% of usability problems. Some of the participants never used a scheduling systems, some had just used one, and others more than one. The reasons for using scheduling system can be academic or professional, or both.

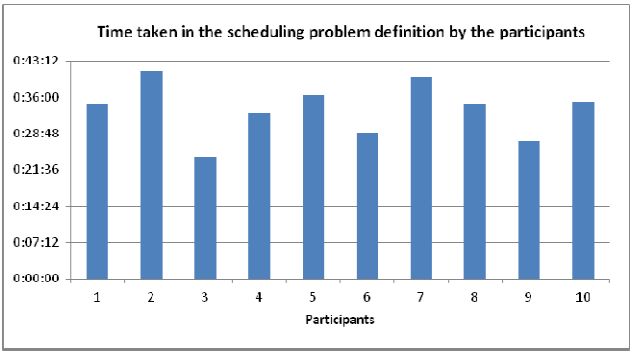


Fig. 7. Time taken in the scheduling definition

After the prototype test the participants filled out a survey, where they were asked about several interaction aspects and about the ease of use and usefulness of the features offered. The participants evaluated positively the prototype in its various aspects: overall structure, ease of use, tools provided, and the articulation between the two modules. After analyzing the results survey we can see that the participants, besides considering the scheduling problem complex, classified the prototype as enough for a scheduling problem definition, as adequate (best rate), and as familiar (Figure 8).

The participants also contributed with comments and improvement suggestions such as the use of hot keys, or the possibility to resize operations.

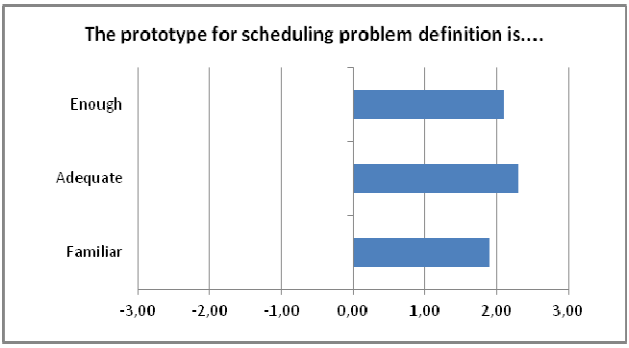


Fig. 8. Prototype global opinion

6 Conclusions and Future Work

This work represents the first iteration of the star life cycle and allowed us to build a user interface that provides an interactive and graphical way for problem scheduling definition.

In addition, the articulation achieved between the two modules contributed to the ease of interaction of the system. The real time validation mechanism enables the user to visually identify failures or errors in setting the scheduling problem.

As future work the improvements identified in usability evaluation will be implemented and the articulation between the developed modules and the scheduling module will also be improved. It is also intended that validation results, besides relying on the use of different colors, can be identified by using shapes in objects representation, thus taking into account users with color-perception deficiencies [2].

Acknowledgments. This work is supported by FEDER Funds through the "Programa Operacional Factores de Competitividade - COMPETE" program and by National Funds through FCT "Fundação para a Ciência e a Tecnologia" under the project: FCOMP-01-0124-FEDER-PEst-OE/EEI/UI0760/2011.

References

1. Pinedo, M.: Planning and Scheduling in Manufacturing and Services. Springer, New York (2009)
2. Hix, D., Hartson, H.R.: Developing User Interfaces: Ensuring Usability Through Product & Process. John Wiley & Sons, Inc. (1993)
3. Hewett, T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G., Verplank, W.: ACM SIGCHI Curricula for Human-Computer Interaction (1996)
4. Galitz, W.O.: The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques. Wiley Publishing Inc. (2007)
5. Karray, F., Alemzadeh, M., Saleh, J.A., Arab, M.N.: Human-Computer Interaction: Overview on State Art. International Journal on Smart Sensing and Intelligence (2008)

6. Nielsen, J.: Usability Engineering. AP Professional (1993)
7. Meech, J.F.: Contextualizing User Interfaces for Complex Systems. PhD's Thesis at Birmingham University (1999)
8. Lekin, http://community.stern.nyu.edu/om/software/lekin/
9. Sistrade Scheduling, http://www.sistrade.pt/pt/Doc_PDF/Sistrade_ERP_Scheduling.pdf
10. JD Edwards EnterpriseOne Production Scheduling, http://www.oracle.com/us/media1/e1-scm-ds-production-scheduling-065817.pdf
11. Piairo, J.E.F.: Desenvolvimento de Interface Inteligente para Suporte à Gestão e Controlo de Produção. Master's Thesis at Faculdade de Engenharia da Universidade do Porto (2012)
12. Pereira, A.M.M.: Aplicação de Meta-Heurísticas ao Problema de Escalonamento em Ambiente Dinâmico de Produção Discreta. PhD´s Thesis at Universidade do Minho (2003)
13. Virzi, R.A.: Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? Human Factors 34(4) (1992)

Perceived Site Security as a Second Order Construct and Its Relationship to e-Commerce Site Usage

Edward Hartono[1], Ki-Yoon Kim[2], Kwan-Sik Na[3], James T. Simpson[4,*], and David Berkowitz[4]

[1] Department of Accounting & MIS
Alfred Lerner College of Business & Economics,
University of Delaware
Newark, DE 19716
[2] School of Business
Kwangwoon University
26 Kwangwoon-gil Wolgye-dong, Nowon-gu
Seoul, 139-701, Republic of Korea
[3] Department of Management Information Systems
Seowon University
241 Musimseoro, Hungduk-gu,
Cheongju-shi Chungbuk, 361-742, Republic of Korea
[4] Department of Management, Marketing and Information Systems
College of Business Administration
University of Alabama in Huntsville
301 Sparkman Drive, Huntsville, AL 35899
hartono@udel.edu, min1203@kw.ac.kr, ksna@seowon.ac.kr,
{simpsonj,berkowd}@uah.edu

Abstract. Understanding *perceived site security* has been identified as a critical issue in current B2C e-commerce practices and research. A literature review reveals that the perceived site security construct has been conceptualized as a multi-dimensional construct; yet, all empirical studies have used uni-dimensional measures. Based on the CIA triad framework, we posit that perceived site security is a second order construct consisting of three distinct aspects of perceived site security: (1) perceived site confidentiality, (2) perceived site integrity, and (3) perceived site availability. The results from this study reveal that the new second order structure of perceived site security measure is superior to both the group factor structure and one factor structure. The results of the structural model also support the nomological validity of the newly conceptualized perceived site security measure.

Keywords: Security; e-commerce, technology acceptance model (TAM), second order factor, CIA triad, measurement.

* Corresponding author.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1075–1085.
DOI: 10.1007/978-3-642-36981-0_102

1 Introduction

Business-to-consumer (B2C) e-commerce occurs when the retail consumer purchases a product through the Website that a seller builds to sell its products through the Internet [1]. More consumers choose to purchase products through B2C e-commerce Websites because of various benefits such as more convenience (any time, any place), lower prices, and wider selection. As a result, more businesses have built B2C e-commerce Websites.

As a public and global network, the Internet is exposed to security vulnerabilities [18]. Unauthorized parties can break into the systems and gain unauthorized access to the B2C e-commerce Websites as well as transaction specific information. Clearly, these incidents will be detrimental to the consumers.

As consumers become more educated and aware of this issue, their decision to use, or not to use, a B2C e-commerce Website to make a purchase may well be influenced by the extent to which they believe that the Website is able to safeguard their personal, confidential information from security breaches [23].

Hence, understanding perceived site security has been deemed essential to the development of successful websites [21]. A number of studies have empirically investigated the impact of perceived site security in B2C e-commerce Website usage [15], [21], [25]. A review of these studies reveals that the perceived site security construct has been measured in a variety of ways. While no unified, agreed-upon method for measuring this construct currently exists, all studies have measured perceived site security as a uni-dimensional construct.

Our study aims to contribute to research in perceived site security by: (1) theoretically examining the dimensionality of perceived site security construct, (2) developing a multidimensional measure of this construct (second order construct), and (3) testing the multidimensional measure of this construct in the context of B2C e-commerce Website usage.

2 Perceived Site Security

Within the context of B2C e-commerce Websites, perceived site security has been defined as the online buyer's belief about the online seller's ability to safeguard his or her information from security breaches during data entry, processing, transmission, and storage [18], [21].

Perceived site security has been measured by previous studies in a variety of ways, all of which are uni-dimensional. Table 1 lists a sample of these measures.

During the last 5 years, the issue of information security has drawn considerable attention among scholars and practitioners in disciplines such as management information systems, computer science, and electrical and computer engineering [22]. A review in information security literature reveals that information security had been viewed widely as having three basic dimensions, (1) confidentiality, (2) integrity, and (3) availability [22].

Table 1. A Survey of the Perceived Site Security Measures[1]

Studies	Indicators of Perceived Site Security
[5], [21], [25]	1. "I would feel secure sending sensitive information across the World Wide Web", 2. "The World Wide Web is a secure means through which to send sensitive information"
[16], [17]	"The Internet e-banking systems:" 1. "Restrict unauthorized access", 2. "Protect customer private data", "Have rigorous security control"
[8]	1. "I feel secure to perform this task on the handheld computer", 2. "There is feedback indicating the information is protected"
[6]	1. "Internet vendors implement security measures to protect Internet shoppers" 2. "Internet vendors have the ability to verify Internet shoppers' identity for security purposes"

[1] See `http://cba.uah.edu/simpsonj/percieved-site-security.pdf` for complete tables, appendices, measure development, measurement model.

Confidentiality refers to the degree to which improper disclosure of information is anticipated and prevented. Integrity refers to the degree to which improper modification of information is anticipated and prevented. Availability refers to the degree to which information is available to authorized subjects when required [24].

These three dimensions are widely known as the CIA triad. They are accepted broadly as important facets of security by all traditions of information-related security such as computer science, management information systems, and computer and electrical engineering. The inclusion of these three dimensions in any security related issue ensures the coverage of the major components of security [22].

Based on the CIA triad framework, we argue that measure of perceived site security should include these three dimensions. Perceived site security, therefore, could be conceived as a multi-dimensional, composite construct, a second-order factor consisting of the commonality among three dimensions: confidentiality, perceived site integrity, and perceived site availability.

Hence, we hypothesize that:

> **Hypothesis 1:** Perceived site security as a second order construct of perceived site confidentiality, perceived site integrity, and perceived site availability is a better construct than perceived site security as a first order factor.

3 Perceived Security and Technology Acceptance Model (TAM)

To test the nomological validity of the newly conceptualized perceived site security construct, we will assess how well this construct relates to other constructs to which perceived site security should be theoretically related. A number of studies have incorporated perceived site security into TAM as an additional belief that is important in determining the actual usage of B2C e-commerce Websites [15], [21]. The results of these studies have demonstrated significant relationships between perceived site security and various TAM constructs.

Therefore, we will incorporate the newly conceptualized construct of perceived site security into TAM to assess how well this construct relates to the other TAM constructs to which perceived site security have been shown to be significantly related. Figure 1 presents our conceptual framework that delineates drivers of actual use of B2C e-commerce Websites. Such drivers include perceived site security, which, in this study, is measured as a second order construct of perceived site confidentiality, perceived site integrity, and perceived site availability.

The seven standard TAM hypotheses that will be re-tested in this study are:

Hypothesis 2. Perceived ease of use positively correlates to perceived usefulness of B2C e-commerce Websites.
Hypothesis 3. Perceived ease of use positively correlates to attitude toward using B2C e-commerce Websites.
Hypothesis 4. Perceived usefulness positively correlates to attitude toward using B2C e-commerce Websites.
Hypothesis 5. Perceived ease of use positively correlates to intention toward using B2C e-commerce Websites.
Hypothesis 6. Perceived usefulness positively correlates to intention toward using B2C e-commerce Websites.
Hypothesis 7. Attitude positively correlates to intention toward using B2C e-commerce Websites.
Hypothesis 8. Intention positively correlates to actual usage of B2C e-commerce Websites.

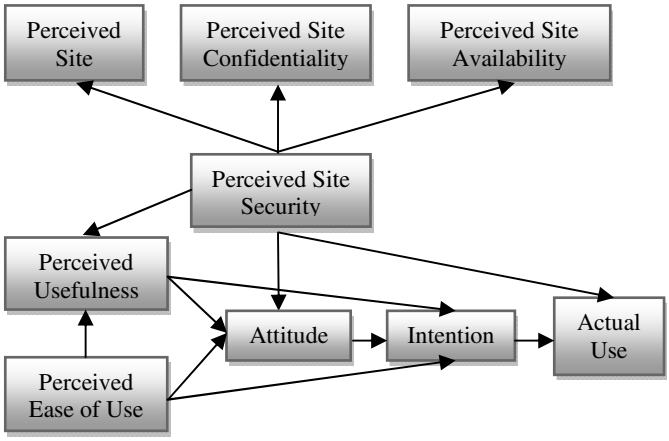


Fig. 1. The Conceptual Framework of B2C E-Commerce Website Actual Usage

4 Perceived Site Security and Attitude

Theory of Reasoned Action (TRA) [9] argued that a person's attitude towards the performance of a specified behavior (e.g., use a B2B website) is determined by the person's salient beliefs about consequences of performing the behavior and his/her evaluation of these consequences. TAM has adapted TRA for modeling user acceptance of information systems [7].

Perceived site security, by definition (see section 2 above), is a belief construct. TRA posits that belief influences attitude. Hence, we have incorporated perceived site security to TAM by simply adding another belief variable that positively influences user attitude toward using a system.

Positive relationship between perceived site security and attitude can be explained by applying the theory of perceived risk in consumer behavior. The fact that buyers and sellers do not interact face-to-face in online transaction causes the buyers to perceive that the sellers are more likely to behave in an opportunistic manner, and vice versa [11]. Hence, the buyers and the sellers will have greater security concerns on online transactions than on conventional transactions [15]. A buyer that perceives that a B2C e-commerce Website has a low level of security, and in turn a high level of risk, will likely form a negative attitude toward using this B2C e-commerce Website which will negatively influence his or her actual usage of this B2C e-commerce Website.

Considering this premise, we hypothesize that:

> **Hypothesis 9.** Perceived site security is positively correlated to attitude toward using the online shopping mall.

5 Perceived Site Security and Perceived Usefulness

Gefen, Karahanna, and Straub [11] suggest that e-commerce application usefulness can be classified into two groups. First, the current or short term usefulness allows the consumers to reduce transaction cost by, for example, purchasing a product without having to visit the bricks and mortar store. Second, the future or long term usefulness prevents the consumers from paying additional transaction costs resulting, for example, from an unauthorized party breaking into the e-commerce database, stealing customers' credit card information, and using it to make unauthorized purchases. Increase in perceived site security should increase the buyer's belief that using an online shopping mall will enable him or her to gain, not only current, but also future net benefits.

Considering this premise, we hypothesize that:

> **Hypothesis 10:** Perceived site security is positively correlated to perceived usefulness of the online shopping mall.

6 Measure Development

Appendix A provides a sample of the measures for all of the constructs used (i.e., factors, latent variables) in the study. All constructs measure consumers' perceptions of B2C e-commerce Websites that they have used. Each construct is measured by using multiple indicators. All indicators were measured on a seven-point Likert-type scale. All of the scales were anchored at 1 as "strongly disagree" and 7 as "strongly agree".

All measures were developed on the basis of similar constructs that have been developed and validated in various studies. The measures of the TAM constructs were adapted from those developed and validated by prior studies**.** The measures of perceived site confidentiality, perceived site integrity, and perceived site availability as the three dimensions of perceived site security were developed on the basis of information security constructs that have been developed and validated by prior studies.

A total of 489 respondents from three organizations (i.e., university, private company and government office) in Seoul, South Korea participated in on site interviews about their experience in using their most frequently used online shopping mall. 436 usable responses were collected for a response rate of 89 percent.

7 Measurement Model

Consistent with the work of Anderson and Gerbing [3], we examined the measures for their content validity, construct reliability, convergent validity, and discriminant validity. This approach has been used in numerous studies.

Since the perceived site security is a second order construct, we evaluated the measures by running confirmatory factor analysis (CFA) on two models [19]: (1) the perceived site security measurement model that includes perceived site security as a second-order construct derived from three first order constructs, namely perceived site confidentiality, perceived site integrity, and perceived site availability, and (2) the combined measurement model that includes all 25 indicators and all 9 constructs.

The Perceived Site Security Measurement Model. We conducted a CFA on this measurement model by using AMOS 6.0. While the χ^2 statistics ($\chi^2 = 56.497$ ($p<0.01$) and χ^2/d.f. = 2.354) suggests a lack of fit between the model of data, the use of χ^2as a measure of model fit for samples that exceed 200 has been widely criticized [13]. Other popular fit indices reported in Table 3 exceed the norms and suggest that the data fits the model very well.

To provide more support to the second-order construct conceptualization, we also conducted CFAs on two alternative measurement models: (1) group-construct model (i.e., three first order constructs), and (2) one-construct model [14], [20] (see diagram on website http://cba.uah.edu/simpsonj/percieved-site-security.pdf). The results reported in Table 2 shows that these two models are significantly inferior to the second-order construct model. Therefore, our empirical results suggest that the data structure fit our first hypothesis that the perceived site security is a second-order construct structure.

Table 2. Goodness-of-fit Indices and χ2 Difference Tests of Second Order Construct Structure to One Factor Structure and Group Factor Structure

Goodness-of-fit Indices

	χ2	d.f.	GFI	RMR	RMSEA	NFI	IFI	CFI
Second Order Construct Model	56.497	24	.973	.064	.056	.970	.982	.982
Group-Construct Model	309.439	27	.875	.445	.155	.834	.846	.845
One-Construct Model	534.795	27	.780	.311	.208	.713	.723	.722

χ^2 **Difference Test**

	χ2 difference	d.f. difference	p
Second Order Construct Structure versus Group Factor Structure	252.942	3	<.01
Second Order Construct Structure versus One Factor Structure	478.298	3	<.01

Combined Measurement Model. We also subjected the combined measurement model to a CFA using AMOS 6.0. This measurement model was deemed to fit the data well as various fit indices (χ^2 = 404.68 (p<0.01) and χ^2/d.f. = 1.73; GFI = .93, RMR = .071, RMSEA = 0.04, NFI = .92, IFI = .97, CFI = .97) exceed the norms (i.e., GFI, NFI, IFI, CFI > .90 [4] and RMR, RMSEA < .08 [12]. All path coefficients were all significant at p < 0.01.

An analysis of composite reliabilities [10] suggests high construct reliability (i.e., > .70). Convergent validity was evidenced by the high average variance extracted (AVE) for each construct [10]. We examine the discriminant validity by comparing the AVE with the square of the correlation coefficients between any two constructs. The AVE for each construct well exceeds the square of the correlations between that construct and any other construct, hence, indicates discriminant validity [10]. Finally, a check for common method variance that has been used in published studies [2] suggests that common method bias is not a serious cause for concern. These analyses taken together suggest that the measures used in this study adequately represent the theoretical constructs they are designed to measure.

8 Structural Model

The structural model used to test the hypotheses consisted of 6 constructs tested in the measurement model. The structural model fit indices indicate acceptable fit with the

data (χ^2 = 461.56 (p<0.01) and χ^2/d.f. = 1.93, GFI = .92, RMR = .11, RMSEA = 0.046, NFI = .91, IFI = .96, CFI = .96).

The original TAM hypotheses are all supported. We find that the relationships among TAM constructs are all positive and significant, hence, consistent with prior TAM studies.

We find that perceived site security is positively and significantly associated with attitude, in support of hypothesis 9, and with perceived usefulness, in support of hypothesis 10 (see Table 3).

Table 3. Results of Structural Model

			Unstandardized Loading	Standard Error	p
Perceived Ease of Use	→	Perceived Usefulness	.59	.09	p < .01
Perceived Security	→	Perceived Usefulness	.37	.19	p < .05
Perceived Ease of Use	→	Attitude	.43	.08	p < .01
Perceived Usefulness	→	Attitude	.31	.06	p < .01
Perceived Security	→	Attitude	.35	.16	p < .05
Attitude	→	Intention	.66	.06	p < .01
Perceived Usefulness	→	Intention	.11	.05	p < .05
Perceived Ease of Use	→	Intention	.15	.07	p < .05
Perceived Security	→	Intention	.17	.13	p = .20
Intention	→	Actual Usage	.59	.06	p < .01
Perceived Site Security	→	Perceived Site Integrity	3.34	.63	p < .01
Perceived Site Security	→	Perceived Site Confidentiality	2.75	.53	p < .01
Perceived Site Security	→	Perceived Site Availability	1.00		

9 Conclusions

In general, the empirical results are encouraging and provide support for the main objective of this study which is the development of a valid and reliable measure of site security that provides a more complete structure of the components of perceived site security.

Based on the CIA triad, we propose three distinct aspects of perceived site security: (1) perceived site confidentiality, (2) perceived site integrity, and (3) perceived site

availability. We posit that perceived site security is best conceived as a second order construct consisting of the commonality of these three first order factors.

The results from comparing the second order structure of perceived site security measure to the alternative group factor structure and one factor structure, demonstrate that the second order structure is superior to the two alternatives on all major evaluation criteria. Moreover, our results reveal that the second order factor structure provides the first reliable and valid, multi dimensional, CIA triad based measure of perceived site security.

The findings of the structural model (see Table 6) further demonstrate the quality of the new measure by supporting the nomological validity of the newly conceptualized perceived site security measure. The second order structure of perceived site security reveals that the new measure behaves as expected when evaluated in a theory based nomological network.

For managers, the results of our study provide a better understanding about perceived site security. Prior studies have demonstrated to managers that buyers' belief toward B2C e-commerce Website security performance significantly impacts actual e-commerce usage. Yet, the results of our study pinpoint the dimensions of this security performance, namely confidentiality, integrity, and availability. As a result, managers may develop strategies to influence buyers' beliefs toward the ability the B2C e-commerce Website to protect each of these dimensions.

For researchers, the results of our study demonstrate the advantages of future researchers' adopting multidimensional measure for perceived site security construct. Prior to this study, security construct was conceptualized as single dimension construct. Future studies should recognize the multidimensional nature of this construct.

The interpretation of our results is subject to certain limitations. First, our empirical results must be considered in the context of the particular subjects included in the study. The exclusive use of Korean respondents has the advantage of excluding unwanted confounding factors resulting from cultural differences; yet the generalizability of the results is likely reduced. Second, in an effort to develop a reasonably parsimonious model, all possible variables used in the models reported in the literature could not be included. These variables include additional dimensions of perceived site security that have been added over the years. Future research might include these additional dimensions. Third, the actual usage is measured subjectively. Future research can replicate or extend this study by using more objective measure of the actual usage. Finally, the use of cross-sectional data allows us to examine only a "snapshot" of the impact of various antecedents on B2C e-commerce Website actual usage. Perhaps future studies will consider the use of longitudinal data which would reveal the dynamic of this phenomenon over an extended period of time.

Acknowledgment. The authors would like to thank the Integrated Enterprise Lab for support of this project.

References

1. Alzola, L.M., Robaina, V.P.: Quality in electronic commerce B2C: Perceptions, expectations and importance. Service Industries Journal 26, 671–687 (2006)
2. Amoako-Gyampah, K., Meredith, J.R.: Examining cumulative capabilities in a developing economy. International Journal of Operations & Production Management 27, 928–950 (2007)
3. Anderson, J.C., Gerbing, D.W.: Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin 103, 411–423 (1988)
4. Bentler, P.M., Bonnet, D.G.: Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin 88, 588–606 (1980)
5. Cheng, T.C.E., Lam, D.Y.C., Yeung, A.C.L.: Adoption of Internet banking: An empirical study in Hong Kong. Decision Support Systems 42, 1558–1572 (2006)
6. Cheung, C.M.K., Lee, M.K.O.: Understanding consumer trust in Internet shopping: A multidisciplinary approach. Journal of American Society for Information Science and Technology 57, 479–492 (2006)
7. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User acceptance of computer technology: A comparison of two theoretical models. Management Science 35, 982–1003 (1989)
8. Fang, X., Chan, S., Brzezinski, J., Xu, S.: Moderating effects of task type on wireless technology acceptance. Journal of Management Information Systems 22, 123–157 (2005-2006)
9. Fishbein, M., Ajzen, I.: Belief, attitude, intention, and behavior. Addison-Wesley, Reading (1975)
10. Fornell, C., Larcker, D.: Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 18, 39–50 (1981)
11. Gefen, D., Karahanna, E., Straub, D.W.: Trust and TAM in online shopping: An integrated model. MIS Quarterly 27, 51–90 (2003)
12. Hu, L.T., Bentler, P.M.: Fit indices in covariance structure modeling: Sensitivity to underparameterization model misspecification. Psychological Methods 3, 424–453 (1998)
13. James, L.R., Stanley, A., Mulaik, S.A., Jeanne, J.M.: Causal analysis: Assumptions, models and data. Sage, Beverly Hills (1982)
14. Kim, S., Stoel, L.: Dimensional hierarchy of retail Website quality. Information & Management 41, 619–633 (2004)
15. Lian, J.W., Lin, T.M.: Effects of consumer characteristics on their acceptance of online shopping: Comparisons among different product types. Computers in Human Behavior 24, 48–65 (2008)
16. Liao, Z., Cheung, M.T.: Service quality in Internet e-banking: A user-based core framework, e-Technology, e-Commerce and e-Service. In: IEEE International Conference on e-Technology, e-Commerce, and e-Service (EEE 2005), Hong Kong, pp. 628–631 (2005)
17. Liao, Z., Wong, W.K.: The determinants of customer interactions with Internet-enabled e-banking services, working paper no. 0701. Department of Economics, National University of Singapore (2007)
18. Pavlou, P., Liang, H., Xue, Y.: Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective. MIS Quarterly 31, 105–136 (2007)
19. Ramani, G., Kumar, V.: Interaction orientation and firm performance. Journal of Marketing 72, 27–45 (2008)
20. Rindskopf, D., Rose, T.: Some theory and applications of confirmatory second-order factor analysis. Multivariate Behavioral Research 23, 51–67 (1988)

21. Salisbury, W.D., Pearson, R.A., Pearson, A.W., Miller, D.W.: Perceived security and World Wide Web purchase intention. Industrial Management & Data Systems 101, 165–176 (2001)
22. Siponen, M.T., Oinas-Kukkonen, H.: A review of information security issues and respective research contributions. The DATABASE for Advances in Information Systems 38, 60–80 (2007)
23. Tao, W., Kim, C., Kim, K.S., Shin, N.: An empirical study of customers' perceptions of security and trust in e-payment systems. Electronic Commerce Research and Applications 9, 84–95 (2010)
24. Tsiakis, T., Sthephanides, G.: The concept of security and trust in electronic payments. Computers & Security 24, 10–15 (2005)
25. Vatanasombut, B., Igbaria, M., Stylianou, A.C., Rodgers, W.: Information systems continuance intention of web-based applications customers: The case of online banking. Information & Management 45, 419–428 (2008)

Appendix A: Samples of Indicators

To each online buyer, we ask the extent to which he or she agrees with the following statements:

To each online consumer, we ask the extent to which he or she agrees with the following statements:

Constructs	Indicators
Perceived Site Confidentiality	Someone uses my Internet ID to read my transactional information. [R]
Perceived Site Integrity	My transactional information is altered. [R]
Perceived Site Availability	I cannot order due to system failure. [R]
Perceived Ease of Use	It is easy to place an order
Perceived Usefulness	This site is very informative
Attitude	It is a smart idea to shop in this site
Intention	I will use this site again
Actual Use	I have frequently used this online shopping mall to search products

* Reverse scale

Accessibility Study in Sites of Public Higher Education Institution in Brazil

Mariana Angelo Lopes Sanches and Lílian Simão Oliveira

Faculdade de Tecnologia(Fatec) – Indaiatuba
– Rua Dom Pedro I, 65 – Cidade Nova
13334-100 São Paulo, Brazil
marianasanches@hotmail.com, liliansimao@gmail.com

Abstract. This paper describes the accessibility study of Fatec sites - public universities in Brazil. It describes the accessibility evaluation done in all Fatec sites using an automatic tool to identify the main problems for people with visual impairment. As a result, it was possible to obtain the most common problems in the websites and suggest some changes in order to make them more accessible.

Keywords: Human Computer Interaction, Accessibility, Websites, Visually Impaired.

1 Introduction

The most fundamental right of the citizen is to be treated with equality and the federal government reassures that through Decree Nº 5.296 from Dec. 02, 2004, which regulates Laws nº 10.048, from Nov.8, 2000, and nº 10.098, from Dec.19, 2000. In Chapter VI – Of the access to Information and Communication, art 47 describes that " [...] accessibility is mandatory on portals and electronic sites of public administration on the Internet, for the use of people who are visually impaired, and that guarantees access to information."

In addition to that, one of the public services that deserves attention is education. In order to reassure the Decree 5.296/2004, which regulated the law 10.098 on Dec.19, 2000 which establishes basic criteria and general rules to foster accessibility, in its article 24: "Teaching institutions of any level, stage or modality, public or private, will make access and use available of all environments for people with any kind of handicap or that have mobility restrictions, including classrooms, libraries [...]". In the same article, first paragraph, subsection II "it makes available to professors, students, civil servants and staff with any kind of handicap technical help which allows access to school and administrative work in equal terms to everyone else". As a result of this adequacy of sites to all kinds of users without any prejudice fosters citizenship and has an impact on educational, social and economic life.

Based on this context, government has made an effort to regulate sites which are of greater interest to the public audience[14]. However, the effort hasn't reached federal

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1087–1094.
DOI: 10.1007/978-3-642-36981-0_103

universities. This work aims at evaluating the degree of accessibility on the sites of a public university because academic sites play a crucial role for students. Nowadays they provide information such as: events at the campus, enrolment, access to the library, catalogues with courses, calendars, standard model for school work and also subjects given in class. Therefore, we realize that people who do not have access to that information is excluded from the academic activities.

According to researches [9] 23,9% of the Brazilian population have some kind of disability and the commonest is visual. A total of 528.624 people have this limitation. These numbers show that among all types of handicap, the one related to sight deserves more attention. Hence, it is possible to assist a greater number of people. Due to numbers, this work was concerned about evaluating the interface of websites to observe accessibility for the visually impaired.

This work investigates what the commonest problems are on the web sites of Brazilian public universities. The research specifically analyses the web sites from FATEC (Faculdade de Tecnologia do Estado de São Paulo - Brasil) from the visually impaired point of view. The specific objectives of this work are: i) Research and understand the main problems of accessibility on web sites for the visually impaired. ii)Investigate the seriousness and frequency of problems. iii) Analyse and suggest improvements for the developers to make them more aware of the issue. Finally, to show that some implementation procedures can interfere greatly on the quality of the site.

This article is organized in the following manner: Topic 2 describes the main concepts for the understanding of the work. Topic 3 presents other related works. Topic 4 describes the project, tools, methods and results. Topic 5 concludes the study.

2 Related Work

As we have continued with the research, we have come to realize that accessibility is being explored in many ways, also for legal reasons, although it hasn't become common practice yet. This has been seen on works that compile data of the current situation of the development of applications from the Accessibility point of view.

Some of the research refers to government public sites, as with the case presented by [14] which evaluated the portal of the Federal Government Network. This work used automatic software to evaluate the degree of accessibility, but the focus was to validate the site links and verify the quantity of existing problems.

At the same time that works such as [3], [4] and [17] present how accessibility on the internet is treated when information is accessed from Brazilian libraries [17] emphasizes the importance of accessibility from virtual libraries. [3] says that libraries from Brazilian federal universities are going through huge problems to comply with standards of inclusion to access to information on their websites. This research evaluates the central sites from federal libraries and has concluded that there is still a lot to be done in order to make users more independent. However, [4] presents in his work the resources of the Louis Braille Library from the Benjamin Constant Institute which assures everyone access to information, specially the visually

impaired. Despite that, he admits that the Louis Braille Library also needs to adapt to the new standard.

Still in the academic environment, there is the work of [7] which analyses the academic websites from Spanish universities. This research shows through Spanish laws that accessibility to the web must be applied to all public institutions, including universities.

Some of the works describe methods of research used to evaluate the accessibility of an application.

One example is a research from [6] which compiled data using questionaires as a method and the data was obtained from people involved in web development. The work presented descriptive statistics which showed there is still a long way to go for the use of the improvement recommendations observed in practice.

On the other hand, related to accessibility to the visually impaired, which is the focus of this research, we find works such as [15], in which there is not only data compiling and automatic tests, but also tests done with the users. By using methodology of qualitative nature, without restrictions to statistics and applying multicase research, the methodology according to [15], apud [2] "enphasizes the description, induction, theory and the study of personal perceptions". In this way, we have the visually impaired point of view that shows the interface problems and the result is a proposal for improvements for the analysed sites. As the work of [10] would say, it is a practical work, which evaluates the sites from the users' point of view, and it can be used as recommendation for developers and designers who are willing to create an inclusive web environment.

By analysing other more recent articles, [16] developed work with some accessibility evaluation methods, which show how to make a more inclusive environment. This work concludes that the accessibility proposals are easy to learn, quick to apply and of low cost.

Work [7] uses the W3C automatic validator, presents improvements that can be implemented in academic sites. [5] did a research about accessibility in European universities using various studies, among which are the ones for the visually impaired. Automatic methods of evaluation were used.

Another work which enphasizes the importance of accessibility in the virtual world is [17], which talks about the importance of having W3C accessibility guarantee, once it is the most complete and assures that the visually impaired are able to fulfill the tasks on the site. On a previous work from the same authors, it was established the relation among usability guidelines that contribute with interface accessibility [18].

The work of [1] made a comparison between the WCAG and e-MAG accessibility recommendations, and it concluded that the international standard is compatible with the Brazilian standard, and it must be used for evaluation of digital interfaces. After this observation, this work chose to use the WCAG recommendations.

After the bibliography compiling, it was observed that there are several methods to evaluate accessibility. Automatic tools of evaluation were used to identify the main problems and according to the level of seriousness of each one of them. Criteria to choose the tool was based on the WCAG recommendations. This study focused on the websites of Faculdade Estadual de Tecnologia of São Paulo, from the FATEC group.

3 The Project

3.1 The Research

The applied research aims at obtaining quantitative data about the accessibility level of the studied websites. The objective is to describe the main problems and suggest some changes so that the visually impaired can have access to information. In order to do the field research, the evaluation scope of public university sites was decreased to a study more focused on technological schools in the state of São Paulo – Brasil.

The Centro Estadual de Educação Tecnológica Paula Souza (CEETEPS) is responsible for Faculdades de Tecnologia(Fatecs), which offer university courses on technology and ETECs, which offer technical courses. CEETEPS is associated to the Department of Economic Development, Science and Technology of São Paulo State Government- Brazil. There are 209 Technical Schools (Etecs) and 56 Technology State Universities (Fatecs), in 159 cities. They started in 1969 and offer 65 university courses in various fields [11].

The system users in this study are enrolled at Fatec, and the visually impaired have extra attention. The students' average age is around 17 and variations can occur according to the course and time they attend.

Automatic validators were used to analyse the Fatecs Websites, and later on a revision about what should be done to improve accessibility. Currently, there are several evaluation tools. However, after a bibliography and experimental analysis [13] [12] [5], Hera was the chosen tool. It verifies the W3C/WAi - WCAG 1.0 guidelines and it is complete and easy to understand. There is an update of WCAG 1.0, version 2.0. However, according to [16] most of the automatic evaluators were not updated and there are many aspects that hinder understanding. Since version 1.0 is complete and the aim of the research is to obtain an initial overview about the selected sites for this research, this was the chosen tool.

3.2 Evaluations

The evaluations were done on the defined websites using the Hera tool. There are many options when using this tool to verify each item, the way the code is viewed , or on the page layout, the errors and warnings. In addition to that, it is possible to obtain HTML or PDF reports which are automatically generated by the tool. Another option that was extensively used to compile data about the commonest problems is the option of navigation through results, in which we have a table showing a summary of the main points to be improved. This table describes which ones are errors and which ones must be manually verified and they are sorted out according to priority. The priority level varies from 1 to 3, where 1 is more critical and must be addressed immediately.

According to [8] there are 56 Fatecs, and 54 have websites. From the 54, 12 sites were not available at the moment the research was done. Therefore, only 42 sites were analysed.

From the results obtained we made a sheet indicating the commonest and most serious problems. After that, the commonest errors were selected according to priority and suggestions were described in order to improve the website.

3.3 Analysis

Data generated from the evaluations was analysed and grouped according to priority. Priority 1 problems were arranged according to the frequency they occurred on the evaluated sites. Table 1 shows this information. Therefore, we verify that most of the Priority 1 problems that were presented consist of small details and adjustments at the time of implementation. Table 1 shows which W3C guidelines were infringed and the frequency in which the errors occurred.

Table 1. Frequency of errors related to W3C guidelines of priority 1

Errors – Related to W3C Guidelines	Frequency of occurrence
Guideline 1 – Check Point 1.1	28 Websites
Guideline 8 – Check Point 8.1	15 Websites
Guideline 6 – Check Point 6.3	9 Websites
Guideline 12 – Check Point 12.1	2 Websites

The first error is related to Ckeck Point 1.1 from W3C Guideline 1 which recommends "Provide equivalent alternatives to auditory and visual content.". This problem occurrs in about half of the analysed sites. According to W3C it is necessary to add textual to non-textual elements such as how to use *<alt>* or *<longdesc>* with explanatory texts for one image or other elements. These non-textual elements can be: images, graphic representations of the text, including symbols, animations, as it is the case of animated GIFs, applets and programmed objects, ASCII art, panels/frames, interpreting programs, images used in lists as signaling points of numbers, spaces, graphic buttons, sounds (reproduced with or without user's interaction), independent audio files, video showings. The objective of this guideline is to assure sites have at least one subtitle to the equivalent indication of the presented graphic element, i.e., to transmit textually what is viewed, in a way that users of assistive technology[2] know what is presented on the screen. Another thing presented by W3C in the same guideline is the use of *<alt>* and *<longdesc>*. The HTML attribute *<alt>* must be used to insert the subtitles of the corresponding texts, as previously presented. In addition to this, there are also cases where the images are just illustrations and do not need to be read. In this case it is advisable to use the command in the following way *<img alt=" ">*, in which it is ignored by the aid technologies. The *<longdesc>* command, on the other hand, is used to insert graphs and tables of analysis, for example, which allows the user to read the long description separately and then return to the initial page.

The second most frequent problem is Check Point 8.1 from Guideline 8 of W3C "Ensure direct accessibility of embedded user interfaces.", whose guideline says: "Make sure the programmable elements such as scripts and applets are directly accessible or compatible with the support technologies. This is the third most infringed guideline among the analysed sites - Check Point 6.3 from Guideline 6 "Ensure that pages featuring new technologies transform gracefully." which recommends "Make sure the pages are usable when scripts, applets, or other programmable objects are deactivated or are not supported. In case that is not possible, enter with equivalent information on an alternative page. "

This means that the pages which have script and applets, i.e., an object of programming that allows an interactive function, are not accessible to all users. Another alternative, for navigators that do not support scripting, is the use of the <*noscript*> command which will present an alternative content to the user. It is not advisable to use JavaScript in link, because the user who do not have support will not be able to access it.

Figure 1 describes the quantity of errors that the developer must correct, represented by the darker line at the bottom. Most sites present the range from 5 to 10 errors. The light color line at the top show the items that must be verified manually with the user. These numbers are higher than the errors and can be found on the range from 35 to 40 points.

According to Figure 1 many of the evaluated items could not be identified as errors or as correctly addressed and need manual verification. The quantitative presented in relation to these topics is high. Therefore, it is advisable to make tests with the user who uses assistive technologies, as it is the case for the visually impaired who uses the screen reader. Another option would be to an evaluation with domain experts -- be it by knowledge of similar projects, or by knowledge obtained through sharing experience with the visually impaired – by doing heuristic tests.

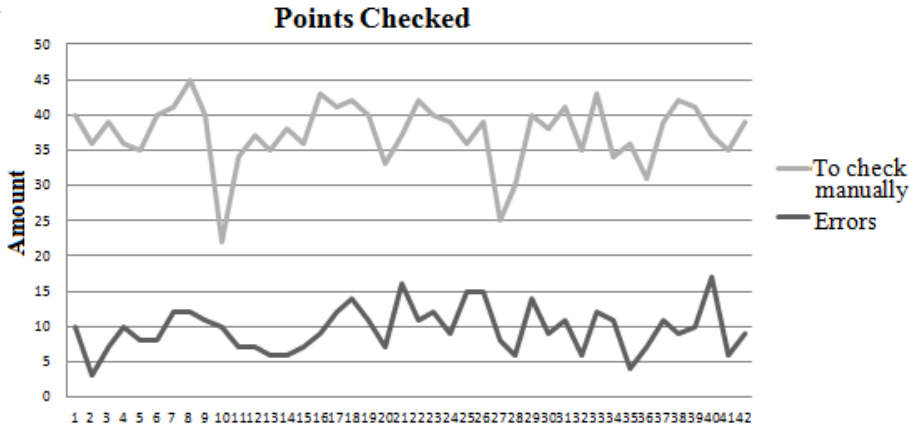


Fig. 1. Graph which represents the quantity of items that must be verified manually and the errors which infringe W3C Guidelines

After obtaining the quantitative, the error average and the items which need another evaluation were verified. It was also possible to identify the highest number of errors and items to be verified on a single website. In addition to that, we can see the lowest quantity of errors which need manual evaluation and their average. Table 2 shows these quantitatives. It shows how essential it is to the developer to be concerned about W3C Guidelines and use the alternative commands in order to make the site more accessible.

Table 2. Average of Errors and Items to be verified manually

Average	Errors	Items to Verify
Higher	17	45
Lower	3	22
Average	9,64	37,43

4 Conclusions and Future Work

Through data obtained during the analysis of the sites, we presented the commonest errors, the items to be verified and the average of incidents from the total of websites of Fatec schools. We concluded that there is a way to go to correct Priority 1 errors, and the developers must be aware of these corrections for future implementations. We suggest improvements that are fit to most sites, and this will mean a significant improvement.

However, it is extremely important that tests with different users be done after implementation, and mainly with the users who make use of assistive technologies. Among them , mainly the ones who are visually impaired because they represent a great number of the disabled population. Once a site is accessible to the visually impaired, it can become accessible to many other people such as: the elderly, children, the mentally impaired who use screen readers and the ones with partial sight.

It is based on this that we leave to future work the analysis with users or other people who are sympathetic to their needs through interviews, task follow-ups and heuristic tests. The Hera tool can also be used in future works to make a comparison of the website before and after. The tool makes the page of analysis results available for 7 days, and it is possible to make site adjustments following WCAG recommendations and the users' suggestions, and in the end compare the results and analysis before the corrections and after the improvements.

References

[1] Bach, C.F., Ferreira, S.B.L., Silveira, D.S.: Accessibility guidelines: A comparison between WCAG e e-MAG - Revista Eletrônica de Sistemas de Informação 8(1), artigo 1 (2009)

[2] Bogdan, R.C., Biklen, S.K.: Investigação Qualitativa em Educação: uma introdução à teoria e aos métodos. Porto Editora, Portugal (1994)

[3] Cortes, M.D.F., Lopes, M.L.: As Bibliotecas Universitárias Federais Brasileiras e a Acessibilidade das Informações em Seus Websites. Revista ACB: Biblioteconomia em Santa Catarina. Florianópolis 13(1), 117–129 (2008)
[4] Ferreira, A.F.B.C.: Biblioteca Louis Braille do Instituto Benjamin Constant: Assegurando ao Deficiente Visual Acesso ao Conhecimento. Revista ACB: Biblioteconomia em Santa Catarina. Florianópolis 282 13(1), 282–290 (2008)
[5] Espadinha, A.C.G.: Modelo de atendimento às necessidades educativas especiais baseado na tecnologia: estudo de caso centrado em alunos com baixa visão. Tese em Motrocidade Humana, especialidade de Educação Especial e Reabilitação – Universidade Técnica de Lisboa – (2010)
[6] Freire, A.P., Fortes, R.P., de, M.: Acessibilidade no desenvolvimento de sistemas web: um estudo sobre o cenário brasileiro. Dissertação de Mestrado – USP – Universidade de São Paulo. São Carlos – SP. 136 f (December 2008)
[7] Fernández, J.M., Roig, J., Soler, V.: Web Accessibility on Spanish Universities. Second International Conference on Evolving Internet. Valencia – Spain (September 2010)
[8] Guia de Internet, 2013 – Centro Paula Souza – Disponível em, http://www.centropaulasouza.sp.gov.br (acessado em: 05 de setembro de 2012)
[9] IBGE – Instituto Brasileiro de – Disponível em, http://www.ibge.gov.br/home/presidencia/noticias/noticia_visualiza.php?id_noticia=2125&id_pagina=1 (acessado em: 02 de Setembro de 2012)
[10] Melo, A.M., Baranauskas, M.C.C.: Design Inclusivo de Sistema de Informação na Web. In: Proceedings of VII Brazilian Symposium on Human Factors in Computing Systems, IHC 2006, Natal – RN (November 2006)
[11] Paula Souza – Site official do Centro Paula Souza – disponível em, http://www.centropaulasouza.sp.gov.br (acessado em: 26 de Novembro de 2012)
[12] dos Rodrigues, A.S., Souza Filho, G.L., Borges, J.A.: Acessibilidade na Internet para deficientes visuais. Revista do Núcleo de Computação Eletrônica. Rio de Janeiro: UFRJ, Disponível (2001), http://intervox.nce.ufrj.br/dosvox/textos/guido.doc (acessado em: 06 Setembro de 2012)
[13] Santana, V.F., Almeida, L.D.A., Baranauskas, M.C.C.: Aprendendo sobre Acessibilidade e Construção de Websites para Todos. Revista Brasileira de Informática na Educação, Porto Alegre – RS 16(3), Set/Dez (2008)
[14] Simão, J.B., Rodrigues, G.: Acessibilidade às informações públicas: uma avaliação do portal de serviços e informações do governo federal. Ci Inf. Brasília 34(2), 81–92 (2005)
[15] Sonza, A.P., Santarosa, L.M.C.: Ambientes virtuais sob a perspectiva de usuários com limitações visuais. Tese de Doutorado em Ciência da Computação – UFRGS – Universidade Federal do Rio Grande do Sul. Porto Alegre – RS. 313 f. Dez (2008)
[16] Tanaka, E.H.: Método baseado em heurísticas para avaliação de acessibilidade em sistemas de informação. Tese de Doutorado em Ciência da Computação - UNICAMP. Campinas – SP. 200 f. Dez (2010)
[17] Torres, E.F., Mazzoni, A.A., Alves, J.B.M.: A acessibilidade à informação no espaço digital -. Ci Inf. Brasília 31, 83–91 (2002)
[18] Torres, E.F., Mazzoni, A.A., Alves, J.B.M.: Conteúdos digitais multimídia: o foco na usabilidade e acessibilidade. Ci. Inf., Brasília 33, 152–160 (2004)
[19] Vilella, R.M.: Conteúdo, usabilidade e funcionalidade: três dimensões para avaliação de portais estaduais de governo eletrônico na web. 2003. 262 f. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal de Minas Gerais, Belo Horizonte (2003)

Exploring the Design Space of Mobile Payment Systems

Rui José[1], Nuno Otero[1,2], Helena Rodrigues[1], Filipe Meneses[1], and Odete Coelho[1]

[1] Algoritmi Center, University of Minho, Portugal
[2] Department of Media Technology, Linnaeus University, Sweden
{rui,nuno.otero,helena,meneses}@dsi.uminho.pt

Abstract. This paper describes an exploratory study on the design space for mobile payments. The study aims at identifying how some features of potential systems may affect the end-user experience and the overall values associated with it. This work takes an interaction design perspective and is not particularly focused on usability or task efficiency issues. Instead, we provide a set of concepts towards the generation of a framework about the Human values involved in mobile payments. The proposed framework and the results from users' interviews provide an added-value for anyone designing mobile payment systems. The resources presented should facilitate the reasoning about the major sensitivities involved in mobile payments as well as initial reflections of their implications.

Keywords: Human-computer interaction, Electronic commerce.

1 Introduction

The design of any digital artifact should always go beyond the mere definition of form and functionality and take into account the ways in which the artifact is going to fit into the larger context of daily life and into the eco-system of already existing services. This requires a thorough understanding of the social milieu that the system is meant to integrate, a clear view of the respective value proposition and the engendered users' experiences [1]. While there is increasing awareness about the ways in which computers and other digital technologies can raise major concerns regarding ethical principles [2], the respective design processes still pay little attention to the multiple social trade-offs involved in the design decisions and particularly to their implications on Human values.

In this research, we study these issues within the scope of mobile payments systems, which we define as payment services performed through the use of a mobile device. Like other digital artifacts, mobile payment technologies should not be designed on the assumption that the frame of reference from previous payment systems can be applied directly to a new technological approach. In traditional payment situations, the technical and social framework in which they are inscribed has evolved and matured over the years and most of the procedures involved are now part of common practices. Mobile payments can have a strongly disruptive effect on many of the practices and respective safeguards people normally resort to when

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1095–1104.
DOI: 10.1007/978-3-642-36981-0_104

making payments with currently existing methods. For example, from a technology perspective, credit cards are known to have multiple risks, but their use is based on a trust model that has evolved over the years to deal with the perception and management of risk by the users and issuers respectively. With mobile payments, values that are now deeply embedded in current payment practices will necessarily be challenged, generating attrition with stakeholders and potentially leading to policy vacuums [2]. If the design is focused only on the mere optimization of technical properties, e.g. efficiency, security and cost, it will miss the major implications that technological options can have on the values of users and society as a whole; for example, by excluding certain social groups, avoiding tax evasion or reducing the turnaround time of money flows.

A thorough approach to the design of new mobile payment systems should thus be grounded on a deep understanding of these implications and deal explicitly with the uncertainty associated with the impact of the new technologies in peoples' values and concerns. In this study, we aim to develop our understanding of these implications by studying the main values associated with the features in different types of payment situations. We have defined a set of 6 reference scenarios, embedded with different properties in regard to key usage, and we have conducted interviews with 30 participants to elicit their perception about the main values involved in each of the scenarios. The result is an initial contribution towards a framework for Human values involved in mobile payment systems that we think may inform the design of new payment technologies by providing a mapping between technological decisions and the respective implications in regard to multiple and possibly conflicting values.

2 Related Work

Kindberg et al. [3] investigated users' trust and security concerns in mobile payments. Their study focused on a possible scenario involving interactions using the concept of an "electronic wallet" to pay for a meal in a simulated restaurant. Their goal was to better understand how to increase users' perceptions of trustworthiness and security. More specifically, their research focused on evaluating the implications on users' trust and sense of security of factors such as: (a) different physical configurations of devices, (b) the utilization of wireless connections and (c) people's previous experiences, or lack of it, in using similar situations. One of their main results showed that ease of use and convenience or social issues may be equally or more important than trust and security issues.

Kristoffersen et al. [4] empirical study investigated users' attitudes towards novel payment solutions, in particular, m-payments in Norway. The authors conclude that m-payment applications, as stand-alone applications, have not been very successful. Users do not seem to particularly value the mobile payment functionality except if it allows for an immediate use of the product. Furthermore, they hypothesise that the payment method and the actual transaction should be tightly coupled with the buying experience. Mallat et al. [5] conducted a study of mobile ticketing service adoption in public transportation. Their study focused on improving traditional adoption models

of mobile technology in ticketing services. Their results suggest that contextual and mobile service-specific features such as budget constraints, availability of other alternatives and time pressure are important determinants of adoption. C. Kim et al. [6] review the literature regarding mobile payment services and analyze the impact of m-payment system characteristics and user-centric factors on m-payment usage across different types of mobile payment users. They propose a m-payment research model to analyze the adoption behaviours of m-payment users and assist managers in implementing appropriate business models and service strategies for different m-payment user groups.

These studies are representative of the type of research that has previously addressed the design of mobile payment systems. They are focused in uncovering the value-chains and qualities that can foster adoption or in understanding the perceptions of trust in regard to different payment scenarios. Our present study, while also considering people's value-chains towards these novel payment methods, is clearly focused on establishing a systematic connection between Human values and common features of mobile payment systems.

The study of the role of Human values in technology design has led to the emergence of frameworks of Human values, such as the Value Sensitive Design framework proposed by Friedman et al. [7][8]–[9]. According to Friedman et al., the framework "...can help researchers uncover the multiplicity of and potential conflicts among human values implicated in technological implementations." (pag. 356). More recently, Nathan et al. [10] utilized the Value Sensitive Design framework in conjunction with inspiring ideas from urban planning and design noir to foster reflection regarding systemic effects (large scale, long term) on persons and society of some digital artifacts appropriations.

3 Methodology

For this study, we have defined six reference scenarios that framed interviews with 30 participants to uncover their perceptions about the values involved in each scenario.

3.1 Scenarios

Scenarios are a valuable and commonly used method in interaction design [12]. They are stories about users and their interactions with artifacts in a specific context that should enable the exploration and discussion of contexts, needs and corresponding requirements. For this study, we devised a set of 6 scenarios that represented different combinations of what we considered to be properties with a particularly significant impact on people's perception about mobile payments, more specifically:

- The presence of a trusted component of the payment infrastructure may strongly influence the perception of trust associated with the payment procedure.
- The presence of a Human mediator may also influence the perception of trust, but it also has implications for the usability of the system.

- The integration of added-value services in the payment process may considerably increase the value for users but it may also increase the process complexity.
- NFC technology may increase the convenience of the payment procedures, but without established practices it may also affect the perception of control.

From the various combinations between these properties, we have selected the 6 scenarios that overall seemed representative of the diversity of possibilities, while corresponding to realistic usage situations. These combinations and the respective scenarios are depicted in Table 1.

Table 1. Scenarios selected from a combination of key properties

Scenario	Infrastructure	Human mediator	Extended services	NFC
1: NFC-enabled device with payment terminal	Yes	Yes	Payment only	Yes
2: Debit card with payment terminal	Yes	Yes	Payment only	No
3: Buying electronic tickets	Yes	Yes	Extended	Yes
4: Vending machine with NFC	Yes	No	Payment only	Yes
5: Internet payment using the mobile phone	No	No	Payment only	No
6: Peer-to-peer using NFC	No	Peer	Payment only	Yes

Interestingly these combinations are very much aligned with relevant payment scenarios also identified in [13]. For each of the selected combinations, we wrote a scenario as a usage story, with a simple photo for further contextualization, but without going into many details. The goal was to create as much situational context as possible without getting too tied to specific technological features. A more detailed description of the scenarios will be made as part of the presentation of the respective results.

3.2 Participants

We then interviewed 30 participants, asking them about their perceptions regarding payment systems in general, and their interpretation of the key issues involved in some of the specified scenarios. We tried to obtain a diverse group of participants with distinct backgrounds, by selecting them at three different locations: a university campus, offices of a local council and a café at the city centre. Although the places chosen were quite heterogeneous in their cultural and social milieu, reflecting a concern with getting opinions and values from a diverse set of people, we do not claim to have a complete cover of the diverse potential user groups of mobile

payments. In relation to demographic data of the sample, although the age range was wide and there are representatives of different age groups, nine participants reported being 19-20 years old. We should also note that twenty-two people reported having completed secondary school, six a university level degree and one basic education (primary school).

3.3 Interviews

The interviews were focused on the elicitation of Human values that participants associated with each of the scenarios. In order to deal with the inherently subjective nature of Human values, while having a solid ground for our methodology, we decided to base our work on the Value Sensitive Design framework. There are mainly two aspects in this framework that were of very relevant for our investigation: the set of values chosen and the general steps of the research process.

In relation to the set of values, this work is based on the values described in Friedman et al. [7], which comprises the following: human welfare, ownership, privacy, freedom of bias, universal usability, trust, autonomy, informed consent, accountability, courtesy, identity, calmness and environmental sustainability. From this list, we decided to exclude "identity" because in the pilot interviews people did not clearly understand the term. We also excluded "privacy", because we wanted to see how participants spontaneously mentioned this particular value.

In relation to the methodological issues, the framework suggests the need to find methods that allow the elicitation of stakeholders' views, assumptions about networks of values and possible trade-offs. This is was the goal of the semi-structured interviews that we designed to explore the perceived values in the reference scenarios.

The interview started with a brief explanation of the study goals and the presentation of a consent form. After this initial introduction the interview proceeded with closed demographic questions regarding age, income and their use of three currently well-established different payment methods: debit card, payshop, a payment method common at local agents and used mainly to pay utility bills, and Internet based purchases using credit cards.

The interview would then focus on description of the scenarios and the formulation of open questions about the values embedded in the payment situations. More specifically, questions were formulated concerning:

- Preferences towards mobile technology mediated payment methods.
- Understanding of the procedures involved in each presented scenario.
- The five most important values from a set of given values regarding the transaction described in the scenario presented.
- Other values that the participants thought of as important given the story told.

When posing the open-ended questions, the interviewer could, at any point, formulate further questions in order to elucidate something said or elicit new ideas and thoughts. The interviews were audio-recorded and transcribed for analysis.

4 Results and Discussion

In relation to the preferred payment method, the majority of the participants chose the debit card (27 out of 30). Regarding the purchases usually done with currently well-established payment methods, participants reported that:

- For debit card payments the range of products and services acquired is quite diverse. In fact, almost all major groups of products and services were considered. People are definitely very used to this method of payment and to its different capabilities.
- In relation to payshop, the six respondents that reported making payments with this method mentioned: paying bills (probably domestic ones) and mobile phones top-ups.
- For Internet based shopping the seven interviewees that said using this method mostly reported acquiring books, music, films, software and travelling goods (flights and accommodation).

The remainder of the results concerns the way in which the various scenarios have elicited different values. Table 2 summarizes the relationship between scenarios and values.

Table 2. Frequencies of the most mentioned values for each scenario

Values	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Total
Universal Usability	4	4	3	3	3	5	22
Ownership and property	2	5	2	2	4	6	21
Trust	5	4	3	3	1	3	19
Accountability	3	4	3	2	2	2	16
Autonomy	1	2	3	3	1	3	13
Human Welfare	3	1	1	1	1	5	12
Informed consent	2	1	2	2	2	3	12
Calmness	2	0	0	1	3	1	7
Freedom from bias	2	2	0	1	0	1	6
Courtesy	2	1	1	0	2	0	6
Environmental sustainability	0	1	2	1	1	1	6

Scenario 1: NFC-Enabled Device with Payment Terminal. Scenario 1 describes a payment mediated by a human and supported by a trusted element that is part of the payment infrastructure (payment terminal). NFC supports the communication between the user and payment infrastructure. The human mediator introduces the payment amount in the payment terminal and selects mobile payment mode. The user approaches his mobile phone to the payment terminal. He receives the payment amount in the mobile phone, confirms and enters the security PIN. He approaches his mobile phone again to the payment terminal and receives a payment confirmation message. After payment transaction, payment infrastructure issues the user receipt.

In this scenario the descriptions of two of the four interviewees clearly mentioned the interventions of the TPA and the mobile phone. However, no details were given regarding the actual mechanisms described. Furthermore, curiously, one considered the payment method an ecological alternative to the classic use of a TPA and the other made a comment regarding its implicit ease of use. Clearly, people experience considerable difficulty in identifying what data elements could be involved in the exchange. In relation to the risks identified, people seemed quite confident that no risks were associated, not even the stealing of the pin number. Maybe the interviewees did not reflect in the possible implications involved outside the particular event described. We mention this possibility since we would expect some sort of reference to the possibility of stealing the pin number. However, a more radical interpretation might suggest that although there are similarities between debit card payments and the scenario presented people did not consider them. Regarding the values deemed more important four are clearly distinguished: universal usability, trust, ownership and accountability. Furthermore, two more values outside the list were mentioned: confidentiality and mobility. However, it should be mentioned that these two terms considered as new values could probably be merged into autonomy and trust. This exemplifies how delicate the whole process of defining and identifying values is. One of the interviewees mentioned privacy as a missing value of our list.

Scenario 2: Debit Card with Payment Terminal. Scenario 2 corresponds to a classical payment using a payment terminal and a debit card. The payment is mediated by a human and is supported by a trusted element that is part of the payment infrastructure (payment terminal). The human mediator introduces the payment amount in the payment terminal and selects debit card payment mode. The user observes the payment amount in the payment terminal, confirms and enters the security PIN. After payment transaction, payment infrastructure issues the user receipt.

In this scenario, people do not seem to have many doubts regarding the nature of the transaction. When asked about what was happening, two of the six respondents actually said: “nothing unusual happened, the customer just paid the bill”. However, the lack of detail of the descriptions raises the question if people are really aware of the underlying procedures involved in the transaction or if they only recognise the needed actions to complete the process. When asked about the data exchanged the most complete answers mentioned the pin number and the amount to be paid. In relation to the possible risks involved in this type of payment: two people considered that no risks were involved, three mentioned the disclosure of the pin number and two interviewees also considered the danger of card cloning. The values that seem to be considered the most important, from the list provided, were: trust (5 times) and universal usability (4 times). Please note that we are considering the 3 or 4 values that were more frequently chosen as the five most important values in the list – this applies for the whole analysis. No other additional values were mentioned when the interviewer asked so.

Scenario 3: Buying Electronic Tickets. In scenario 3, all the transactions in scenario 1 are repeated. However, in this scenario, an extra functionality is provided to the user, which corresponds to the issuing of electronic tickets that are stored in the mobile phone for subsequent validation, as for example in the transports context.

The descriptions of the events presented in the scenario focus on the use of the mobile phone and the explicit reference to ease of use and efficiency of the process: "the user utilized the mobile phone to pay the tickets instead of carrying money in his pocket, he just approached the mobile to the TPA and made the payment in a fast and simple way" or "the user chose a fast and practical payment method". One of the interviewees mentioned that the payment method could have involved some sort of application present in the mobile phone and the Internet. In relation to the data exchanged, the participants were not clear. Only one referred an explicit type of data: the mobile phone number. In general the participants mentioned the "personal data" and "bank account details" without further specification. The entities involved were supposedly the buyer, banks and the vendor. It seems curious because no one mentioned possible mediating entities. People seem to believe that the transactions are done seamlessly between these three agents (buyer, seller, and the banks). As we can see from the other scenarios this situation is similar for the other situations described. The participants considered that the risks involved were: communication failure between the mobile phone and the machines responsible for the identification of the purchase. Only one participant mentioned the possible theft of personal data. Regarding the values chosen as most important, the interviewees pointed out: universal usability, trust, autonomy and accountability. It should be mentioned that the frequencies of the values chosen is more spread in this scenario, all the 4 values being referred to were picked three times. Comparing with the previous scenario the difference resides in the inclusion of autonomy (chosen 3 times) and exclusion of ownership (chosen only two times). One of the interviewees mentioned privacy as a missing value of our list.

Scenario 4: Vending Machine with NFC. Scenario 4 differs from scenario 1 in that it does not assume the existence of a human mediator. The user interacts with a self-checkout vending machine. The user chooses the intended good in the vending machine that shows the payment amount. The user approaches his mobile phone to a signaled area in the vending machine and receives the payment confirmation. There is no need to enter the security PIN as it refers to a micropayment.

In this scenario the interviewees described the events in a very shallow way. In fact, the statements point to the novelty of the situation instead of the details of the transaction and event. This points out, once more, the difficulties of explaining new technologies to end-users when utilizing non-detailed scenarios and not explicitly probing for answers. Although no great detail is found when asked about data exchange, three of the interviewees explicitly mentioned the current account balance of the buyer, something not referred to in the previous scenarios. In relation to the potential risks involved, one of the interviewees refers to privacy, while another mentions broken transaction. In what concerns the values chosen, universal usability, trust and autonomy were the preferred ones. However, three of the interviewees explicitly mentioned privacy when asked to consider additional values.

Scenario 5: Internet Payment Using the Mobile Phone. Scenario 5 corresponds to a classical Internet payment with no human mediator, infrastructure element or NFC. The user issues the products through the Internet and later receives a message in his mobile phone with payment instructions. The user enters the pin and processes the payment. He later receives the payment receipt by email.

Only one of the interviewees described the process in some detail: "The buyer purchased a product using the Internet and providing personal data like the pin number. He also received the proof of payment". Two of the four interviewees mentioned the security concerns without explicitly stating why and how. In relation to the identification of the data exchanged, once again the responses are quite vague: pin number, personal data (without stating what), and bank account details. The risks involved in this transaction and reported reflect the concerns regarding security and seem to be closely associated with the explicit mentioning of the Internet. In what concerns the values chosen, the preferences were: universal usability, autonomy and trust. One interviewee mentioned mobility as one value to be added and privacy was not mentioned by any of the participants.

Scenario 6: Peer-to-Peer Using NFC. Scenario 6 corresponds to a payment with a peer device. There is more than a person involved, but neither of them has a special role in the mediation. The user issuing the payment selects the payment option. He approaches his mobile phone to the payee's phone. In this NFC interaction he receives and confirms the payment value. In the second NFC interaction, payment is issued and the user receives the receipt.

In this case the interviewees seemed a bit clueless about the whole process. Two of them explicitly stated "I never heard of this, I do not know" while the three other vaguely mentioned "a combination of services", "the Internet was used" and "the payment was effectuated via mobile phone". Clearly, people did not know what to say and it seems no analogies emerged. However, when asked about the data exchanges, although two of the participants did not provide details, two gave quite elaborate answers: "mobile networks, account details of both agents, and the banks were involved. The data of both agents were used to login to their bank accounts using the mobile phones and corresponding networks" and "the exchange was done using the bank account identification number, card codes, the networks and the banks". In relation to the risks perceived, four of the participants chose not to answer while one enumerated risks associated with Internet based payments. The values more frequently chosen were: ownership and property, universal usability and, surprisingly human welfare. However, two of the interviewees added safety, while one mentioned ease of use and another mentioned honesty.

5 Conclusions

The results of our study have shown, first of all, how the different properties in our payment scenarios can considerably influence the perception that people have about the key values involved. Universal usability, ownership and property, trust and accountability were the values more frequently chosen. However, they were not distributed equally across all the scenarios: ownership and property was not particularly relevant for scenario 1, 3 and 4; trust was the most chosen value for

scenario 5 and 6 (although safety and honesty were specifically mentioned). Privacy was not mentioned very often, although people may conflate this particular value with trust and did not give it its due relevance. Interestingly, in the peer-to-peer transaction of scenario 6, in which there is not trusted component from the infrastructure, trust and honesty were the most strongly elicited values, suggesting that people would probably only consider this type of transactions with people they trust.

These connections between the most elicited values and the properties of the scenarios may be seen as starting point towards a more comprehensive mapping between features of mobile payment systems and their implications on Human values. Further research is needed to clarify how to increasingly improve the alignment between technological decisions in the design of mobile payment systems and the sensitivity to the values involved.

Acknowledgments. This work was co-funded by "Agência de Inovação" and the national QREN program through the COMPETE program, under the project MOBIPAG - National Initiative for Mobile Payments (project 13847).

References

1. Sellen, A., Rogers, Y., Harper, R., Rodden, T.: 'Reflecting human values in the digital age', Commun. ACM 52(3), 58–66 (2009)
2. Johnson, D.G.: Computer Ethics. In: Floridi, L. (ed.) Philosophy of Computing and Information, pp. 65–75. Blackwell Publishing (2004)
3. Kindberg, T., Sellen, A., Geelhoed, E.: Security and Trust in Mobile Interactions: A Study of Users' Perceptions and Reasoning. In: Mynatt, E.D., Siio, I. (eds.) UbiComp 2004. LNCS, vol. 3205, pp. 196–213. Springer, Heidelberg (2004)
4. Kristoffersen, S., Synstad, A., Sorli, K.: Users' perception of mobile payment. International Journal of Knowledge Management Studies 2(1), 74–95 (2008)
5. Mallat, N., Rossi, M., Tuunainen, V.K., Öörni, A.: An empirical investigation of mobile ticketing service adoption in public transportation. Personal Ubiquitous Comput. 12(1), 57–65 (2008)
6. Kim, C., Mirusmonov, M., Lee, I.: An empirical examination of factors influencing the intention to use mobile payment. Computers in Human Behavior 26(3), 310–322 (2010)
7. Friedman, B.: Value-sensitive design. Interactions 3(6), 16–23 (1996)
8. Friedman, B., Peter, J., Kahn, H.: Human values, ethics, and design. In: Jacko, J.A., Sears, A. (eds.) The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, pp. 1177–1201. L. Erlbaum (2003)
9. Friedman, B., Kahn, P.H., Hagman, J., Severson, R.L., Gill, B.: The Watcher and the Watched: Social Judgments About Privacy in a Public Place. Human–Computer Interaction 21(2), 235 (2006)
10. Nathan, L.P., Friedman, B., Klasnja, P., Kane, S.K., Miller, J.K.: Envisioning systemic effects on persons and society throughout interactive system design. In: Proc. of the 7th ACM Conf. on Designing Interactive Systems, Cape Town, South Africa, pp. 1–10 (2008)
11. Carroll, J.: Five reasons for scenario-based design. Interacting with Computers 13, 43–60 (2000)
12. Kreyer, N., Pousttchi, K., Turowski, K.: Mobile Payment Procedures: Scope and Characteristics. e-Service Journal 2(3), 7–22 (2003)

SketchyDynamics: A Sketch-Based Library for the Development of Physics Simulation Applications

Abílio Costa[1] and João Paulo Pereira[1,2]

[1] Computer Science Department, School of Engineering (ISEP), Polytechnic of Porto, R. Dr. António Bernardino de Almeida 431, Porto, Portugal
[2] Knowledge Engineering and Decision Support Group (GECAD), School of Engineering, Polytechnic of Porto, R. Dr. António Bernardino de Almeida 431, Porto, Portugal
amfcalt@gmail.com, jjp@isep.ipp.pt

Abstract. Sketch-based interfaces provide a powerful, natural and intuitive way for users to interact with an application. By combining a sketch-based interface with a physically simulated environment, an application offers the means for users to rapidly sketch a set of objects, like if they are doing it on piece of paper, and see how these objects behave in a simulation. In this paper we present SketchyDynamics, a library that intends to facilitate the creation of applications by rapidly providing them a sketch-based interface and physics simulation capabilities. SketchyDynamics was designed to be versatile and customizable but also simple. In fact, a simple application where the user draws objects and they are immediately simulated, colliding with each other and reacting to the specified forces, can be created with only 3 lines of code.

Keywords: Sketch-Based Interfaces, Gesture Recognition, Physics Simulation, Rigid Body Dynamics, Education, Entertainment.

1 Introduction

Using pen and paper to draw or sketch something in order to express an idea is very common and also very natural for us. By using this concept in user interfaces one can make the interaction process more natural and spontaneous.

In this paper we propose SketchyDynamics, a programing library to aid in the creation of applications for 2D physics simulations in which the user interacts directly with the scene using a "pen and paper" style interaction. Thus, instead of selecting from a menu which objects compose the scene to be simulated, the user can simply draw them directly into the scene. We hope that developing this library will provide a boost for developers to create new applications around this concept, be they for educational purposes, like an application used to teach physics with an interactive whiteboard, or for entertainment purposes, such as a physics-based game where the user draws parts of the scene in order to reach a goal.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1105–1116.
DOI: 10.1007/978-3-642-36981-0_105

The library supports three gestures to draw rigid bodies and other three to define connections between them. The first three gestures are used to produce rectangles, triangles and circles, which can be created by drawing these symbols directly. Also, the user can draw a zigzag to connect two bodies with a spring, an alpha to pin a body over another and a small circle to define a rotation axis between two bodies. Since both the circle body and the rotation axis relation use the same gesture, we only have in fact five gestures to recognize, presented in Figure 1.

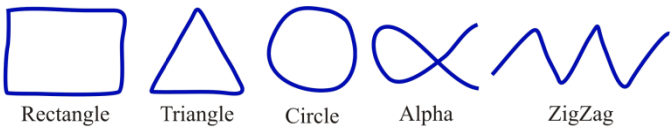


Fig. 1. Set of gestures used in our library

Although there are already several applications that combine physics simulation with a sketch-based interface, most of them have a specific scope and audience. As a library, SketchyDynamics is intended to be used in different types of applications and does not have a definite scope. We hope that our work helps developers create new and exciting applications with little effort in combining the physics simulation with the sketch-based interface.

In the next section we present an overview of the results achieved in the sketch recognition field and also works that combine sketch-based interfaces with rigid body physics simulation. Section 3 gives a little insight into a previous evaluation whose purpose was to select the sketch recognizer that best integrates with our library. In section 4 we present our library, its technical characteristics, along with its functionality. Section 5 discusses a preliminary informal evaluation and section 6 concludes this paper and presents potential future work.

2 Related Work

This section presents some of the related work in the sketch-based interfaces domain and is divided into two subsections. The first subsection will address the work done in the sketch recognition field, while the second presents some examples of applications that result from the combination of sketch-based interfaces with rigid body physics simulation.

2.1 Sketch Recognizers

Given the potential of automatic sketch recognition, a lot of work has been done in order to develop recognizers capable of dealing with the intrinsic ambiguity of hand-drawn sketches. Since there is a wide variety of sketch recognition algorithms, it is only natural that there's also diversity in their characteristics. Examples of these characteristics are the ability to be trained to recognize new gestures, the capacity to

recognize multi-stroke gestures or the sensitivity to the gesture's orientation, scale or drawing direction.

Rubine's recognizer [1], a trainable gesture recognizer, classifies each gesture using a linear classifier algorithm with a set of distinct features. The recognizer is very flexible since features can be easily added or removed to make the recognizer fit the application needs, as proven by Plimmer and Freeman [2]. The major limitations of Rubine's recognizer are its sensitivity to the drawing direction, scale, and orientation and inability to identify multi-stroke sketches. Pereira *et al.* [3] made some modifications to Rubine's recognizer in order to make the algorithm accept multi-stroke sketches, but only when drawn with a constant set of strokes, as pointed out by Stahovich [4]. Pereira *et al.* also present a way to make the algorithm insensitive to drawing direction.

CALI [5] is an easy to use multi-stroke recognizer that uses Fuzzy Logic and geometric features to classify gestures independently of their size or orientation. CALI divides gestures into two types: shapes and commands. Shapes can be drawn (and recognized) using solid, dashed and bold lines, while commands are only recognized with solid lines. Since CALI is not trainable, adding new gestures is not an easy task, involving analysis of which features characterize and distinguish the new gesture and hand-coding these features. To solve this limitation the authors also present a trainable recognizer but it has a lower recognition rate and requires numerous training templates for each gesture class[1].

Wobbrock *et al.* [6] present the $1 Recognizer which aims to be easy to understand and quick to implement. It is insensitive to scale and orientation of sketches, but sensitive to their drawing direction. One major advantage of $1 Recognizer is the simplicity to add support for new gestures, requiring only one training template per gesture class to be effective. Furthermore, the authors also explain how to make the recognizer sensitive to scale or orientation, for some or all gesture templates.

In order to solve some of the limitations of the $1 Recognizer, such as not being able to recognizing multi-stroke gestures, sensitivity to the drawing direction, and problems recognizing uni-dimensional gestures such as lines, Anthony & Wobbrock extended it and created the $N Recognizer [7]. Despite the improvements over the $1 Recognizer, $N has problems recognizing gestures made with more strokes than those used in the training templates. Also, it is not well suited to recognize "messy" gestures like a scratch-out, commonly used for erasing-like actions.

Lee *et al.* [8] present a trainable graph-based recognizer that is insensitive to orientation, scale and drawing direction and is able to recognize multi-stroke gestures. Since the recognizer uses statistical models to define symbols, it handles the small variations associated with hand-drawn gestures very well. Despite being a trainable recognizer, it requires all training templates of a gesture class to be drawn with a consistent drawing order or consistent orientation.

Vatavu *et al.* [9] present a trainable recognizer that uses elastic deformation energies to classify single-stroke gestures. The recognizer is naturally insensitive to

[1] A gesture class represents a unique gesture, but can be made from multiple representations of that gesture, i.e. multiple templates.

gesture scale and orientation, since the same gesture has similar curvature functions independently of the drawing orientation or size, but is sensitive to drawing direction and starting point within the gesture.

Sezgin and Davis [10] present a multi-stroke sketch recognizer, based on Hidden Markov Models (HMM), that is capable of recognizing individual sketches in complex scenes even if the scene is not yet completed, i.e. while it is being drawn, and without the need to pre-segment it. On the other hand it can only recognize sketches in their trained orientations, thus being sensitive to orientation. Since the recognition relies on the stroke order of the trained templates, it is not well suited for domains where the stroke ordering cannot be predicted. Also, because HMMs are suited for sequences, it cannot recognize single-stroke sketches, unless they are pre-segmented.

2.2 Physics Simulation with Sketch-Based interfaces

The idea of using a sketch-based interface to create and manipulate a simulated scene is not something new. For example, ASSIST [11] is able to recognize sketches and convert them to mechanical objects which can then be simulated. The system recognizes circles and straight-line polygons (simple or complex) made of single or multiple strokes. The recognition is done incrementally, while the user is drawing, which makes the system feel quicker and also gives an instantaneous feedback to the user, since hand-drawn lines are converted to straight lines and colored according to the type of object recognized. When an improper interpretation of a gesture is made, the user is able to correct it using a list of alternative interpretations. In ASSIST, users can also pin one object over another with a rotational axis by drawing a small circle, or anchor objects to the background by drawing a small cross. After finishing the sketch, the user can press a "Run" button to transfer his design to a 2D mechanical simulator that runs and displays a simulation of the designed scene.

Another application, "Free-Hand Sketch Recognition for Visualizing Interactive Physics" [12] enables users to draw simple 2D objects and simulate how these objects behave in 3D. The application is able to recognize four types of objects: lines, circles, rectangles, and triangles. When the gesture cannot be recognized a small dialog is presented, requesting the user to specify the desired gesture. After creating an object, the user is able to anchor it so that it remains static during the simulation. The design process consists of three modes: the "Ink" mode where the user can draw new objects; the "Select" mode, where a circle selects the enclosed objects; and the "Erase" mode, used to remove objects. Despite the designing being done in 2D, the physics simulation is 3D and the user is able to move the camera and also move objects in 3D space.

There are also games that take advantage of a sketch-based interface and a physics simulated environment to entertain the player. One popular example is Crayon Physics Deluxe [13], a puzzle game where the main objective is to guide a ball so that it touches all the stars in each level. Instead of controlling the ball directly, the user needs to draw objects that influence the ball, leading it to the stars. The user can draw rigid bodies with any shape and connect them with pivot points and ropes. Since the

simulation is always running, sketched objects are simulated and interact with other objects right after being drawn. The game has a "children's drawing" theme, with a background that resembles a yellow paper sheet and crayon-like sketches, both characteristics that make it successfully adopt the pen-paper paradigm. Crayon Physics Deluxe also includes a level editor and an online playground, so users can create their own levels and submit them online.

3 Sketch-Based Recognition Evaluation

Due to the high importance of having good gesture recognition, since the user must feel the interaction to be as natural and unrestrictive as drawing with a pen on a paper, the gesture recognizer used in SketchyDynamics was selected based on previous evaluation [14] [15]. The evaluation was conducted using real gesture samples drawn by 32 subjects, with a gesture set specifically arranged for our library (Figure 1).

For the evaluation process we developed an application to collect gesture samples from the subjects, process them, and compute the recognition results. With this tool we evaluated Rubine's recognizer, CALI and the 1$ Recognizer, concluding that for our gesture set CALI achieved the highest recognition rates.

With this evaluation we were also able to improve recognition rates by tweaking the templates and the recognizer's implementation to our specific gesture set.

4 The SketchyDynamics Library

SketchyDynamics is a programing library that aims to simplify the implementation of 2D physics simulation applications with sketch-based interfaces. Using 2D graphics and physics simulation means that the user sketch (in 2D) produces a 2D object, which resembles the pen-paper paradigm and simplifies user interaction.

Out of the box, SketchyDynamics provides an interface for the user to interact with an application along with recognition and processing of user actions such as drawing, moving, scaling and removing rigid bodies and their joints. SketchyDynamics also deals with the physics simulation of these elements and visually represent them on the computer screen along with other user interface elements. Thus, a developer can integrate these features in an application with almost no effort.

4.1 Architecture

A major concern when designing SketchyDynamics was to make it versatile, so that developers can create all kind of applications, but at the same time simple enough to enable rapid prototyping. For example, with only 3 lines of source code a developer can create a simple test application where the user can draw objects and see their simulation, while they collide with each other and react to the specified "gravitational force". With a dozen more lines the developer is able to add a background body

where the user is able to attach objects, or a ground body so that drawn bodies have something to fall onto.

As stated previously, we use CALI as the gesture recognizer since it yielded the best results in our evaluations.

For the physics simulation SketchyDynamics uses the Box2D physics engine. Despite using Box2D, SketchyDynamics does not encapsulate it or hide it from the programmer. Instead programmers have access to all Box2D objects and functionality so they are able to parameterize them according to the application's needs.

Although bodies and joints are created automatically by the library when the user draws them, the application is also able to programmatically create and remove them (along with their visual representations). Furthermore, SketchyDynamics also gives the application full control over the simulation state.

To render the bodies simulated by Box2D and any other visual elements we used the OpenGL API. Despite that, SketchyDynamics was designed so that a developer can easily use another API. This is achieved by ensuring that all OpenGL-specific code is encapsulated in a few classes, thus creating a conceptual abstraction layer.

While implementing the OpenGL abstraction we took the opportunity to add some "graphics library" functionality. For example, a programmer can easily create polygons by defining their vertices and then apply geometric transformations to them, toggle their visibility on screen, among other operations, all done in an object-oriented manner. Additionally, the library provides scene query functionality and easy texture management for the developer. To render each object SketchyDynamics offers three rendering queue layers so that each individual object can be drawn on the background, on the front (as a user interface element) or in the middle of these two layers. Furthermore, the depth or order of each object inside each layer can also be specified.

Another design decision that resulted from the OpenGL abstraction was the incorporation of the window creation process inside SketchyDynamics, thus reducing the effort on the developer's side. Moreover, SketchyDynamics delivers events received by the window, like mouse and keyboard inputs, to the application using the observer pattern, thus letting the developer take actions based on the user input.

4.2 User Interaction

In order to best represent the pen-paper paradigm, the user interaction was designed to take advantage of systems with a touchscreen and stylus. Thus, the user only needs to press and move the stylus to interact with the system, without needing extra buttons[2]. Furthermore, no menus are used and most of the interaction is done by sliding the stylus across the screen. Although it was designed with that type of devices in mind, SketchyDynamics also works well with a traditional computer mouse.

There are two types of objects the user is able to create: bodies and joints. Bodies are rigid objects that are simulated according to physics laws while joints are used to connect bodies. Figure 2 shows various bodies and three types of joints.

[2] In a traditional mouse system this means that only the left mouse button is needed.

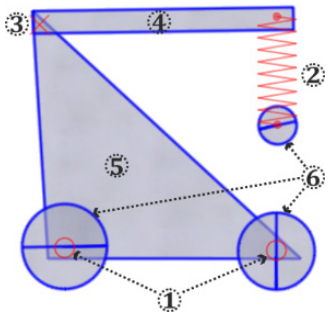


Fig. 2. Various types of joints and bodies: 1) revolute joints; 2) spring joint; 3) weld joint; 4) rectangular body; 5) triangular body; 6) circular bodies

It is also important for the user to be able to manipulate the objects to a certain degree so SketchyDynamics lets the user change an object's position, scale, and orientation, or even delete it.

4.2.1 Creating

The creation of an object, be it a body or a joint, is done by drawing it. So, for example, if users want to create a rectangle body, they simply draw the rectangle on the screen. SketchyDynamics then recognizes the rectangle and its properties, like size and orientation, and creates the physical and visual representations of it.

SketchyDynamics supports four types of bodies: rectangles, triangles, circles and freeform bodies. When the user input is recognized as a rectangle, triangle or circle, it is represented in a beautified manner, as illustrated in Figure 3. Otherwise, when the input is not recognized, it is interpreted as a freeform and represented in a simplified manner (with fewer vertices) for performance reasons.

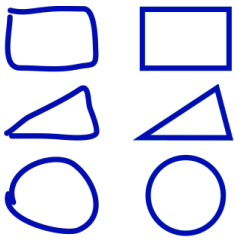


Fig. 3. Example of drawn shapes (left) and respective beautified representations (right)

The user can also connect two bodies with three different joint types: weld, revolute and spring. Weld joints connect two bodies at a specific anchor point, preventing any relative movement between them. Like weld joints, a revolute joint connects two overlapping bodies at a specific point but allows the bodies to rotate freely around that point. Spring joints try to keep a constant distance between two connected bodies, based on the distance at the time the joint was created, stretching and shrinking like a real spring.

Just like creating bodies, the creation of joints is done by drawing them. Drawing an alpha gesture over two bodies connects them with a weld joint with an anchor at the gesture's intersection, while drawing a small circle creates a revolute joint anchored at the circle's center. To create a spring joint, the user draws a zigzag gesture starting in one body and ending in another one, defining the two spring's anchor points as the start and end points of the gesture.

Regarding the visual representation of joints, the weld and revolute joints are represented by a small cross and by a small circle, respectively, on the joint anchor point while the spring joint is displayed as a zigzag line starting in one anchor point and ending on the other, stretching and shrinking subject to the distance between the bodies. The object presented in Figure 2 was constructed using joints of the three types.

In order to better deal with the ambiguity in hand-drawn gestures, a guesses list is presented whenever the user executes a gesture. The guesses list shows all the available objects so that the user can choose an object other than the recognized one. The objects corresponding to gestures identified as matching by CALI recognizer appear bigger and first in the list, since they are the most probable choices, followed by the remaining objects. The guesses list feature can be disabled by the developer, in which case the most probable object is always selected.

Depending on the application-specific setup passed to SketchyDynamics, objects can be created while the physics simulation is in a paused state or while it is running and thus making other objects react instantly to the new object. This instantaneous simulation mode is useful for applications where the user interacts with a live environment as usually happen in games.

4.2.2 Selecting

For an object to be manually manipulated by the user, it first needs to be selected. When any object is selected the physics simulation is paused so that the user can easily edit it without being interrupted by other moving bodies. If the simulation was running before the selection of an object, it will resume after all objects are unselected.

Objects are selected by tapping on them with the stylus (or left-clicking them with a mouse), and can be deselected with the same action. This makes selecting multiple objects an intuitive process since users only need to keep tapping on the objects they want to select. It is also possible to unselect individual objects when there are multiple objects selected. When an object is selected, its lines assume a distinctive color, returning to the original color after being unselected. As shown in Figure 4, this gives a clear feedback regarding the object's state. Also, tapping on an area of the screen with no objects or on an object configured as non-selectable, deselects all selected objects. Non-selectable objects are useful to create the application's scenery, which the user cannot manipulate but may be able to interact with, for example by connecting a user-made body to a scenery object.

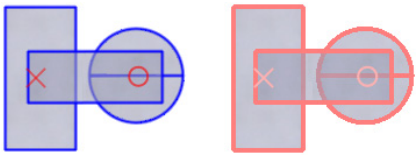


Fig. 4. Set of objects in unselected (left) and selected (right) states

When there are multiple bodies connected by joints and one of them is selected, all the other connected bodies are automatically selected, as long as they are selectable objects. This feature was introduced in order to improve the usability of the system, since we found that when multiple bodies are connected the user typically wants to manipulate them as a whole.

4.2.3 Moving

A selected body or joint can be moved by pressing over it and dragging the stylus. The object will move in sync with the stylus as long as the user keeps it pressed on the screen.

When there are multiple objects selected they all move in a synchronized manner, regardless of which object was pressed by the stylus.

4.2.4 Scaling and Rotating

Scaling and rotation of bodies is done simultaneously in a single action. As the action to move an object, scaling and rotation is done by pressing and dragging the stylus, but instead of pressing inside the selected body, the user needs to press outside it. As the user drags the stylus, the selected bodies scale and rotate based on the stylus initial and current positions. Only bodies can be rotated or scaled, so this operation is not applicable to joints.

The scale factor is calculated based on the current distance from the stylus position to the body center and the initial distance (before dragging the stylus). Regarding rotation, it is done based on the angle between two imaginary lines: the line from the current stylus position to the body's center, and the initial line (before dragging the stylus). Thus, moving the stylus closer or farther from the body scales it while moving the stylus around the body rotates it.

When multiple bodies are selected, they are all subject to the same rotation and scaling factor, but instead of using the body's center point as the reference point, the geometric average of all individual center points is used.

4.2.5 Removing

Since removing objects is an important operation that contributes to user's creative freedom, it was designed to be simple, intuitive, and to have a low impact on the user's cognitive load. In fact, removing an object is a just special case of moving it.

When an object starts being moved by the user, a large rectangle with a trash bin icon slides down from the top of the screen, sliding back up and off-screen when the object cease to be moved. If the stylus enters the trash bin area while moving any object, the trash bin icon turns red. If the user lifts the stylus while on this rectangle,

all the selected objects are removed. Figure 5 shows the trash bin area in context of a simple, almost empty, application, and also the trash bin icon representations before and after the stylus drags an object onto it. We choose to keep this area hidden unless the user starts moving objects to improve the use of screen real estate, since objects can only be deleted when they are being moved by the user.

Joints can also be removed by simply being moved outside any of the two bodies they connect, without the need to move them to the trash bin rectangular area, although the trash bin works for joints too.

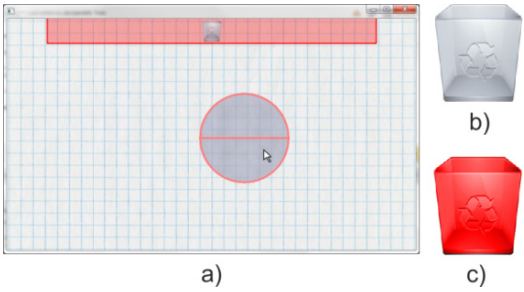


Fig. 5. a) simple application showing the trash bin area in context; b) trash bin icon in its normal state; c) trash bin icon when an object is dragged inside the trash area

5 Preliminary Evaluation

On a preliminary informal evaluation, we noticed that SketchyDynamics is easier to use in devices with touchscreen and stylus, in comparison to traditional mice, which comes with no surprise since these devices are more suited for drawing and sketching.

Regarding the interaction with the system, creating objects is an intuitive action because it resembles the use of a pen and a paper. Also, the sliding trash bin area makes it easy to figure how to remove an object.

The automatic selection of connected objects when the user selects an object made the interaction faster and more convenient. However, it would also be useful to be able to select only one object from a set of connected objects, and scale or rotate it around the joint anchor point, instead of the object's center.

Another aspect worth mentioning is that while having the rotation and scale operations done simultaneously reduces the user's cognitive load and can improve the interaction speed in some situations, doing only one of these operations without triggering the other is complicated.

A more thorough usability evaluation of SketchyDynamics' functionalities was also conducted [15], with a group of 8 subjects (students, teachers and researchers). Using a prototype application developed with SketchyDynamics, each subject was asked to create a complex scene, consisting of 17 bodies (9 rectangles, 3 triangles, 3 circles, and 2 freeforms), and 11 joints (6 weld joints, 3 revolute joints, and 2 springs)[3].

[3] A video demonstrating the creation of such scene can be found at
http://youtu.be/1niigTt_m_I

To construct that scene, subjects took an average of 9 minutes and 12 seconds, with a standard deviation of 3 minutes and 14 seconds, which are very encouraging results. After creating that scene, the subjects filled out a survey form to expose their perception of SketchyDynamics. All major features of SketchyDynamics were assessed in the survey with scores above 1 value, in a scale from -3 (awful) to +3 (excellent), showing that SketchyDynamics pleased the users and is on the right track. Furthermore, the questions regarding the overall perception of SketchyDynamics showed that subjects feel that it is easy to use and is appropriate for creating physically simulated scenes, with both questions receiving answers with average scores of at least 2 values (also from -3 to +3).

6 Conclusions

We have presented a library capable of speeding up the development of applications by providing developers a sketch-based interface combined with physics simulation. The library also provides facilities in managing the graphical side of the application and dealing with user input.

In an effort to make the library suitable for the widest range of applications we are working on adding more functionality into it, such as a new rope-like joint.

During a preliminary evaluation we noticed that it would be useful to be able to select an individual body from a set of connected bodies and transform it using the joint anchor point as a reference. This poses some design problems since an object can have multiple joints (which one would be used?). The problem further increases if there is more than one selected object. Before implementation, further study on how to overcome these problems is needed.

Another interesting feature to have would be the existence of object hierarchies, in which transformations applied to one object are replicated onto its child objects, but not the opposite. This hierarchy could be based on the depth of the objects.

Regarding the difficulty of doing a scale or rotation operation individually, a possible solution would be the use of a modifier key to restrict the action to a single operation. Every time this key is pressed, the system could check if the mouse movement was mainly radial or tangential, doing only a scale or rotation operation, respectively. This concept could also be applied to restrict the movement of objects to horizontal, vertical and 45 degree translations.

References

1. Rubine, D.: Specifying Gestures by Example. SIGGRAPH Computer Graphics 25(4), 329–337 (1991)
2. Plimmer, B., Freeman, I.: A toolkit approach to sketched diagram recognition. In: Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI.but not as We Know It (BCS-HCI 2007), vol. 1, pp. 205–213 (2007)
3. Pereira, J., Branco, V., Jorge, J., Silva, N., Cardoso, T., Ferreira, F.: Cascading recognizers for ambiguous calligraphic interaction. In: Eurographics Workshop on Sketch-Based Interfaces and Modeling (2004)

4. Stahovich, T.: Pen-based Interfaces for Engineering and Education. In: Sketch-based Interfaces and Modeling, pp. 119–152 (2011)
5. Fonseca, M., Pimentel, C., Jorge, J.: CALI: An online scribble recognizer for calligraphic interfaces. In: AAAI Spring Symposium on Sketch Understanding, pp. 51–58 (2002)
6. Wobbrock, J., Wilson, A., Li, Y.: Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In: 20th Annual ACM Symposium on User Interface Software and Technology (UIST 2007), pp. 159–168 (2007)
7. Anthony, L., Wobbrock, J.: A lightweight multistroke recognizer for user interface prototypes. In: Graphics Interface 2010 (GI 2010), pp. 245–252 (2010)
8. Lee, W., Kara, L., Stahovich, T.: An efficient graph-based recognizer for hand-drawn symbols. Computers & Graphics 31, 554–567 (2007)
9. Vatavu, R.-D., Grisoni, L., Pentiuc, S.-G.: Gesture Recognition Based on Elastic Deformation Energies. Gesture-Based Human-Computer Interaction and Simulation 5085, 1–12 (2009)
10. Sezgin, T., Davis, R.: HMM-based efficient sketch recognition. In: 10th International Conference on Intelligent User Interfaces (IUI 2005), pp. 281–283 (2005)
11. Alvarado, C., Davis, R.: Resolving Ambiguities to Create a Natural Computer-Based Sketching. In: Proceedings of IJCAI 2001, pp. 1365–1371 (2001)
12. Kamel, H., Shonoda, M., Refeet, M., Nabil, R.: Free-Hand Sketch Recognition For Visualizing Interactive Physics, `http://code.google.com/p/sketch-recognition-simulation-tool` (accessed 2012)
13. Purho, P.: Crayon Physics Deluxe, `http://crayonphysics.com` (accessed 2012)
14. Costa, A., Pereira, J.: SketchTester: Analysis and Evaluation of Calligraphic Gesture Recognizers. 20º Encontro Português de Computação Gráfica (20ºEPCG) (2012)
15. Costa, A.: SketchyDynamics: Apoio à Produção de Sistemas Baseados em Interfaces Caligráficas para a Simulação da Dinâmica de Corpos Rígidos (M.S. thesis). School of Engineering (ISEP), Polytechnic of Porto, Portugal (2012)

Designing User Learning Experience in Virtual Worlds: The Young Europeans for Democracy Serious Application

Gonçalo Cruz[1,2], Ana Maia[1,3], Leonel Morgado[1,2], Benjamim Fonseca[1,2], Hugo Paredes[1,2], Fernando Bessa[1,4], Clara Rodrigues[5], and Paulo Martins[1,2]

[1] University of Trás-os-Montes e Alto Douro,
Apartado 1013, Quinta de Prados,
5001-801 Vila Real, Portugal
{goncaloc,margaridam,leonelm,benjaf,
hparedes,fbessa,pmartins}@utad.pt
[2] INESC TEC - Associate Laboratory coordinated by INESC Porto
(Institute for Systems and Computer Engineering of Porto), Campus da FEUP,
Rua Dr. Roberto Frias, 378,
4200-465 Porto, Portugal
[3] CIDTFF - Research Centre Didatics and Technology in Education of Trainers,
University of Aveiro,
Campus Universitário,
3810-193 Aveiro, Portugal
[4] CETRAD - Centre for Transdisciplinary Development Studies,
Rua Dr. Manuel Cardona,
5000-558 Vila Real, Portugal
[5] Future Balloons Unipessoal LDA,
Avenida 5 de Janeiro 12/2C, Cova,
3090-705 Figueira da Foz, Portugal
clara.rodrigues@gmail.com

Abstract. In this paper, we present the user experience design of Young Europeans for Democracy (YED) serious application. Facing the actually need to support young citizens understanding and participation in Europe, schools have a key role on European studies integration. Accordingly to the growing opportunities of applying technology for training and education, we propose a serious application in a open-source 3D Virtual World environment solution, adopting a problem-based learning approach, role-playing dynamics and group work tasks. The main goal was to design a meaningfully user experience, increasing the students motivation and engagement for learning different knowledge on the topics of European Studies. Some preliminary results are presented, based on participatory observation within the use of the serious application.

Keywords: European Studies, Virtual Worlds, Open Simulator, User learning experience, Human-Computer Interaction.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1117–1128.
DOI: 10.1007/978-3-642-36981-0_106

1 Introduction

Schools are organizations entrusted with education of our children and youths, including their citizenship, particularly as European union citizens. Teachers are the connectors between Society – its organizational, historical, social, cultural, legal and economic dimensions – and students as users of those references. They have to promote awareness on how European politics, democracy, free market or free mobility, intercultural dialogue, the Euro, and other innovations changed the meaning to be an European citizen. This is a hard job because of its nature, very intangible, and also due the discomfort of teachers with European studies, or even the lack of motivation by students.

We contend that the pedagogical approaches to develop awareness and understanding of the values involved in European citizenship have to go beyond the lecture format and should take a different kind of approach: students need opportunities to perceive those values in different learning contexts and ways, through problem-based learning, role-playing dynamics and group work tasks. Assuming this, we present the ongoing project Young Europeans for Democracy (YED), which intends to provide schools and teachers with innovative resources and know-how related to new forms of teaching about European Union. Promoting youths as active agents of their own learning, and giving them the opportunity to experience their role as European citizens in an immersive 3D learning environment, we have developed a serious application in the Open Simulator 3D Virtual World.

The Section 2 presents related research work on European Studies integration, 3D Virtual Worlds for serious applications and User learning experience in 3D Virtual Worlds, all of which lie at the foundation of our study. Afterwards, in Section 3, we describe the main phases of the YED project, focusing on the user learning experience design within the development of the serious application. And finally, in Section 4, we identify some preliminary results and establish the research agenda for future improvements of the work.

2 Background

2.1 European Studies Integration

Accordingly to the European Commission, a past study revealed the pessimism of European Union citizens about the situation in their country and the future as causes indicating the rapidity of changes taking place, namely: the large uncertainties about the consequences, whether economic or technological developments; the fear regarding the use and effects of new technologies; the risks produced by pollution and genetic manipulation; the feeling of loss of values and social cohesion - especially the growth of poverty, loss of importance of forms of solidarity and friendliness, isolation of young, growing insecurity and fear of the effects of immigration; and the pernicious results of economic liberalism - whose effects are felt with special

virulence in essential public services (health, education, public transport) and social cohesion in a society increasingly more subordinated to the power and interests of finance and big business [1].

Considering that, in a more recent study [2], young Europeans demonstrated clear pessimism about their expectations for the next 20 years, though less severe than among older citizens. It seems indisputable that Europe is steeped in what threatens to be the largest ever faced deadlock since the Schumann Declaration [3]. Given the skepticism of young Europeans about their future, a situation that endangers the credibility of European institutions and their rulers. It is necessary that schools and teachers develop initiatives and approaches aimed to provide ways of awareness and understanding of the historical, social, legal, cultural, economic and political dynamics, in which Europe is steeped.

If that is critical to the Society, schools have to promote European Integration studies as a cross-disciplinary field, covering the study of the origin and evolution of the European Union, as a whole, in all of its aspects. These include the analysis of both the internal and external dimension of European integration, including the European Union' role in the dialogue between peoples and cultures, in the youths mobility and integration into the labor market, in the citizens fundamental rights promotion, and others.

2.2 3D Virtual Worlds for Serious Applications

Virtual Worlds (VWs), such as Second Life or Open Simulator, are computer-generated environments, in which multiple users navigate, interact and communicate having a form of embodied representation [4]. Although not been created with that purpose, they offer many advantages for the development of different kind of serious applications, such as training [5]. In this sense, with the growing opportunities of applying educational technology, the use of these environments for training and education was also increased [6][7][8][9].

There is a possibility of participants playing different roles [10] in a problem-based learning methodology [11], with group dynamics and task-based approach [12], promoted by an active and collaborative knowledge construction process, where students can control their own learning with a high level of engagement [13]. That can be adopted for experiential and social learning, particularly in areas such as self-development, thinking skills, and critical civic conscience [14][11].

In fact, it is possible to apply a wide range of learning strategies, considering the versatility of these environments [15]. Having an representation of themselves, through an avatar, students being able to move within space and context with multiple sensory inputs, such as visual, auditory and tactile [13]. They also can communicate and participate in real time events and tasks, building objects and simulated- reality scenarios, seeing immediately the results of their own interaction with the environment and peers in a 3D way [14][11].

From a comparative point of view, virtual worlds such as Second Life and Open Simulator can't be confronted: the first is a social world and the second is a server software. However, Open Simulator can be a good solution to adopt. This platform is both open source and freely available for installation at organizations, such as schools

own servers, allowing them to start their own project, provided only to a specific educational community, and solving problems related to content protection, users age, backup policies, currency costs, among others [16].

2.3 User Learning Experience in 3D Virtual Worlds

As we saw above, a set of learning affordances from VWs are also pointed out by Dalgarno and Lee [17]. These environments can facilitate learning experience tasks that would be impractical to undertake in the real world, the transfer of knowledge and skills to real situations through contextualization of learning, and they also lead to increased intrinsic motivation, engagement, and effective collaborative learning than is possible with 2-D alternatives.

Although, some negative findings suggest that these environments promote great social presence rather than cognitive, showing that learning was not enhanced and the technology became a distraction [18][4]. For instance, in order to achieve higher levels of cognitive presence, the activities must be well structured and the student roles clearly defined [19].

There is still a lack of validated design principles and guidelines for developing learning spaces within VWs. User experience evaluation studies are scarce, and only a few studies have been reported in the literature related to interaction design and usability [20][21]. There are some studies related to social and motivational aspects of learning in VWs [22][18], that identifies learners pre-analysis and task-orientation as a path to increase student engagement in 3D VWs. Other researchers have focused on interaction with technological features of VWs, e.g., multisensory representation can help learners' interest, fun, ability to navigate [23][24], interaction with data gloves [25], etc.

An empirical research identifies interaction design and usability as key factors influencing the learning experience [26] . From the fields of urban planning, Human-Computer Interaction, Web usability, geography and psychology, the authors propose a cross-disciplinary set of design principles and guidelines for 3D learning spaces: 1- Conform the affordance of learning spaces and objects; 2- Have a story or narrative for the space; 3- Use real world metaphors; 4- Consider ambience and aesthetics of the learning space; 5- Realism for familiarity and comfort; 6- Orient the user at the landing or entry point; 7- Avoid spaces that can trigger phobic reaction or don't provide and easy exit; 8- "Design the form following the function".

3 The Young Europeans for Democracy Project

The Young Europeans for Democracy (YED) project aims to support the educational community, at vocational and secondary schools, in the process of European studies integration to promote an active citizenship between Portuguese youths. From our team prior experience in the field, we propose the design of 3D virtual learning scenarios through a problem-based learning methodology, with role-playing and group dynamics approaches. The main goal was the development of a serious application to train young students and their teachers, in the achievement of a wide range of knowledge and skills related to the European Union studies.

The YED project is a partnership between the University of Trás-os-Montes e Alto Douro (UTAD) and 5 vocational/secondary schools from Vila Real, Northeastern Portugal. During the project design and implementation, a set of different stakeholders will be involved. From the UTAD, a multidisciplinary team of 10 researchers (Professors, Tutors, Programmers and Designers) will design and implement the project with 11 teachers and 236 students, aged between 15 and 18 years old. Started on February 2012, the project execution is scheduled for one year, based on a five-step methodology.

3.1 Need Analysis

The first step was designed to collect and understand the students needs, regarding European citizenship and European studies integration. It aimed to identify the learning priority areas, in which the design and implementation of the serious application are supported - a similar process was developed with SME managers [27].

For that purpose, 4 workshops were made with 70 students of the 5 schools. These one hour sessions have included group tasks with brainstorming dynamics, open discussions and corresponding question exercises, in order to measure the students' prior knowledge, skills and perceptions on the topics of European Union historical, political, economic, social, legal, geographical and cultural dimensions. The approach was to present a wide range of problematic scenarios, regarding the European students' context and daily life, to be solved or discussed by them.

Based on a participatory observation methodology, we have identified a strongest lack of motivation by students to engage in the proposed learning tasks. We also understand, in a general way, that students show poor perceptions and didn't have the satisfactory thinking skills in order to discuss and solve the problems (less observed when they dealing with the historical, social, geographical and cultural dimensions of European Union). Therefore, we have defined a set of main learning priority areas and goals, constituted by a multidisciplinary field, that addresses a wide range of skills and knowledge, namely: 1- European Union history and organization (Know and understand the major treaties and EU symbols, the role of EU political, juridical, and economic institutions in its social context, etc.); 2- European fundamental rights (Interpret, understand and apply the fundamental rights as European youths to their own daily life's); 3- Intercultural dialogue (Know the EU role in the promotion of the main EU Member-states cultural references, like language, monuments and historical facts, personalities, gastronomy, religion, economic and political system, etc.); 4- Students and workers mobility (Analysis the opportunities that EU programs offer to European youth students and workers in other Member-states); and 5- Single currency and internal market (Confront the advantages and disadvantages of Euro currency and interpret the origin causes of European internal market crisis).

We also use the virtual world Second Life (SL viewer version 2.0) in the sessions, only to familiarize and provide the first contact with a similar environment to Open Simulator, and identify students' needs towards the use of technology. In general, the students' engagement was clearly higher than the expected and they easily have

executed a set of simple tasks, like communication with peers (eg. chat and IM) and in-world orientation (eg. movement controls, teleport, interaction with objects and camera control tasks).

3.2 Teachers Professional Training

We developed a set of training sessions to the eleven teachers involved in the project, into a different courses: 1 - We delivered a 25 hours b-learning course, with particular focus on teaching skills for the main priority learning areas and goals, previously identified in the first step. From open and participatory discussions about the topic contents, to the exercise of learning design skills, particularly with the use of education technology; 2- We have also delivered a 3 hours short-course, focused on the use of 3D virtual worlds as educational environments. The main goal was to present and exercise different possibilities of in-world learning scenarios, from simple activities (like in the first step with the youths), to a more complex tasks, like objects building and avatar customization (eg. forms, textures, positions, etc.).

3.3 Design and Development of the Serious Application

Based on some user experience design principles and guidelines presented in the literature [21][25][26], and accordingly to the main priority learning areas and goals, previously identified in the first step with students, we designed the serious application. Adopting a problem-based learning approach, role-playing tasks and group work dynamics, we propose a set of 5 different designed in-world and well structured learning activities (different spaces, learning goals, ways and kinds of interaction, score and feedback systems, time, and workload expectation) within a narrative through different simulated real-world scenarios (students have to achieve an historical perspective of their identity as youth European Union citizens; identify the necessary documents to travel to other European Union Member-states, geographically located, in order to explore different cultural aspects; help community in the promotion of the fundamental rights that they have as European Union citizens, regarding their own daily life's context and routines, etc.).

The serious application covers 8 virtual islands in UTAD's Open Simulator server, which one called by the name of different European Union Member-states, namely: Portugal, United Kingdom, Belgium, France, Italy, Spain, Greece and Germany. In Portuguese Island, is where the majority of the learning spaces (four) are located and is also where the learning process starts and finishes. We use a set of figures bellow, to represent and describe the learning spaces and the user behavioral within all the tasks. Due technical conditions, we defined teams with a maximum of 4 students (1 avatar per student), assigning each one with a different color (7 colors in the total) to play and achieve the learning tasks - that allow students to know which objects they may interact to identify and execute/solve the learning task/problem.

Firstly, Figure 1 presents the entry point of the serious application, located in the Portuguese Island and called "Terreiro do Paço" (a famous Portuguese square in Lisbon).

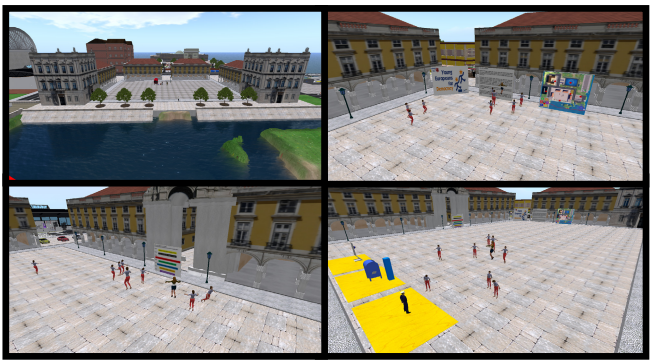


Fig. 1. Terreiro do Paço: The entry point of the serious application, where students know the main goal and rules of the whole game (eg. main narrative, group identification, general scoring and feedback system, objects and interaction ways - accordingly to related affordances and functions that cross-cuts all the learning spaces in the island). Learners are actively engaged in learning some basic controls of the viewer software, useful to interact with objects and characters within the tasks: movement control, camera control, chat, IM, inventory, notecards, touch-based landmarks and teleport, etc. For example, they have to interact with the Map of the Island, mailboxes, teleport buttons (similar to the start buttons of each activity), and with a Portuguese sailor that receives them with a "Welcome aboard" notecard.

Secondly, Figure 2 presents the learning space of the first activity, also located in the Portuguese Island and called "European Union Square".

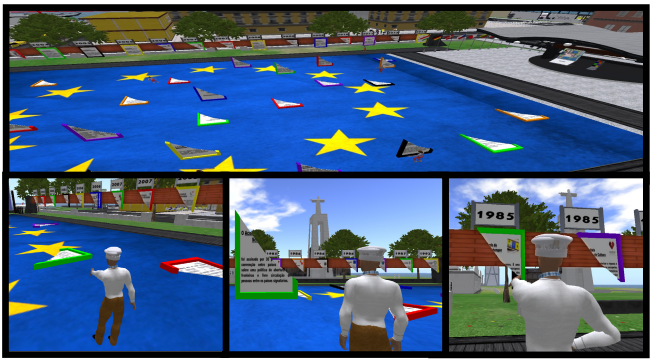


Fig. 2. European Union Square: The learning space of the first activity, where students have to know historically in which facts, symbols, treaties and missions consists the European Union of today, building a 3D collaborative timeline based on a puzzle-solving way. Each team have to touch in a set of 5 pieces of information, scattered around the floor, identified by different colors, and match them with their pairs (placed in the timeline dates).

Thirdly, Figure 3 presents the learning space of the second activity, also located in the Portuguese Island and called "Citizen Shop" (like Portuguese real-world "Lojas do Cidadão").

Fig. 3. Citizen Shop: The learning space of the second activity, where students have to find their situation (around its walls) in order to travel to a European Union Member-state (eg. looking for a job due the Portuguese financial crisis; winning a travel package; doing a school exchange; doing an inter-rail, acquire external health care services, etc.). Thereby, they have to check (touching) the necessary documents to travel in a main colored panel (each group have one), accordingly to their different situations.

Fourthly, Figure 4 presents the learning space of the third activity, also located in the Portuguese Island and called "Member-states Park".

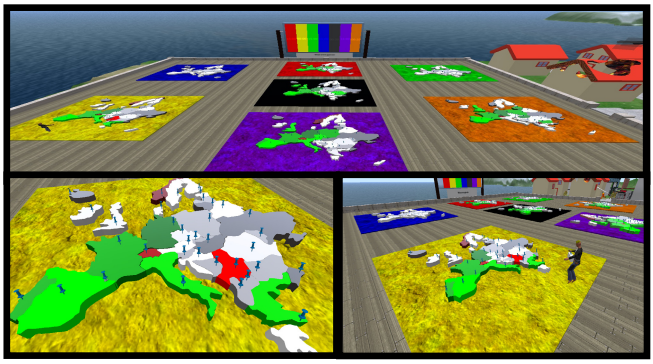


Fig. 4. Member-states Park: The learning space of the third activity, where students have to identify the Member-states and non Member-states of European Union in their team's map. The interaction with the objects is also touch-based, and each object represents an European country.

Fifthly, Figure 5 presents the learning space of the fourth activity, which in turn is located on the different islands of the application (United Kingdom, Belgium, France, Italy, Spain, Greece or Germany).

Fig. 5. United Kingdom: One of the different learning spaces of the fourth activity (which changes per group and depends on the situation found in the second activity). The students have to travel to a European Union Member-State, and when arrive there will be a game-clues style based on a multiple-choice question panel to solve. The main goal is to explore the cultural references of each country/space, designed with real-world objects, monuments and symbols (eg. Big Ben, British Museum, Londoner Routemaster bus, etc.), and find useful clues, distributed within the island, in order to select the correct solution/answer of the proposed problems/questions in the panel.

Finally, Figure 6 presents the learning space of the last activity, which is located in the Portuguese Island and called "Police station".

Fig. 6. Police station: The learning space of the fifth and last activity. After the activity 4, students get back to Portugal and meet a Portuguese police member that needs help, in order to solve different problems related to citizen rights around the island. There are 6 different real-world simulated case-scenarios outside the police station, distributed in the environment, which the teams have to interpret. For this, students have to interact (touch-based) with different characters (clearly identified with a white animated effect), in order to collect notecards with the case-descriptions. Posteriorly, students have to take into account the European Union fundamental rights, identifying the problematic cases (in which the rights are disrespected), and match them with a set of citizen rights, presented with numbers on multiple panels, inside the police station. For that purpose, students should edit each notecard with the matched numbers of the rights. Finally, students have to send all the edited notecards to a blue mailbox.

4 Preliminary Results and Research Agenda

We are actually in the fourth main phase of the YED project, delivering workshops with teachers and their classes, where students immerge within their youth citizenship 3D journey. We have followed different session programs, according to the availability of each teacher and class in the project. Some preliminary results are presented here, based on participatory observation within the use of the serious application.

Our finds are in agreement with other two studies [19][20], which suggest that learning was not enhanced and the technology became a distraction. In the majority of the learning spaces, students seems to be less focused on the learning priority content than in the social interaction between other students or secondary content. For instance, when tutors or peers are reading and explaining the instructions for each task, the other students don't stop to chat or editing the avatar appearance. Moreover, in the first activity, we observe that a higher average of students didn't read the information in the pieces as expected, and just simply try to match them without knowing the content.

Another limitation that we found are related to the feedback system. In spite of the students provided with performance feedback information (textual and visual represented) when they interact with the objects in the different tasks, the lack of feedback in beginning and ending of the activities seems to leave some students adrift and confused, with no awareness when the activity started or ended. However, positive aspects has been found. In general, students show high engagement in the use of the 3D Virtual World and a clearly satisfaction related to the different activities, real-world simulated scenarios, and types of user interaction with objects and characters within the spaces. This seems to be related with some of the Minocha & Reeves design principles [26] that we followed in order to improve the learning experience, such as designing for affordance of the learning spaces, for storytelling, using real world metaphors, consider realism for familiarity and orient the user at the entry point.

Thus, in this first analysis, we noticed the existence of a high level of incidental learning - at least this is some level of engagement that might have some benefits over not engaging at all. However, this surface level of knowledge leads us to think in alternative ways of getting a more in-depth engagement (provided through well-structured activities with continued feedback over all the steps of each task).

We are collecting data and deepen the analysis on the impact of the project through different instruments and methods. From pre- and post-test, to interviews and questionnaires on the students and teachers expectations, perceptions, and satisfaction regarding the impact of the 3D serious application in their motivation and engagement for learning and teaching about European Union. Thus, and through the innovative nature of the YED project, we want to measure in-depth, from the user experience point of view, all the potential and limitations of the adopted learning design and user interaction approach.

Acknowledgments. YED is sponsored by the European Commission's Education & Culture Directorate-General, under the Lifelong Learning Programme, contract nr. 357639-LLP-2011-PT-AJM-ICS.

References

1. S.A.R.L., O.: Perceptions of the European Union: A qualitative study of the public's attitudes to and expectations of the European Union in the 15 Member States and in 9 candidate countries. General report, European Commission (2001)
2. Commission, E.: Youth in Europe: a statistical portrait. In: Eurostat Statistic Books. Publications Office of the European Union, Luxemburg (2009)
3. Schuman, R.: The Schuman Declaration. In: Nelson, B., Stubb's, A. (eds.) The European Union: Readings on the Theory and Practice of European Integration, 2nd edn., Macmillan Press Ltd., London (1998)
4. Omale, N., Hung, W., Luetkehans, L., Cooke-Plagwitz, J.: Learning in 3-D multiuser virtual environments: Exploring the use of unique 3-D attributes for online problem-based learning. British Journal of Educational Technology 40(3), 480–495 (2009)
5. de Freitas, S.: Serious virtual worlds: A scoping study. Technical report, Joint Information Systems Commitee. Bristol, UK (2008)
6. Baker, S.C., Wentz, R., Woods, M.: Using virtual worlds in education: Second Life® as an educational tool. Teaching of Psychology 36(1), 59–64 (2009)
7. Lopes, A., Pires, B., Cardoso, M., Santos, A., Peixinho, F., Sequeira, P., Morgado, L.: System for Defining and Reproducing Handball Strategies in Second Life On-Demand for Handball Coaches' Education. In: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, ED-MEDIA 2009, vol. 2009, pp. 3954–3961. AACE, Chesapeake (2009)
8. Morgado, L., Varajão, J., Coelho, D., Rodrigues, C., Sancin, C., Castello, V.: The Attributes and Advantages of Virtual Worlds for Real World Training. The Journal of Virtual Worlds and Education 1(1), 2159–9238 (2010)
9. Frias, P., Cruz, R., Fernandes, R.: Second Language Teaching in Virtual Worlds: The Case of European College Students under the ERASMUS Programme. In: Vincenti, G., Braman, J. (eds.) Multi-user Virtual Environments for the Classroom: Practical Approaches to Teaching in Virtual Worlds. IGI Global (2011)
10. Vasileiou, V., Paraskeva, F.: Teaching Role-Playing Instruction in Second Life: An Exploratory Study. Journal of Information Technology and Organizations 5, 25–50 (2010)
11. Vosinakis, S., Koutsabasis, P., Zaharias, P.: An Exploratory Study of Problem-Based Learning in Virtual Worlds. In: 2011 Third International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES), pp. 112–119. IEEE Computer Society, Washington DC (2011)
12. van der Lan, A.S., Schoutend, P., van den Hooff, B., Feldberg, F.: Modelling the Metaverse: A Theoretical Model of Effective Team Collaboration in 3D Virtual Environments. Journal of Virtual Worlds Research 4(3) (2011)
13. Perez, L.: Challenges and Opportunities in Using MUVEs in K-12 Environments. In: Tettegah, S., Calongne, C. (eds.) Identity, Learning and Support in Virtual Environments, pp. 45–55. Sense Publishers, Rotterdam (2009)
14. Dawley, L.: Social network knowledge construction emerging virtual world pedagogy. On the Horizon 17(2), 109–121 (2009)

15. Leonard, L.G., Sherblom, J.C., Withers, L.A.: Communication Challenges and Opportunities for Educators Using Second Life. Higher Education in Virtual Worlds Teaching and Learning in Second Life. Emerald Group Publishing Limited (2009)
16. Vilela, A., Cardoso, M., Martins, D., Santos, A., Moreira, L., Paredes, H., Martins, P., Morgado, L.: Privacy challenges and methods for virtual classrooms in Second Life Grid and OpenSimulator. In: 2010 Second International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES), pp. 167–174. IEEE Computer Society, Washington DC (2010)
17. Dalgarno, B., Lee, M.J.W.: What are the learning affordances of 3-D virtual environments? British Journal of Educational Technology 41(1), 10–32 (2010)
18. Lim, C.P., Nonis, D., Hedberg, J.: Gaming in a 3D multi-user virtual environment (MUVE): Engaging students in Science lessons. British Journal of Educational Technology 37(2), 211–231 (2006)
19. Kanuka, H., Rourke, L., LaFlamme, E.: The influence of instructional methods on the quality of online discussion. British Journal of Educational Technology 38(2), 260–271 (2007)
20. Zaharias, P., Belk, M., Germanakos, P., Samaras, G.: User Experience in Educational Virtual Worlds. In: CHI 2011, ACM 978-1-4503-0268-5/11/05, Vancouver (2011)
21. Hasler, B.S., Buecheler, T., Pfeifer, R.: Collaborative Work in 3D Virtual Environments: A Research Agenda and Operational Framework. In: Ozok, A.A., Zaphiris, P. (eds.) OCSC 2009. LNCS, vol. 5621, pp. 23–32. Springer, Heidelberg (2009)
22. Barab, S.A., Thomas, M., Dodge, T., Carteaux, R., Tuzun, H.: Making learning fun: Quest Atlantis, a game without guns. Educational Technology Research and Development 53(1), 86–108 (2005)
23. Sato, M., Liu, X., Murayama, J., Akahane, K., Isshiki, M.: A Haptic Virtual Environment for Molecular Chemistry Education. In: Pan, Z., Cheok, D.A.D., Müller, W., El Rhalibi, A. (eds.) Transactions on Edutainment I. LNCS, vol. 5080, pp. 28–39. Springer, Heidelberg (2008)
24. Minogue, J., Jones, M.G., Broadwell, B., Oppewall, T.: The impact of haptic augmentation on middle school students' conceptions of the animal cell. Virtual Reality 10(3), 293–305 (2006)
25. Mikropoulos, T.: Presence: a unique characteristic in educational virtual environments. Virtual Reality 10(3), 197–206 (2006)
26. Minocha, S., Reeves, A.J.: Interaction Design and Usability of Learning Spaces in 3D Multi-user Virtual Worlds. Information and Communication Technology, 157–167 (2010)
27. Dominguez, C., Varajão, J., Morgado, L., Oliveira, I., Sousa, F.: SME Managers' Most Important Entrepreneurship and Business Competences. In: Quintela Varajão, J.E., Cruz-Cunha, M.M., Putnik, G.D., Trigo, A. (eds.) CENTERIS 2010. CCIS, vol. 110, pp. 274–282. Springer, Heidelberg (2010)

The Behaviour Assessment Model for the Analysis and Evaluation of Pervasive Services

Bernhard Klein, Ivan Pretel, Ulf-Dietrich Reips,
Ana B. Lago, and Diego Lopez-de-Ipiña

DeustoTech –Deusto Institute of Technology, Universidad de Deusto
Avenida de las Universidades 24, 48007, Bilbao, Spain
{bernhard.klein,ivan.pretel,reips,
anabelen.lago,dipina}@deusto.es

Abstract. Several mobile acceptance models exist today that focus on user interface handling and usage frequency evaluation. Since mobile applications reach much deeper into everyday life, it is however important to better consider user behaviour for the service evaluation. In this paper we introduce the Behaviour Assessment Model (BAM), which is designed to gaining insights about how well services enable, enhance and replace human activities. More specifically, the basic columns of the evaluation framework concentrate on (1) service actuation in relation to the current user context, (2) the balance between service usage effort and benefit, and (3) the degree to which community knowledge can be exploited. The evaluation is guided by a process model that specifies individual steps of data capturing, aggregation, and final assessment. The BAM helps to gain stronger insights regarding characteristic usage hotspots, frequent usage patterns, and leveraging of networking effects showing more realistically the strengths and weaknesses of mobile services.

Keywords: Mobile services, user acceptance, human-computer interaction, technological acceptance model, log data analysis, service design, living lab.

1 Introduction

User acceptance in field trials is still mostly evaluated through questionnaires and focus interviews. Mobile applications are, however, much stronger related to real mobile behaviour as people carry their devices with them. Because of the dependency of mobile applications' functionalities on the user situation answers to general questions about the application can often not easily be given.

A relative new approach for mobile services is the observation of application usage through data loggers. A data logger records application events or errors jointly with other usage or system related metadata. To support daily activities successfully, mobile applications should not interrupt the activities, provide a reasonable interaction/benefit ratio to the user, and provide community leveraging beyond exploitation of personal experience. Analysing usage hotspots, usage frequency and usage type allow researchers to speculate about potential strengths, weaknesses or even problems for the surveyed service.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1129–1140.
DOI: 10.1007/978-3-642-36981-0_107

In this work, we present a behaviour assessment framework that describes the systematic collection of behaviour data and guides researchers in their log data analysis. With such an analysis approach researchers can gain more insights about first and long term service impressions, acceptances issues correlated with the user experience and the success of subsequent product evolvement steps.

The article is structured as follows. The next section discusses related works. Section 3 describes the method how to capture, aggregate, and represent data. In Section 4 the behaviour assessment model is defined. A preliminary case study is given in Section 5. Section 6 concludes the article.

2 Related Work

In order to perform a study focused on mobile services the first step is to compare, analyse and find the differences and connections between existing data loggers, concepts and conclusions related to the mobile services evaluation field.

Lab-based evaluation frameworks log information in a controlled environment using specific devices and specific users. The main advantages of the lab-based frameworks are the highly controllable environment and the collection of data, which is cheap and easy. However, the context, which is the most influential factor in the mobile services field, is not considered and it can hardly be simulated. Many simulation tools produce highly inaccurate results because of the context. Furthermore, several agents also alter the results of user experiments. The experts who lead the experiment and the tasks performed by the users can not only alter the execution of experiments but also evoke situations that would never happen in real environments. The users may also add biased results during the execution of the experiments [1] because they suffer several problems such as test-anxiety [1]: during the task performance the highly test-anxious person divides his attention between self-relevant and task-relevant variables; due to the self-focussed attention the user of the mobile service may not show real behaviour. Further, in many tasks such as phone calls, it would be subjectively annoying for many users to be in a room with observing researchers.

On the other hand the field-based evaluation frameworks (see Table 1) capture information in real environments. They commonly use added cameras and human observers to capture information from the interactions. Furthermore, this kind of framework tries to bring the lab to the field. For example, the *Usertesting* platform [2] not only brings methods like the think-aloud verbal protocol but also records the user's feedback with a webcam; finally it reproduces the interaction again enabling the annotations during it. Using this kind of techniques means that although the task is performed in real environment, it is changed and consequently, the interaction altered. Another tool related to *Usertesting* is the *Morae Observer* [3] tool. It captures all the interaction data and indexes it to one master timeline for instant retrieval and analysis; it generates graphs of usability metrics. Both tools are focused on the interaction because they are centred on capture of screen interaction and the user's feedback through filming the face or recording comments. Another group of tools such as *ContextPhone* [4] and *RECON* [5] are focused on the context capture. They capture

the surrounding environment through mobile sensors. This capturing technique retrieves a lot of real data without influencing the interaction but the user's feedback is lost. In order to fill the lack of the user's feedback other tools like *MyExperience* [6] and *SocioXensor* [7] use techniques like self-reports, surveys and interviews mixed with the context capture. These tools are quite powerful and flexible because the user has at any time the complete control about when participate in an application acceptance survey. In case, he has been interrupted in the survey he can resume it to a later point of time.

Table 1. Properties of the logging tools

Tool	Capture techniques	Data	Report
Usertesting	Screen, webcam and microphone	Interaction, user information and user's feedback	Reproduce the screen interaction
Morae Observer	Screen, webcam and microphone, observer	Interaction, user information and user's feedback	Reproduce the interactions and calculate graphs
ContextPhone	Mobile sensing and interaction event logging	Interaction, device status and environment	Mobility patterns detection
RECON	Interaction event logging and mobile sensing	Interaction, device status, user information, user's feedback, and environment	Trace Data analysis Engine
MyExperience	Wearable hardware sensing, mobile sensing, audio recording and user surveys	Interaction, device status and environment	Performance analysis, SMS usage and mobility analysis
SocioXensor	Interaction event logging, survey, interview	Interaction, user, device status and environment	SQL database

To sum up, to acquire valid interaction data about mobile services, it is essential to capture objective information to solve questions like when, where, how long, etc. users are really interacting with a service. These questions can hardly be determined with a lab-based framework. The field-based evaluation frameworks can provide deeper and more objective information, but the added agents such as cameras and invasive evaluation methods (e.g. think-aloud verbal protocols) have to be removed. In order to do so, the best way to capture interaction data is by registering information through a mobile device using a tiny capture tool. This tool should log the context via the built-in mobile sensors and logging the key interaction events.

3 Mobile Service Assessment through Behaviour Analysis

A framework for automatically logging and processing data for evaluation has been developed. In the following we briefly explain the different behaviour capturing and aggregation phases and the architectural requirements.

3.1 Data Logging and Aggregation Overview

As can be seen in Fig. 1 the framework distinguishes four main phases:

1. *Data Capture*: A data logger component installed separately on the mobile device records event and error data triggered by the mobile service. Examples for logging data are: service start and stop times, UI events e.g. buttons pressed, screen transitions, any changes in settings and erroneous data entries, exceptions and any unexpected system behaviour. These data are complemented with additional user contexts (e.g. provider and subscriber data), service information (e.g. queries/results, content data, screen stay duration) and device contexts (e.g. location data) for further evaluation.
2. *Transfer Protocol*: Logging data is periodically (e.g. daily) transferred to an analysis component hosted on the Internet. To minimize the influence on mobile service performance the transfer process is only started if the mobile device remains in an idle execution state.
3. *Data Aggregation*: The analysis component parses the incoming logging data and interprets the raw data log format with a parser. A filter process removes out-of-bound values, spatio-temporal inconsistencies, and entries that do not conform to preset criteria. Following this filtering step the log data are aggregated through clustering analysis.
4. *Data Visualization*: From the results tables, graphs and diagrams are generated for the researcher. Furthermore, the entire log is automatically annotated so that each entry is written out for human readability and annotated to get basic derived information such as duration and transitions.

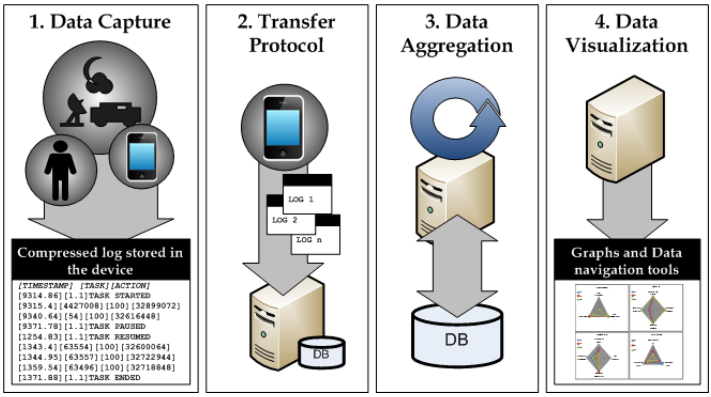


Fig. 1. Graphical description of the process

3.2 System Architecture

The Neurona evaluation framework [8] was extended to meet the BAM requirements. This platform shown in Fig. 2 is based on three main components: the mobile device component, the connector component, and the analysis server component.

The Mobile Device component is software installed in the user's mobile and logs user interactions; it is formed by the Logger/App interface, Logger Module and Context Information Module. The Logger/App interface is a tiny software library used to send interaction events to the logger module. The logger module stores the interaction data and shows brief questionnaires about the interaction experience to capture the user's feedback; these questionnaires are shown at the end of the interaction to not disturb the experience. Another element is the context information module, which provides context information acquired from the built-in mobile sensors and the mobile Operative System.

The Analysis Server component is hosted in a web server; this component is formed by the Data Aggregation Module, the Visualization Module, the Applications Manager and the Usergroup Administration. The Data Aggregation Module receives logged data and calculates normalized information to store it in the system database. The expert who wants to check the normalized information can do it using the Visualization Module; which shows advanced graphs. The Applications Manager enables the expert to register into the system, update and remotely configure prototype applications. The Usergroup Administration module registers users and devices, assigns applications and exposes several administration options related to the relations between users, applications and experts.

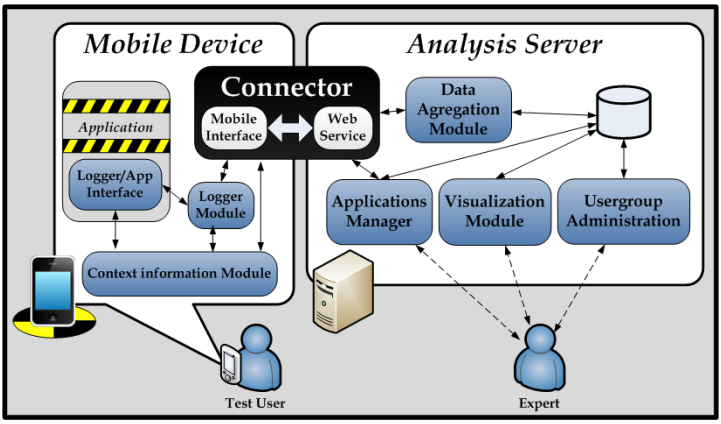


Fig. 2. System architecture

Finally, the connector between the explained elements transfers the logged information generated by the Mobile Device component to the Analysis Server component. It is divided in two main elements: the mobile interface and the server web service. Basically the mobile interface checks the state of the device and if the user is not interacting with the device it sends logged data to the web service hosted in the Analysis server. To minimize the required transfer bandwidth logging data is encoded in memory saving format and decoded later to a human readable format when the logging data has been received by the Analysis Server.

4 The Behaviour Assessment Model

A proven performance assessment method considering concurrent aspects has been the Balanced Scoreboard (BSC) approach. Aligning each of the dimensions systematically helps get a better impression about different influencing factors.

4.1 Dimensions of the Behaviour Assessment Model

The BAM is based on balanced set of behaviour categories which are orthogonal:

- *Planned and spontaneous execution scenarios*: According to Ajzen [9] people in unfamiliar situations often prefer to plan their activities, whereas people in familiar environments rely on their capabilities and thus react more spontaneously. Adequate mobile services have to support both scenarios; offering a remote and location based execution of their services (see Colbert [10]).
- *Service actuation and service interaction*: It is generally accepted that a seamless integration of mobile services in daily life is only given if mobile services raise users' attention in periods where the user is not interrupted, if the interaction efforts keeps a reasonable balance with it intended service benefit.
- *Central provision and community networking*: Mobile services targeting a broad proportion of the audience are better provided through a central provider. However, with peer-to-peer infrastructures people can also provide services to smaller user groups e.g. friend group or only provide them for a short time frame.

This leads to following six different dimensions illustrated in Fig. 3:

1. *Remote activity discovery*: This dimension is based on the categories Planned Execution Scenario and Service Actuation. In order to fulfil end-users need to plan activities ahead of a trip; users require the capability to explore the service offer according to given properties. The retrieval quality depends on the query power e.g. different search concepts and the query success rate. An example is a map based discovery tool, which retrieves services according to locations selected on a map.
2. *Situation-aware activity recommendations*: The dimension founds on the categories Spontaneous Execution and Service Actuation. As mobile services are much stronger correlated with the daily life of end-users an important requirement is to raise their attention to an adequate service offer in a seamingless manner. A successful implementation depends on the reasoning power (that compare the current users' context and the intended service context) and the number of directly consumed services (reasoning success).
3. *Mobile activity creation*: The categories Planned Execution and Service Interaction define this dimension. Complex mobile services require often too much knowledge from the user to execute them easily on the spot. Therefore, services should offer any type of service creation, personalization or reservation functionality so that they can be consumed better in time constrained situations. The editing complexity and the service content quality are important indicator examples to determine this dimension.

4. *On-the-spot activity support*: The dimension is constructed through the categories Spontaneous Execution and Service Interaction. Since users on the move often follow other real-world activities it is important that the attention needed to execute the service is kept to an absolute minimum. The navigation complexity (effort) and the quality of the content provided by the service are important indicator examples.
5. *General platform activity services*: This dimension stems from the categories Planned Execution and Central Provision. All general service aspects influencing the provision quality e.g. power consumption and error handling account for this dimension.
6. *User-created activity services*: This dimension is founded on the category Spontaneous Execution and Community Networking. Tools that consider community behaviour can help in structuring the knowledge space further and lead to more transparency in the community. Examples are best-of ranking lists, member reputation lists and content recommender systems. For instance car sharing opportunities can be more easily evaluated by users and improve their selection. Suitable example indicators are the lurker ratio (active community participation) and the degree of community transparency achieved with previously mentioned community services.

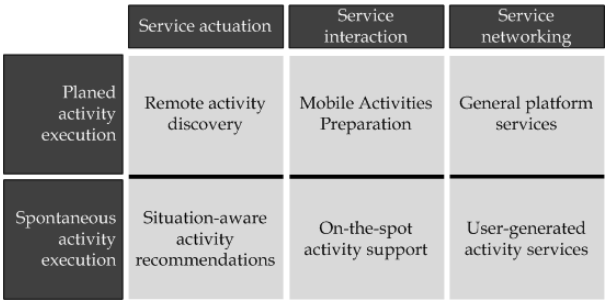


Fig. 3. Dimensions of the behaviour model

4.2 The Balanced Scoreboard Assessment Approach

These six dimensions focus on realistic service usage. This emphasises the valuation of a service by the way how end-users apply services to solve given problems. Such behaviour patterns have the potential to tell us about underlying reasons why specific service fail or become well accepted. Recording such behaviourally relevant data also allow the emulation of service usage in respect to given user´s context. Both aspects are important for developers to continuously improve the service. According to the BSC approach, the intention is to find a few aggregated indicators that quantify a given dimension. The indicator must meet the requirements of reasonability and measurability. A general problem of social surveys is to translate the indicators into precise measures. The abstract classes of measurement types, correspond hereby with different event and error logging data types. To achieve comparability between different numerical scales of measurements e.g. an event/error frequency scale, a

function has to be defined which maps selected scale areas on specific quality rating values. Since humans perceive the influence of various indicators for a given dimension differently, weight coefficients are used to balance the influence of individual indicators. Both mapping function properties and weight coefficients can be obtained through a profiling questionnaire prior to the field trials.

Finally, the results of an analysis and evaluation are typically held in a spreadsheet for detailed analysis and visualised by a radar chart for a summarised representation (see Fig. 4). For visualisation by a radar chart, the six dimensions are equally arranged. The scaling is adapted appropriately according to the distribution of the measurement results with its positive orientation towards the origin. For a better visualisation of the consequences of the results, each scale can be subdivided in fulfilled (positive centre areas), and not fulfilled (negative edge areas).

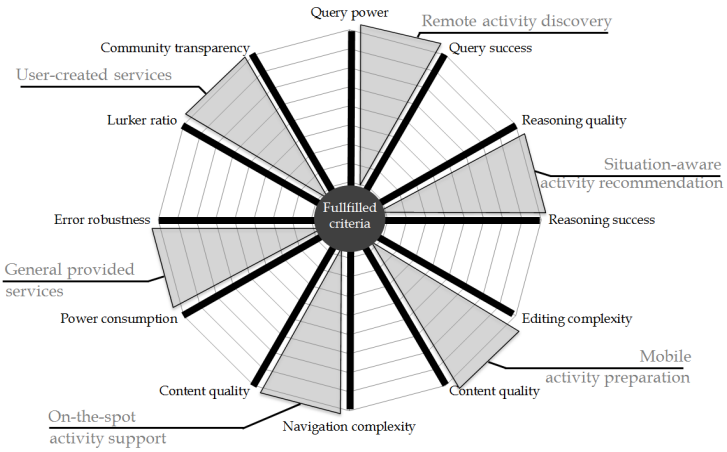


Fig. 4. Visualization of the behaviour model with six dimensions (grey colour) and example indicators (black colour)

5 Case Study of the MUGGES System

Mobile User Generated Geo Services (MUGGES) is a European research project with the goal of evaluating peer-to-peer service concepts based on Global Positioning Systems for mobile phones. MUGGES provides an infrastructure to create, publish, provide and consume mobile micro-services directly from mobile devices (see Klein et al. [11]). As part of the project four application prototypes were developed, which allow the description and sharing of places and routes between users. A field trial was conducted with early adopters in real environments. In the following we will demonstrate how the BAM framework can help to identify benefits and best practices for MUGGES type of applications.

5.1 The Assessment Process

Applying the BAM analysis technique requires specific preparation steps. These include the definition of indicators for each dimension, correlating them with available logging data, appropriate balancing of these measurements with weight factors, the execution of field trials and representing the results. In the following the assessment process is explained in more detail for each phase (see Fig. 5):

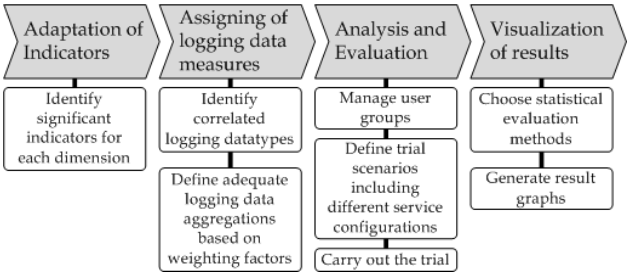


Fig. 5. Different phases of the assessment process and correlated activities

- *Adaptation of indicator structure*. First, adequate indicators have to be identified which represent a specific dimension of the BAM model. Since this is an intuitive process it is recommended to develop this in collaborative process among developers and potential end-users. The design is completed with a proof of plausibility of the defined criteria.
- *Assignment of logging data types*. In this phase, logging data types are correlated with the indicators. In order to achieve a balanced view of all indicators and logging data types weighting factors are applied. Questionnaires can be used to determine the individual importance of each dimension and corresponding logging data types to modify the weight factors accordingly.
- *Analysis and evaluation*. In order to learn more about the characteristic service behaviour it is recommended to conduct a series of experiments. Goals for every single indicator should be determined before the actual start of the field trials, in order to compare these to the empirical results. Then, initial and long-term service usage should be compared in order to identify entry barriers. In a subsequent step, the logging data of trial users shall be clustered according to the technical diffusion model from Rogers [12] to see how technical experience can influence mobile service usage. By comparing innovators, early adopters, early majority, late majority and laggards a reasonable priority list of future service modifications can be determined. Finally it is also important to analyse logging data from different trials in order to see to what extend applied service modifications have lead to an acceptance improvement.
- *Visualisation of evaluation results*. The results of the test group are analysed and evaluated with statistical methods and visualised according to the radar graph approach.

Generally an iterative evaluation approach is recommended starting from early prototypes up to the final mobile service. In order to compare the development

progress it is important not to vary the measurement criteria. It is assumed that the explanatory power of the BAM model increases with a stronger concretization of the mobile service during the development cycle.

5.2 Visual Evaluation of the MUGGES System

We demonstrate the advantages of the BAM with a small example based on the field trial executed for the MUGGES project. In this field trial logging data from 30 potential end-users have been collected during a 2 week period. Each study participant was given a mobile phone with the MUGGES software installed. The users were given specific tasks e.g. describing their favourite shopping route or leisure places. The connector component transfers periodically event data to the analysis server for further evaluation.

For the dimension remote activity discovery and situation-aware activity recommendation query power (measured as average number of applied search keywords or average distance of user/service location at the query time) and query success (measured as average search-consumer ratio) is relevant indicators. The dimension mobile activity preparation, on-the-spot activity support is determined by the indicators editing/navigation complexity (measured through the average screen stay duration per service) and service content quality (measured by the average content length, average number of comments and average update duration per service). The dimension General Platform Services is defined by the indicators power consumption (measured as consumed energy units per day) and error recovery quality (reciprocal number of occurred errors per day). And finally, the dimension user-created activity services are defined by the lurker ratio (measured as provider-consumer ratio of a consumed service) and the consumed service quality (measured through average rating of consumed services).

The MUGGES system has been generally be well accepted as the average mugglet creation rate dropped only insignificantly after the trial kick-off and stayed roughly at about 60 mugglets created per day. Still the MUGGES infrastructure revealed some weaknesses. Applying the BAM approach (with a rating range from 0 – very good till 3 very bad) a service provider can come, for example, to the following simplified conclusions concerning the following dimensions illustrated by Fig. 6:

- *Remote activity discovery*: With increasingly more created mugglets users applied more sophisticated search approaches (from simple template, keyword-based and map-based search) to compensate the small screen size. The discovery function seem to work well for the majority of the trial users (rating 1.5).
- *Situation-aware activity recommendation*: The overall distance between the mugglet location and the trial user has been quite far (up to 1 km). Besides the sparse distribution of the mugglets another reason has been the bad performance of the location technology. Provider could conclude that the recommendation service is not sufficient (rating 2.5) for the current spontaneous usage scenario.
- *Mobile activity preparation*: The mugglet creation process took a lot of time, not short enough to create mugglets on-the-go. People compensated this by

distributing the creation process in several phases. The mobile activity preparation is not sufficient (rating 3.0) in the current development stage.

- *On-the-spot activity support*: The mugglets in general have high information intensity for the user, as they come with a environment map, text descriptions, comments and photos. Above that, the real-time notification feature helped people to stay up-to-date. Mugglets thus have been very useful (rating 1).
- *General provider services*: The peer-to-peer service sharing approach has lead to an high power consumption and the error rate has been quite high. Service provider may conclude that device-to-device sharing is a bad option (rating 2.5) and moving mugglet sharing into the Internet cloud may be a better option.
- *User-created services*: MUGGES usage has been high since users could create their own personal service based on the offered service templates. Especially in later stages during the project service ratings have been found very useful (rating 1) to identify popular services.

The radar graph shows some important weaknesses. Recommender systems, the mugglet creation process and the provider infrastructure still make an everyday usage difficult (see Fig. 6). Comparing these logging data results with the questionnaires conducted after the trial backs these findings. But more importantly, user perceptions were not always clear enough to pinpoint the exact problems with the MUGGES infrastructure. The evaluation with the BAM is more differentiated and considers some critical aspects that influence the acceptance of this mobile service significantly.

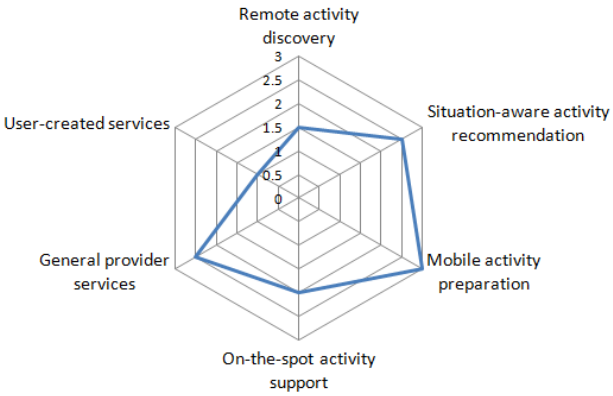


Fig. 6. Evaluation of the MUGGES System

6 Conclusions and Future Work

We introduced the BAM as an instrument for the analysis and evaluation of the user acceptance for mobile services. The BAM is characterised by a structure that helps to identify systematically a balanced set of important, individually measurable and

independent acceptance criteria. The application of the BAM is guided by a process model that supports all phases from the development of acceptance criteria over the measurement of relevant indicators to the evaluation and visualisation of the derived results. Using the BAM reveals several insights:

- *First and permanent usage patterns*. Analyzing the radar graph at the beginning of the trial and later phases of the trial shows can show entry barriers of the mobile service. Results obtained in later stages show how people exploit mobile service strengths but also compensate potential weaknesses of the service.
- *Usage patterns for different technical adoption groups*. According to Rogers technical diffusion model user groups are divided in innovators, early adopters, early majority, late majority and laggards. Clustering logging data according to these groups may reveal interesting insights how the technical experience influences service usage. These observations are especially valuable to define a priority of feature improvements for the mobile service.
- *Behaviour changes in different product development stages*. As the development of the mobile service evolves comparing results with earlier trials can help to confirm if the applied feature modifications fulfil the intended improvements.

References

1. Reips, U.-D., Stieger, S.: Scientific LogAnalyzer: A Web-based tool for analyses of server log files in psychological research. Behavior Research Methods, Instruments, & Computers 36, 304–311 (2004)
2. Cassady, J.C., Johnson, R.E.: Cognitive Test Anxiety and Academic Performance. Contemporary Educational Psychology 27, 270–295 (2002)
3. UserTesting.com - Low Cost Usability Testing, http://www.usertesting.com
4. Morae usability testing tools from TechSmith, http://www.techsmith.com
5. Raento, M., Oulasvirta, A., Petit, R., Toivonen, H.: ContextPhone: A Prototyping Platform for Context-Aware Mobile Applications. IEEE Pervasive Computing 4(2), 51–59 (2005)
6. Jensen, K.L.: RECON: Capturing Mobile and Ubiquitous Interaction in Real Contexts. In: Proceedings of MobileHCI 2009, Bonn, Germany (2009)
7. Froehlich, J., Chen, M.Y., Consolvo, S., Harrison, B., Landay, J.A.: MyExperience: A System for In situ Tracing and Capturing of User Feedback on Mobile Phones. In: MobiSys 2007, June 11-14, pp. 57–70. ACM, San Juan (2007)
8. ter Hofte, H., Otte, R., Peddemors, A., Mulder, I.: What's Your Lab Doing in My Pocket? Supporting Mobile Field Studies with SocioXensor. In: CSCW 2006, Conference Supplement, Banff, Alberta, Canada, November 4-8 (2006)
9. Pretel, I., Lago, A.B.: Capturing Mobile Devices Interactions Minimizing the External Influence. In: Proc. of UBICOMM, pp. 200–205 (2011)
10. Ajzen, I.: The theory of planned behavior. Organizational Behavior and Human Decision Processes 50(2), 179–211 (1991)
11. Colbert, M.: A diary study on rendezvousing: Implications for position-aware computing and communications for the general public. In: Proc. of Group Conference (2001)
12. Klein, B., Perez, J., Guggenmos, C., Pihlajamaa, O., Heino, I., Ser, J.: Social acceptance and usage experiences from a mobile location-aware service environment. In: Ser, J., et al. (eds.) Mobile Lightweight Wireless Systems, vol. 81, pp. 186–197. Springer, Berlin (2012)

A Study of Biometric Authentication Adoption in Health Services

Paulo Rodrigues and Henrique Santos

Centro Algoritmi, Universidade do Minho
paulomrrodrigues@yahoo.com,
hsantos@uminho.pt

Abstract. Authentication technologies using biometrics already play a major role in the context of security systems. Some techniques are used in several services not only to ensure physical access to facilities but also to provide some thoroughness in access control. This study addresses the perception of users regarding the use of biometric authentication technologies. The results derived from a survey indicate a trend of acceptance of such technologies on the access control to health services.

Keywords: Biometric, technology adoption, identification, authentication.

1 Introduction

Society is increasingly dependent on new authentication technologies. We see this phenomenon growing at a global level and governments are turning towards to biometric authentication technologies to provide identity assurance[1].

Researchers in this field state that the need for security will be even more intense, which will have a strong impact on the use of biometric technologies by making them the most popular method of user identification and /or authentication [2]. In recent years we have observed a gradual researching effort to increase security and improve the user experience [3].

This study aimed to assess the perception of health services users regarding the use of biometric authentication technologies, which is fundamental to derive proper user requirements supporting the development of adequate (both in security and flexibility dimensions) user interfaces, within health ecosystems.

For this purpose we made a literature review about the state of the art around these technologies and used, as a theoretical model for drafting a proper questionnaire, the Technology Acceptance Model (TAM) proposed by Davis[4], and the extension of TAM from James et al.[3]. We obtained 596 valid questionnaires, from a random sample stratified by sex and age, of users of health services in a hospital.

This study is the first phase of a most comprehensive study which aims to assess the practicability of using biometric technologies for user authentication within health information systems services. The preliminary results indicate that the majority of participants recognize advantages towards the adoption of such systems.

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1141–1148.
DOI: 10.1007/978-3-642-36981-0_108

2 Literature Review and Theoretical Background

2.1 Biometric Technologies

Biometrics has been used throughout the existence of man as the most common (the only one) method of identification. Every day we use the voice dialogue, and face-to-face interaction, to recognize friends and family. Like other characteristics, such as odor (newborn babies identify their mothers), sounds and motion.

These technologies are usually classified as physical or behavioral. The physical characteristics include hand geometry, fingerprints, iris, retina, facial characteristics, DNA, ear and lip motion recognition [5, 6]. In the second case, we are focused on the activities carried out by an individual, such as signatures, voice (actually sharing both characteristics), keystroke dynamics and gait [5, 6]. Some of these activities occur on a voluntary basis and other mechanized through its repeated application [7].

Biometrics to be considered qualified must meet the following principles [8–10]: *universality, uniqueness, measurability, performance, acceptability and protection*.

2.2 Biometric Authentication

Depending on the application context, an authentication system can use the following methods: identification and verification [8, 11].

The main difference between verification and identification is primarily defined by the matching samples process in the system. Verification refers to the confirmation or denial of an alleged identity of an individual, while the identification refers to the problem of establishing (or denying) the identity of an individual against a population previously enrolled within the system [12, 13].

Identification is a more complex problem in terms of design and implementation, given the number of sample comparisons required, especially for large populations [14]. Verification is typically performed as a login process, requiring only a one-to-one comparison and supporting the decision on the result of a similarity function.

2.3 Vulnerabilities, Risks and Challenges

This kind of technology makes life more difficult for impostors and somehow enforces non-repudiation. In addition, users are not exposed to phishing like attacks, and so they are not so exposed to social engineers threats [8].

However, biometrics should not be considered a secret [8], it is possible someone to copy or forge a biometric trait, such as: voice mail, a glass, or even a lost photo with resolution enough to handle a iris sample [15].

Alterman [16] reinforces this concern, stating that when we provide our biometric characteristics for registration in a database, we are constrained that after a fraudulent access we cannot immediately recover our identity.

Forging biometrics is not so easy, requires experience, money, time and privilege access to the system[14]. According to the European Report conducted in 2005 [17] the errors and abuses of identity will be less frequent with the use of biometrics. But

when they occur there are potentially more dangerous and far more serious personal and social repercussions, given the difficulty of recovering identity. And in large scale deployments, identity theft can be a bigger threat [16, 18].

As a consequence, it has been growing the concerns to identify the vulnerabilities of biometric authentication systems, aiming to create countermeasures that can be implemented to prevent identity theft. It is necessary to attend to: the registration of users, process management, rules of implementation(strong and well defined) and use of international standards to ensure scalability and interoperability of systems [19].

In addition, the administrators of the biometric systems must comply with the legal and moral principles inherent to the function they perform, given the information they handle. It is important that the legal aspects are guaranteed [10], so users will not feel disillusioned with the system and lose confidence.

2.4 Theories on Adoption of New Technologies

Information technologies have a radical impact on users, at work, and represent today an important role in organizations [20]. Usually the factors associated with the adoption of new technologies are related with individual beliefs, attitudes, subjective norm, and perceptions of behavioral control, perceived usefulness and its perceived ease of use. Human factors (personality, culture and society) have a significant bearing on how a technology is accepted, optimized and properly exploited in order to bring competitive advantage to organizations [20].

In this study we took particular attention to TAM[4], considered one of the most influential theories in the area of research in information systems regarding the adoption of information technologies [21]. It is considered one of the most powerful theories of this research area [22–24], applied in different domains, targeting different technologies, which indicates that it is a theory with some robustness [21].

3 Study Methodology

3.1 Measure Items and Sampling

The data collection instrument was developed adapting the constructs of TAM 3[25], and adoption studies of such technologies. We also consider James et al.[3], which showed that perceived need for security also influence user acceptance. Besides that, we consider items like reliability, security, privacy, and items to identify advantages of such systems in healthcare ecosystems.

3.2 Survey Distribution and Data Collection

The theoretical population of this study focuses on health services users. For this purpose it was selected the population of the Northeastern Hospital Center of Bragança in Portugal. The questionnaires were randomly distributed to urban and rural areas to include a comprehensive sample of the population.

The sample was obtained through stratified random sampling, originating homogeneous subgroups by sex and age. The final sample is made of simple random elements belonging to each of the homogeneous subgroups. We defined three age groups: 18 to 25 years; 26 to 50 years; and 51 to 60 years. The study population obtained from the CHNE database for the cluster aged 18 to 60 years corresponds to 83,577 users. The sample was calculated for a confidence interval of 95%, with a sampling error of 4%. After calculating the proportion of each age group by sex, we obtain the distribution of subjects as shown in Table 1.

Table 1. Distribution of subjects by age and sex

Male				Female			
18-25	**26-50**	**51-60**	**Total**	**18-25**	**26-50**	**51-60**	**Total**
6692	23224	9339	39255	7391	26718	10213	44322
	↓	↓	Sample calculation		↓	↓	
48	167	65	280	54	191	71	316

4 Preliminary Results

In a first analysis, and concerning the basic knowledge about biometric technologies, it appears that gender is not a relevant factor. As shown in Table 2, 390 of the respondents expressed knowledge about such technologies and even 62 say they use it in the workplace. Only 205 individuals, representing 34.5% of respondents, said that they have limited information, at least before completing the questionnaire.

Table 2. Perception of biometric authentication technologies for sex and age groups

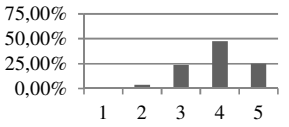
Knowledge, familiarity	Age	Sex	
		Male	**Female**
Insufficient	18-25	24	23
	26-50	46	58
	51-60	28	26
I am informed	18-25	24	28
	26-50	96	111
	51-60	32	37
Use at work	18-25	0	2
	26-50	25	22
	51-60	5	8

4.1 Perceived Usefulness

On this item, it appears that the majority have the perception of the usefulness of such technologies in any authentication process within a health service. From the analysis, we found that only 15.4% of respondents have no opinion, and the values of dissent have little meaning in the context. Table 3 illustrates the results obtained from the items used to measure the perception of usefulness.

Table 3. Frequencies for the perception of usefulness

		Frequency	%
1	Strongly Disagree	2	,3
2	Disagree	8	1,3
3	Undecided	92	15,4
4	Agree	330	55,4
5	Strongly Agree	164	27,5

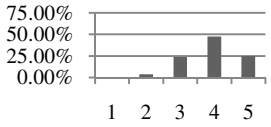


4.2 Perceived Ease of Use

Although in the previous construct (Table 3) it was found that the perceived utility is seen as an asset, the number of respondents having no opinion on the ease of use increased (Table 4). This phenomenon is related to the lack of experience in the use of such technologies. Except for the individuals that use in the workplace, representing 10.4%, the rest can only relate this to the use of fingerprints in the registration process of the citizen card, or through observation, particularly through television or even where public employees use biometric authentication systems to register the entry and exit of the respective service. One of the lessons to be drawn from the data presented in Table 4, points to the need to experiment the technology in order to form an opinion on the ease of use.

Table 4. Frequencies for the perception of ease of use

		Frequency	%
1	Strongly Disagree	3	,5
2	Disagree	21	3,5
3	Undecided	141	23,7
4	Agree	283	47,5
5	Strongly Agree	148	24,8

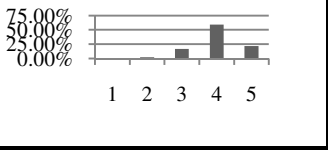


4.3 Behavioral Intention to Use

The previous two constructs reflect the general behavioral intention to use, assessing the positive or negative attitude in relation to the voluntary adoption of such technologies by an individual. Table 5 illustrates a scenario where 76,9% of respondents manifests intention of use these technologies. The reasons associated with the indecisiveness of 21.1% of respondents may be related to lack of information, lack of experimental technology, fear of innovative technologies, among other reasons that are not easily identifiable, unless a different interpretative study is conducted aiming to determine which may lead to the voluntary adoption. This will be the concern of this study at the next stage.

Table 5. Frequencies for behavioral intention to use

		Frequency	%
1	Strongly Disagree	1	,2
2	Disagree	11	1,8
3	Undecided	126	21,1
4	Agree	373	62,6
5	Strongly Agree	85	14,3

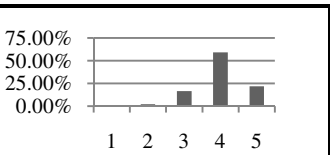


4.4 Reliability and Security Assurance Plus

This item is part of a set of questions that were posed to understand the vision of respondents in relation to biometric authentication technologies, regarding reliability and safety assurance, within health services. It appears that most of respondents see this type of technology as an enforcement of reliability and security in the health service authentication process, assuming that, for example, falsification of documents and exchange of identity is more difficult (Table 6). Other important impact is on limitation of some errors that occur in services (e.g. prescription errors).

Table 6. Reliability and security assurance

		Frequency	%
1	Strongly Disagree	2	,3
2	Disagree	14	2,3
3	Undecided	97	16,3
4	Agree	313	52,5
5	Strongly Agree	170	28,5

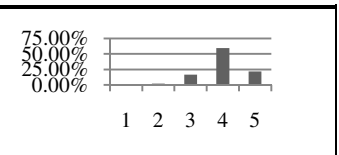


4.5 Privacy and Security

Issues related with privacy and security concerning healthcare units, are crucial when it comes to handling of privileged data. Related to this item, 85.7% (Table 7) of respondents agree that the data associated with identity should be adequate, relevant and not surplus to the objective for which it were collected and processed. The storage of biometric data should allow some sort of owner control and should be implemented following standards that help increase security while maintaining the privacy of users as a guarantee. It is clear the concern about issues related to this item, giving a clear image of fear about the process of implementing this type of technology.

Table 7. Privacy and Security

		Frequency	%
1	Strongly Disagree	-	-
2	Disagree	5	,8
3	Undecided	80	13,4
4	Agree	353	59,2
5	Strongly Agree	158	26,5

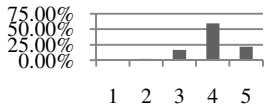


4.6 Advantages in Implementing Such Technologies

In Table 8, 81% of respondents consider the utilization of such technologies as an added value in health services, assuming that such systems provide more reliable and secure access, keeping it simpler and easier to use. If we compare tables 5, 6, 7 and 8, we found that the respondents foresee the use of such systems as advantageous as a means of authentication in a particular health service. Most of the respondents also found it more comfortable to use this type of systems in the future rather than carry various documents for authentication. We assume that this authentication method would satisfy users and health services providers in the same way.

Table 8. Advantages in the implementation of such systems in health

		Frequency	%
1	Strongly Disagree	1	,2
2	Disagree	13	2,2
3	Undecided	99	16,6
4	Agree	353	59,2
5	Strongly Agree	130	21,8



5 Conclusions and Work in Progress

Results clearly point that respondents are willing to adopt these technologies. The majority sees advantages in the process and reinforces expectations towards the willingness to adopt such systems. But we also found that some respondents are reluctant when it comes to provide an opinion on the items evaluated in various constructs.

It will be important in the next phase of this study to examine the arguments and strategies that can influence the adoption process. Data collection will be done through semi-structured interviews to allow a more open interaction with the interviewees, including a simulation using biometric authentication technologies. We expect to recognize behaviors influencing users more resistant to adopt this type of technology and understand how users' behavior can be influenced towards the voluntary adoption of biometric authentication systems. It is expected to provide a model to optimize the process of acceptance while keeping ethical principles.

Acknowledgments. This work was funded by FEDER through Programa Operacional Fatores de Competitividade – COMPETE, and by national founds through FCT – Fundação para a Ciência e Tecnologia, under project: FCOMP-01-0124-FEDER-022674.

References

1. Ho, G., Stephens, G., Jamieson, R.: Biometric Authentication Adoption Issues. In: ACIS 2003 Proceedings, pp. 1–12 (2003)
2. Savastano, M.: Biometrics: Modelling the Body. In: Tenth International Conference on Computer Modeling and Simulation uksim 2008, pp. 14–17 (2008)

3. James, T., Pirim, T., Boswell, K., Reithel, B.J., Barkhi, R.: Determining the Intention to Use Biometric Devices: An Application and Extension of the Technology Acceptance Model. Journal of Organizational and End User Computing 18, 1–24 (2006)
4. Davis, F.D.: A technology acceptance model for empirically testing new end-user information systems: Theory and results (1986), http://hdl.handle.net/1721.1/15192
5. Jain, A.K., Bolle, R.M., Pankanti, S.: Introduction to biometrics. In: Jain, A.K., Bolle, R., Pankanti, S. (eds.) Biometrics, pp. 1–41. Kluwer Academic Publishers (1999)
6. Liu, S., Silverman, M.: A practical guide to biometric security technology. IT Professional 3, 27–32 (2001)
7. Fairhurst, M.: Document identity, authentication and ownership: the future of biometric verification. In: Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003), pp. 1108–1116. Citeseer (2003)
8. Prabhakar, S., Jain, A.K., Pankanti, S.: Biometric recognition: Security and privacy concerns. IEEE Security & Privacy 1, 33–42 (2003)
9. Yun, Y.: The "123" of Biometric Technology. Synthesis Journal 2002 3, 83–96 (2002)
10. Zorkadis, V., Donos, P.O.: On biometrics-based authentication and identification from a privacy-protection perspective: Deriving privacy-enhancing requirements. Information Management & Computer Security 12, 125–137 (2004)
11. Bolle, R.M., Connell, J.H., Pankanti, S., Ratha, N.K., Senior, A.W.: Guide to biometrics. Springer, New York (2003)
12. Magalhães, P.S., Santos, H.D.: Biometria e autenticação. Actas da 4 Conferência da Associação Portuguesa de Sistemas de Informação, p. 17. APSI, Porto (2003)
13. Vashek, M., Řiha, Z.: Security of biometric authentication systems. In: International Conference on Computer Information Systems and IndustrialManagement Applications (CISIM), pp. 19–28 (2010)
14. Jain, A.K., Ross, A., Pankanti, S.: Biometrics: a tool for information security. IEEE Transactions on Information Forensics and Security 1, 125–143 (2006)
15. Watson, A.: Biometrics: easy to steal, hard to regain identity. Nature 449, 535–535 (2007)
16. Alterman, A.: "A piece of yourself": Ethical issues in biometric identification. Ethics and Information Technology 5, 139–150 (2003)
17. Rejman-Greene, M.: Privacy Issues in the Application of Biometrics: a European Perspective. In: Wayman, J., Jain, A.K., Maltoni, D., Maio, D. (eds.) Biometric Systems, pp. 335–359. Springer, London (2005)
18. Delaitre, S.: Risk management approach on identity theft in biometric systems context. In: 1st International Conference on Availability, Reliability and Security, pp. 20–22 (2006)
19. Dimitriadis, C.K., Polemi, D.: Risk Analysis of Biometric Systems. In: Fernández-Medina, E., Castro, J.C.H., Villalba, L.J.G. (eds.) Proceedings of the 2nd International Workshop on Security in Information Systems, pp. 23–32. INSTICC Press, Porto (2004)
20. Chan, S.L.: Information technology in business processes. Business Process Management Journal 6, 224–237 (2000)
21. Lee, Y., Kozar, K., Larsen, K.: The Technology Acceptance Model: Past, Present, and Future. Communications of the Association for Information Systems 12, 752–780 (2003)
22. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly 13, 319–340 (1989)
23. Taylor, S., Todd, P.A.: Understanding Information Technology Usage: A Test of Competing Models. Information Systems Research 6, 144–176 (1995)
24. Venkatesh, V., Davis, F.D.: A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 46, 186–204 (2000)
25. Venkatesh, V., Bala, H.: Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences 39, 273–315 (2008)

Cell Life: A Biology Game to Support Biology Classrooms

Teresa Futscher de Deus[1] and Pedro Faria Lopes[2]

[1] ISCTE – University Institute of Lisbon,
[2] ISCTE – University Institute of Lisbon & ADETTI-IUL
{mtmpf,pedro.lopes}@iscte.pt

Abstract. This paper describes a biology game headed to biology students. The game was developed to be a learning tool supporting biology classrooms. The players need to have the basic knowledge of a living cell and the game context involves: the cell constitution, the transport mechanisms, the metabolic reactions, the protein biosynthesis, the growth and the cell division. The game enables the students to dynamically interact with the cell environment and learn the concepts involved in the official programme (guidelines) of biology studied in the two final years of high school in the Portuguese educational system.

Keywords: multimedia learning, computer graphic visualization, biology classrooms support, computer game.

1 Introduction

From very early years, nowadays children like to play digital video games. It is called the game generation. There are several game categories, like action/adventure, sports, racing/driving, strategy, first person shooter, etc. There are also educational games in several areas of learning. These games bring dynamic to the learning process in basic knowledge (colors, words, and numbers) as well as in scientific areas such as mathematics, biology or languages [1, 2, 3, 4]. Learning a new domain, through conventional methods, requires study, work and teachers guidance. But digital games can be used as a complement of the classes offering an entertaining way to help learning and teaching. Youngsters are usually attracted by games that make their job easier in understanding the class material and amuse themselves.

Beyond the learning process, digital games promote the development of reasoning, social skills, self-confidence, memory training, attention, concentration and language by listening to speaking characters, read messages and/or objectives.

Games for health care purposes have been researched and applied with considerable success [5, 6]. Further research has shown that by using visual tools, students can perceive large amounts of information in a relatively short time and construct their own personal visualization of a process [7].

This paper starts with an introduction of educational digital games and definition of the project. In section 2, information it is presented about the impact of games on

Á. Rocha et al. (Eds.): *Advances in Information Systems and Technologies*, AISC 206, pp. 1149–1155.
DOI: 10.1007/978-3-642-36981-0_109

student learning and biology related multimedia applications are referenced as related work. In section 3, the design, the development and the game related information is explained. And, finally, section 4 presents the conclusions.

1.1 Project Definition[1]

The main goal is to build an educational biology game that supports the learning course material of classrooms that students learn in the Portuguese educational system in the two final years of high school. The information exposed in this game follows the biology manuals given in classrooms of "Curso Científico - Humanístico de Ciências e Tecnologias" [8, 9].

This project is a game prototype containing animations to make it fun and appealing, maintaining at the same time an educational goal with scientifically validated content. Using action and trivia quiz, the player has to complete several tasks, at a cellular level, about biology. It also allows players to embrace an experience of amusement, objective goals, knowledge and interesting facts. The storyline of the game is: "Control the energy cell level! Are you up to the challenge?".

The game scenario takes place in the inner life of an animal cell and the player has to control the energy of the cell in several molecular and biological processes.

2 Impact of Games on Student Learning

A major challenge to biology teachers is to teach molecular and cellular processes so that students can comprehend and understand their complexity: it is a difficult task. Students who viewed biology simulation animations report a higher level of interest in the course material, show greater memory retention and have overall better comprehension than students who did not view those animations [10]. Studies in high school and graduate-level courses proved that the use of multimedia content in teaching has a positive impact [11]. This can also be extended into other kind of multimedia such as games, as described next.

The opportunity to use digital games to create learning experiences is very interesting, allowing the acquisition and/or enhancement of knowledge through entertainment [12]. Games have changed the learners' cognitive skills in order to players to process a lot of information at the same time [13]. Teachers frequently use games to help students experience and have a deeper understanding of the material they study [14].

A game is an assessment [15]. Every time there is a problem to solve and if the player can't solve it, the player tries again until he passes the test. In school, this can be the most painful part, but in a game it can be fun because it is handled in a very different way [15]. This is one of the most positive aspects of using games as learning tools. In the next section, we present some multimedia applications and games that inspired our project.

[1] This project was made in the context of the Master Degree in Computer Science, Multimedia, in ISCTE - University Institute of Lisbon.

2.1 Biology Related Multimedia Applications and Games

There are several examples of multimedia applications and related material available for teaching and simulation purposes on biology. Some of them were used as inspirational work and base ideas to build this project.

In the web portal, Molecular Movies [16], with a collection of scientific animations on cell and molecular processes, one of the animations is from Harvard University, The Inner Life of a Cell [17], a fantastic award-winning animation conceived by biology students. This animation served as an example that it is possible to simulate accurately biology processes.

Cellcraft [18] is an educational and very fun game to learn about the cell's structure and how it survives in an hostile environment. The player can interact with the cell and its components while reading the function of each component and interact with the micro-environment. The results of Cellcraft showed that the gameplay was effective increasing low-order-thinking and changing students' science attitudes [19].

Control of the Cell Cycle [20] is another game about the different phases in ordinary cell division. The cell's tissue is destroyed and the player has to complete all tasks correctly to make new cells. The content of this game inspired the creation of mini-game 5 in our "Cell Life" project.

3 Game Design

The game is composed by five mini-games, the player has to maintain the cell energy and follow its life cycle until its growth and division. To complete the game successfully, it is necessary to have a background in biology theory. The requirements of the game are the following:

1. Be able to identify the components that constitute an animal cell, how they work and what their functions are;
2. Be familiarized with the plasma membrane and regulate what comes into and out of the cell through the existing transportation mechanisms;
3. Understand the metabolic reactions and processes involved in the cell respiration, specially the Aerobic Respiration;
4. Have the capacity to create a complementary RNA (Ribonucleic acid) and use the base pairs of nucleotides;
5. Comprehend protein biosynthesis, how a specific amino acid chain is produced and know the mRNA (messenger RNA) function;
6. And, finally, recognize the process of cell growth and division, called cell cycle, and be able to control it.

3.1 Game Development

Developed in Adobe Flash® [21], "Cell Life" is a 2D game with more than 10 different and interactive scenarios.

The game design and usability follows a collection of common known heuristics [22]. The artwork was drawn very similar to the diagrams and schematics of official biology manuals for purposes of memory retention and learning; the objectives of the game, following in great proximity the official programme contents, have to be complex, specific and concrete. The content of the game is distributed in five different mini-games.

The game initializes with the creation of a cell. In the first mini-game (Fig. 1. a), the player has to complete the task of getting the components that constitute a cell. The player has to collect the different components – nucleus, E.R., plasma membrane, lysosome, ribosome, centrosome, cytoskeleton, Golgi body and mitochondria – through clicks with the mouse. To make the level more challenging, there are also viruses the player should not collect – they make the cell energy and score go down.

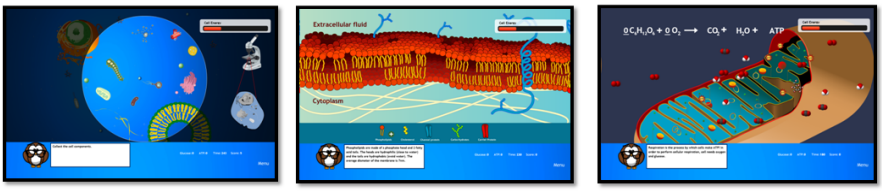


Fig. 1. a. left: mini-game 1; b. center: mini-game 2; c. right: mini-game 3

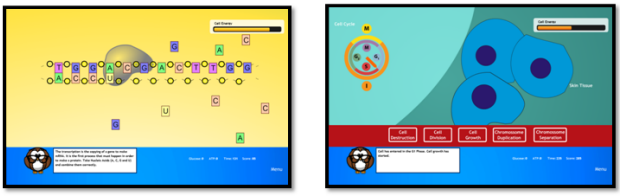


Fig. 2. a. left: mini-game 4; b. right: mini-game 5

In the second mini-game (Fig. 1. b), the membrane is damaged and, without the plasma membrane, the cells are unable to maintain homeostasis[2] and will die. All life depends on the player and what the player finds about the plasma membrane structure and its function. Through drag and drop, the player should reconstruct the membrane with the available supplies: phospholipids, cholesterol, channel proteins, carbohydrates and carrier proteins. Meanwhile, the different transport mechanisms are introduced and illustrated by the components animation.

After this, the third mini-game emerges (Fig. 1. c). In order to perform the cellular respiration, the player has to grab the necessary elements, which are glucose and oxygen, with a point and click. To add complexity to this level, other elements like oxygen, carbon dioxide and ATP (Adenosine TriPhosphat) are in the screen, but the player should not get them or they will break the energy flow and lose points.

The fourth mini-game is about the nucleic acids transcription and translation (Fig. 2. a). The player has to combine, in the correct order, the sequence with the base

[2] Property of the cell regulating its internal environment and maintaining stability.

pair of nucleotides. There is a RNA polymerase and a RNA chain. There are available nucleotides to combine correctly. After the transcription is done, the player has to take the result – the mRNA (messenger RNA) – to the proteins synthesis site, the ribosome. Next, the player has to take an amino acid and code it in order to perform the translation. The player has a mini codon[3] chart where one can see which amino acid combines with what tRNA (transfer RNA) and place it in the right anticodon. In this way, the ribosome reads the DNA (deoxyribonucleic acid) sequence until it gets a stop codon.

The last mini-game is the most interesting one: the player has to make a new cell (Fig. 2. b). The player has five options to perform the cell growth and division: cell destruction, cell division, cell growth, chromosome duplication and chromosome separation. The player also has available a diagram illustrating the cell cycle. And if the player chooses the correct option, the cycle goes along with the cell phase.

3.2 Sound

Today, music is mandatory in a game. The purpose of music is to attract enthusiasm and/or influence/change the heart rate of players and/or change their mood while they are feeling the game. The music for this game was carefully chosen from a music library [23]. There are different songs for the menu, for the mini-games and for the quizzes (that appear at the end of each mini-game) in order to motivate the player and capture his attention.

3.3 Content Evaluation

During the contents specification, the "Cell Life" game got the scientific contribution of two PhD biology researchers:

- Helena Deus [24], a PhD biologist and currently post-doctoral researcher at the Digital Enterprise Research Institute (DERI), in Galway, Ireland within the Health Care and Life Sciences Unit [25], provided useful guidelines for the scientific soundness of the content;
- Janet Iwasa [26], PhD Lecturer in Molecular Visualization, at Harvard Medical School [27], provided several useful papers about biological animation and simulations.

The final version of the game was submitted for evaluation to a medical doctor, the two PhD biology researchers that were involved since the beginning and two biology teachers that teach the official programme. All these specialized people approved the scientific content, the graphic design and the interaction in the games.

Furthermore, the biology teachers evaluated content and pedagogical issues, providing important feedback so that the game could be a useful learning tool. Finally the teachers found the project appropriate to biology students of 10^{th} and 11^{th} grade of the Portuguese educational system.

[3] Sequences of nucleotide triplets.

3.4 Tests and Discussion of the Results

Individual tests were conducted in an informal way to get quick feedback about the game interaction that solved first usability issues. A selection of 8 testers was chosen with ages between 16 and 23 years old: five of the feminine gender and three of the masculine gender. The testers with 16 years old were biology students for the specific audience. The others had biology knowledge or graduation.

During the tests, the testers showed a good progress regarding the interactivity. They quickly understood the goal of the game. The majority considered the game had a high difficulty level, which stands for a good outcome because the idea of learning is usually associated with difficulty and they considered that the game helped to visualize biological processes and learn about it. Due to time development constraints, it was not possible to get results about the concrete learning process while playing the game, but a pilot test is being planned to test for this purpose with learning students.

Since the game presents a lot of information, the testers also expressed the opinion that the game would work best as a teaching tool used in the classroom.

4 Conclusion

To create the game it was necessary to introduce novelty, create challenge, connect goals with the action of the player and bring meaning to the experience of playing in the area of animal cell biology.

The representation of the world game is responsible for the creation of personal visualization in the youngsters'/players' minds and, for this reason, the quality of the perceived information has to be solid. The user interface and the usability tests were fundamental to get the perception that this goal was achieved and that the project led to a hopefully successful learning tool.

The content development in terms of biology items to address and their correlation for problem solving were so extensive and complex that it took a huge amount of time to compile and validate the information and its gameplay dynamics. Without this, the game would be of very little interest. As it took many man-hours, a direct consequence was less time available for other tasks, mainly the specific learning process evaluation with classroom students, to be done in the near future.

The medical doctor thought the game complexity inline with real life study. The PhD biologists validated the content and thought it well integrated. The biology teachers considered the game very complete, highly adequate and in line with the official programme. The level of complexity achieved and the remarks from the specialists that tested the game lead us to be confident that this biology game can be a useful tool to support biology classrooms.

References

1. Jardim, D., Lopes, P.F., Alexandre, I.M.: Math4Kids – Aprender Conceitos de Matemática Brincando. In: Proceedings of the Artech 2010 Conference (2010)
2. Vidal, B.: Interactive simulations for health education: a case study on the cardiovascular disease in Portugal. Instituto Superior Técnico - Universidade Técnica de Lisboa (2011)

3. Deus, T.F., Marques, T., Lopes, P.F., Alexandre, I.M., Alexandre, P.: Learn English With Inspector Flops. In: Proceedings of the Artech 2012 Conference (2012)
4. Lopes, P.F., Moreira, M.V., Pereira, H.: Jogos Multimédia Educativos: Estratégias de Desenvolvimento. 3º Simpósio Internacional de Informática Educativa (2001)
5. Dutta-Bergman, M.J.: Primary sources of health information: comparisons in the domain of health attitudes, health cognitions, and health behaviors. Pub. Med. 16(3), 273–288 (2004)
6. Graf, D.L., Pratt, L.V., Hester, C.N., Short, K.R.: Playing Active Video Games Increases Energy Expenditure in Children. The Official Journal of the American Academy of Pediatrics 124(2), 534–540 (2009)
7. Linn, M.C., Songer, N.B., Eylon, B.S.: Shifts and convergences in science learning and instruction. Macmillan, Riverside (1996)
8. Silva, A.D., Gramaxo, F., Santos, M.E., Mesquita, A.F., Baldaia, L., Félix, J.M.S.: Universo de Vida Biologia 10º ano. Porto Editora, Porto (2004)
9. Silva, A.D., Gramaxo, F., Santos, M.E., Mesquita, A.F., Baldaia, L., Félix, J.M.S.: Universo de Vida Biologia 11º ano. Porto Editora, Porto (2004)
10. McClean, P., Johnson, C., Rogers, R., Daniels, L.: Molecular and Cellular Biology Animations: Development and Impact on Student Learning. Cell Biology Education 4, 169–179 (2005)
11. Thatcher, J.D.: Computer Animation and Improved Student Comprehension. JAOA - The Journal of the American Osteopathic Association 106(1) (2006)
12. Squire, K.: Games to Teach Project Year End Report. iCampus Committee, s.l., Cambridge (2001–2003)
13. Sabella, R.A.: GuardingKids.com: A Practical Guide to Keeping Kids Out of High-Tech Trouble. Educational Media Corporation, s.l. (2008)
14. The 2007 Horizon Report. The New Media Consortium, s.l. (2007)
15. Gee, J.P.: Big Thinkers: James Paul Gee on Grading with Games, Edutopia (2008)
16. MolecularMovies.org: A Portal to Cell & Molecular Animation (2012), http://molecularmovies.com
17. The Presidents and Fellows of Harvard College: The Inner Life of the Cell. Harvard University (2007), http://multimedia.mcb.harvard.edu/
18. Cellcraft: Exploring the Cell Through Computer Games. HASTAC/MacArthur Foundation Digital Media & Learning Competition (2009)
19. Dunlap, P., Pecore, J. L.: The Effects of Gaming on Middle and High School Biology Students' Content Knowledge and Attitudes toward Science. Wake Forest University, Winston-Salem NC (2009)
20. The Control of the Cell Cycle- Credits. Nobelprize.org (2012), http://www.nobelprize.org/educational/medicine/2001/credits.html
21. Adobe Systems Incorporated: Adobe Flash Professional CS5 (2012), http://adobe.com
22. Clanton, C.: An interpreted demonstration of computer game design. In: CHI 1998 Conference Summary on Human Factors in Computing Systems. ACM, New York (1998)
23. MacLeod, K.: Incompetech by Kevin MacLeod (2012), http://incompetech.com
24. Deus, H.: Helena Deus CV(2012), http://lenadeus.info
25. Digital Enterprise Research Institute (DERI), Health Care and Life Sciences Unit (2012), http://hcls.deri.ie/
26. Iwasa, J.: Janet Iwasa, Ph.D (2012), http://cellbio.med.harvard.edu/faculty/iwasa
27. Harvard Medical School (2012), http://cellbio.med.harvard.edu/

Author Index