Lecture Notes
in Business Information Processing 127

Series Editors

Witold Abramowicz
John Domingue
Krzysztof Węcel (Eds.)

Business Information Systems Workshops

BIS 2012 International Workshops
and Future Internet Symposium
Vilnius, Lithuania, May 21-23, 2012
Revised Papers

Springer

Volume Editors

Witold Abramowicz
Poznań University of Economics
Poznań, Poland
E-mail: w.abramowicz@kie.ue.poznan.pl

John Domingue
The Open University
Milton Keynes, UK
E-mail: j.b.domingue@open.ac.uk

Krzysztof Węcel
Poznań University of Economics
Poznań, Poland
E-mail: k.wecel@kie.ue.poznan.pl

ISSN 1865-1348 e-ISSN 1865-1356
ISBN 978-3-642-34227-1 e-ISBN 978-3-642-34228-8
DOI 10.1007/978-3-642-34228-8
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012949471

ACM Computing Classification (1998): J.1, H.3.5, H.4, I.2

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

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

Preface

BIS Workshops give researchers the possibility to share preliminary ideas and first experimental results and to discuss research hypotheses. It was initiated to focus attention on topics that have not yet found a place in canonical research, though they may experience fruitful discussion when confronted with a well-focused audience. Discussions held during presentations allow for improving the paper and preparing it for publication. From our experience, workshops are also a perfect instrument to create a community around very specific research topics, thus offering the opportunity to promote it.

Four workshops took place during BIS 2012. They covered such topics as applications and economics of knowledge-based technologies (AKTB), business and IT alignment (BITA), enterprise systems for higher education (ESHE), and formal semantics for future enterprise (FSFE). Additionally, the Future Internet Symposium was organized in conjunction with BIS 2012.

The 25 articles contained in this volume are an extended version of papers accepted for BIS workshop as well as for the Future Internet Symposium. There were 63 submissions for all mentioned events. Based on reviews, the respective Workshop Chairs accepted 24 papers in total, yielding an acceptance rate of 38%. There is also one invited paper hosted by the BITA workshop.

We would like to express our thanks to everyone who made the BIS 2012 workshops a success: all Workshop Chairs, members of the Workshop Program Committees, authors of submitted papers, invited speakers, and finally all workshop participants. We cordially invite you to visit the BIS website at http://bis.kie.ue.poznan.pl/ and to join us at future BIS conferences.

August 2012

Witold Abramowicz
John Domingue
Krzysztof Wecel

AKTB 2012 Workshop Chairs' Message

The 4th Workshop on Applications of Knowledge-Based Technologies in Business (AKTB 2012) was organized in conjunction with 15th International Conference on Business Information Systems (BIS 2012) held in Vilnius, Lithuania. It continued the successful series of AKTB workshops held in Poznan (2011, 2009) and Berlin (2010).

The mission continuously followed by the AKTB workshops is to invite scientists and practitioners to share their research expertise and knowledge, to provide new insights in modeling-advanced enterprise solutions, and to demonstrate experimental results of computational intelligence methods applied in various areas of business-related information processing.

AKTB 2012 kept the tradition of focusing on innovative and robust computational solutions validated by experimental research and based on the in-depth domain knowledge of various business areas.

The main topics of the workshop are concentrated on serving the needs of contemporary business enterprises by applying intelligent knowledge-based technologies:

- Advanced knowledge-based business information systems
- Computational intelligence for business (artificial neural networks, fuzzy systems, expert systems)
- Decision support systems in business enterprises, financial institutions, and e-management
- Knowledge- based models of data mining in business
- Business process and information requirements analysis
- Information technologies and software developments for business process modeling
- Agent-based and embedded systems in business applications
- Information systems in e-business, e-banking, and marketing
- Online trading by using evolution-based methods, neural networks, and rule-based systems
- Advanced computational approaches to portfolio optimization and selection
- Analysis of financial time series
- Estimations, modeling, algorithms of application of investment strategies in financial markets
- Advanced research and case studies of application computational methods in banking, insurance and credit risk evaluation, and company rating systems

In all, 18 articles were submitted to the AKTB2012 workshop. Each paper was evaluated in a double-blind review process by at least two independent reviewers of the Program Committee. The highest ranked nine articles were accepted for presentation during the conference and the second stage of reviewing before including them in the conference proceedings.

Each reviewer evaluated the quality of the article by taking into account the criteria of relevance of the article to the workshop topics, the adequacy of the

article title and the content, its originality and novelty, the coherence of the methodological background, the substantiation and validity of the conclusions, and the quality of presentation of the paper.

The following 24 outstanding researchers who represent prestigious scientific institutions from 15 countries joined the Program Committee:

Rajendra Akerkar	Western Norway Research Institute, Norway
Marie-Aude Aufaure	Ecole Centrale Paris, MAS Laboratory, France
Lia Bassa	Foundation for Information Society, Hungary
Tânia Bueno	Instituto de Governo Eletrônico, Brazil
Saulius Gudas	Vilnius University, Lithuania
Remigijus Gustas	Karlstad University, Sweden
Jaco Jordaan	Tshwane University of Technology, South Africa
Ferenc Kiss	Foundation for Information Society, Hungary
Irene Krebs	University of Technology Cottbus, Germany
Dalia Kriksciuniene	Vilnius University, Lithuania
Audrius Lopata	Vilnius University, Lithuania
Dhrupad Mathur	SPJCM, United Arab Emirates
Arturas Mazeika	Max Planck Institute of Informatics, Germany
Razvan Petrusel	Babes-Bolyai University, Romania
Tomas Pitner	Masaryk University, Czech Republik
Vytautas Rudzionis	Vilnius University, Lithuania
Virgilijus Sakalauskas	Vilnius University, Lithuania
Eric Schoop	TU Dresden, Germany
Mikhail Simonov	ISMB, Italy
Darijus Strasunskas	NTNU, Norway
Réz Tamas	HandSoft Ltd., Hungary
Margaret Tan	Nanyang Technological University, Singapore
Danuta Zakrzewska	Technical University, Lodz, Poland
Du Zhang	California State University, USA

We appreciate the expertise and quality of work of the Program Committee members, whose reviews provided in-depth analysis of the submitted research works and highlighted valuable insights for the authors. The high standards followed by the reviewers ensured the high quality of the workshop event, excellent presentations, intensive scientific discussions, and added value to the post-conference workshop proceedings.

We would like to express our gratitude for the joint input to the success of AKTB 2012 to all authors of submitted papers, members of the Program Committee, Vilnius University and Department of Information Systems of the Poznan University of Economics, and to acknowledge the outstanding efforts of the Organizing Committee of the 15th International Conference (BIS 2012).

June 2012

Virgilijus Sakalauskas
Dalia Kriksciuniene

BITA 2012 Workshop Chairs' Message

A contemporary challenge for enterprises is to keep up with the pace of changing business demands imposed on them in different ways. There is today an obvious demand for continuous improvement and alignment in enterprises, but unfortunately many organizations do not have proper instruments (methods, patterns, best practices etc.) to achieve this. Enterprise modeling and business process management are two areas belonging to a tradition where the mission is to improve business practice and business and IT alignment. In this tradition the alignment process is usually manifested in taking a business from one state into another improved state, i.e., a transformation of the business and its supporting IT into something that is regarded as better. A challenge in business and IT alignment is to move beyond a restraining focus with one tradition or technology. We need to be aware of and be able to deal with a number of dimensions of enterprise architecture and their relations in order to create alignment. Examples of such dimensions are: organizational structures, strategies, business models, work practices, processes, and IS/IT structures. Among the concepts that deserve special attention in this context is IT governance. An effective IT governance aligns IT investments with overall business priorities, determines who makes the IT decisions, and assigns accountability for the outcomes. There are ordinarily three governance mechanisms that an enterprise needs to have in place, (1) decision-making structures, (2) alignment processes, and (3) formal communications.

BITA 2012 was the third workshop on this subject following an event in 2011, which was located at the 14th International Conference on Business Information Systems in Poznan. The workshop aimed to bring together people who have a strong interest in business and IT alignment, and encouraged a broad understanding of possible approaches and solutions for business and IT alignment, including IT governance subjects. We invited researchers and practitioners from both industry and academia to submit original results of their completed or ongoing projects. Specific focus was on practices of business and IT alignment, i.e., on case study and experiences papers.

The workshop received ten submissions. The Program Committee selected five submissions. We had an invited speaker at the workshop, Wim Van Grembergen, who gave a speech with the title "Enterprise Governance of IT and the COBIT 5 Framework".

We thank all members of the Program Committee, authors, and local organizers for their efforts and support.

July 2012

Ulf Seigerroth
Kurt Sandkuhl

Organization

Chairs

Ulf Seigerroth	Jönköping International Business School, Sweden
Kurt Sandkuhl	Jönköping University, Sweden

Program Committee

Ulf Seigerroth	Jönköping University, Sweden
Kurt Sandkuhl	Rostock University, Germany
Mathias Ekstedt	KTH, Sweden
Remigijus Gustas	Karlstad University, Sweden
Stijn Hoppenbrouwers	Radboud University Nijmegen, The Netherlands
Håvard Jørgensen	Commitment AS, Norway
Marite Kirikova	Riga Technical University, Latvia
John Krogstie	Norwegian University of Science and Technology, Norway
Andreas Opdahl	University of Bergen, Norway
Erik Proper	Radboud University Nijmegen, The Netherlands
Khaled Gaaloul	Radboud University Nijmegen, The Netherlands
Wolfgang Molnar	Radboud University Nijmegen, The Netherlands
Janis Stirna	KTH, Sweden
Anne Persson	Skövde University, Sweden
Mikael Lind	Viktoria Institute, Sweden
Tomasz Kaczmarek	Poznan University of Economics, Poland
Nikolay Shilov	SPIIRAS, Russia
Klas Gäre	Jönköping International Business School, Sweden
Vivian Vimarlund	Jönköping International Business School, Sweden
Steven De Haes	University of Antwerpen, Belgium
Wim Van Grembergen	University of Antwerpen, Belgium

ESHE 2012 Workshop Chairs' Message

Enterprise systems, such as ERP systems, for example, are an important part of the IT infrastructure of today's enterprises worldwide. Hence, the integration of these systems plays an important role in the area of educational environments. Because of their high potential of illustration, visualization, and simulation of business and decision-making processes for learners, enterprise systems have a high significance for future pedagogic innovations within higher education. The primary challenge is to tap this potential in the education of enterprise systems. currently, there are different approaches (e.g., case studies) to train students in these systems. Furthermore, there are also technical solutions that support the learning process in the field of enterprise systems.

After the First ESHE Workshop held in Poznan, Poland, in conjunction with the 12^{th} International Conference on Business Information Systems (BIS) in April 2009 and the Second ESHE Workshop hosted in Valencia, Spain, in conjunction with the Second International Conference on Computer Supported Education (CSEDU) in April 2010, the Third Workshop on ESHE took place in conjunction with INFORMATIK 2011 in Berlin, Germany, in October 2011. The current 4^{th} ESHE Workshop was held in Vilnius, Lithuania, in conjunction with the 15^{th} International Conference on Business Information Systems (BIS) in May 2012.

The 4^{th} Workshop of Enterprise Systems in Higher Education (ESHE) 2012 aimed to continue the innovative concepts in the field of the creation of new and adequate learning environments and also at possible technical solutions, which offer technology-enhanced learning environments for teachers and learners. The main objective of the workshop is to bring researchers and practitioners together to explore the issues and challenges related to enterprise systems in higher education.

The following submissions were accepted for publication by the Program Committee of the ESHE 2012 Workshop:

" Serious and Simulation Games – A Definition Approach," by Bastian Kurbjuhn

"Cloud-Based Online Learning Platforms," by Stefan Bensch

"Repository-Based ERP Case Studies: A Study About Chances and Benefits of Agile Case Study Development," by Hans-Jürgen Scheruhn, Stefan Weidner, and Sandro Sicorello

"Interactive Learning – Teaching IT Project Management Using an Explorative Role Play," by Stefan Weidner

We would like to thank all contributing authors for their submissions and also take the chance to thank the members of our Program Committee for the successful cooperation the last four years.

May 2012

Jorge Marx Gómez
Klaus Turowski
Liane Haak
Dirk Peters

Organization

Program Committee

Maria Manuela Cruz Cunha	Polytechnic Institute of Cavado and Ave, Portugal
Danco Davcev	Sts. Cyril and Methodius University, FYROM
Sabine Graf	Athabasca University, Canada
Gustaf Juell-Skielse	Royal Institute of Technology, Sweden
Helmut Krcmar	Technische Universität München, Germany
Johan Magnusson	University of Gothenburg, Sweden
João Luís de Miranda	Instituto Politécnico de Portalegre, Portugal
João Eduardo Quintela Varajão	University of Trás-os-Montes e Alto Douro, Portugal
Karin Rebmann	Carl von Ossietzky Universität Oldenburg, Germany
Hans-Jürgen Scheruhn	Hochschule Harz, Germany
Stefan Smolnik	EBS Business School, Germany
Frank Teuteberg	Universität Osnabrück, Germany
Stefan Weidner	Otto-von-Guericke-Universität Magdeburg, Germany
Holger Wittges	Technische Universität München, Germany

FSFE 2012 Workshop Chairs' Message

The Formal Semantics for the Future Enterprise Workshop was created with the main goal of bringing together researchers to share their knowledge and experience in the formal aspects of the future enterprise. Nowadays, many companies undergo a lot of changes, at all levels. In this respect, the workshop seeks to create, promote, and spread the formal application fundamentals that will enable the transition to those future enterprises. During this workshop, we discussed and tackled advanced formal approaches in enterprise-related semantics and enterprise virtualization, the transition from data to knowledge, as well as the issues related to representing and storing the knowledge of the enterprise, using semantics and ontologies. Another important aspect is collaboration, with a focus on the formal aspects of communication technology and security. Also, the workshop focused on the formal approach to business processes and decision-making automation.

The Program Committee selected four submission for presentation at the workshop. The paper of P. Bertoli et al. discusses SMART, a solution designed to support operative business process management under the premises of fragmented software systems and complex dynamics depending on non-controllable heterogeneous events and actions. SMART offers a flexible and agile modeling approach, where external services, information sources, and proprietary support systems can be meaningfully and smoothly integrated according to business logics, so as to result in operative consoles that can be used as effective coordination tools. The paper of L.G. Cretu discusses the enterprise architecture (EA) approach as a means to manage the enterprise transformation complexity. The author has developed a methodology to transform EA descriptions made in plain natural language into machine-readable EA models using Semantic Web technologies. The paper of A. Butoi et al. proposes a protocol to provide data storage confidentiality, in the context of virtualization to cloud resources. The protocol splits enterprise data into small encrypted chunks and disperses them into data storage volumes, handling separately the read/write operations. The paper of C.I. Muntean et al. explores the meaning of Twitter hashtags, in order to retrieve connections that might exist between different hashtags and their textual representation and to grasp their semantics through the main topics they occur with.

We thank all members of the Program Committee, authors, and local organizers for their efforts and support.

Gheorghe Cosmin Silaghi
Razvan Petrusel
Robert Buchmann

Future Internet Symposium 2012 Chairs' Message

While we cannot be sure about the concrete shape of the Future Internet (FI), we can expect with confidence that it will continue touching more and more aspects of our lives and our environment, and that it will provide novel features that are often denoted as smart.

The smart applications in the FI are expected to bring people, things, and data together in innovative ways. The smart applications will come from diverse areas such as entertainment, health, utilities, transport, mobility, and logistics. The FI will encompass the "Internet of Content," the "Social Internet," the "Internet of Services," the "Internet of Things," and other concepts that are being proposed. However, a simple analysis of the anticipated smart applications and FI concepts reveals a set of conflicting requirements. Hence, shaping the FI to serve such diverse needs and capabilities is a challenging task.

The 2012 Future Internet Symposium focused on the requirements of novel FI application domains and the problems of the current Internet in addressing those requirements, and it proposed solutions to these problems and the steps necessary to make the smart FI a reality. Here, we publish two selected peer-reviewed articles presented at the symposium. The first one, by Anna Fensel et al., introduces a conceptual framework that applies a number of technologies on the Web and in social media to support effective communication of e-offerings. The second one, by Alex Oberhauser et al., addresses recommendations of mobile services to users with the help of knowledge representation and information-retrieval techniques over combined social Web data sources.

In addition to the presentations of peer-reviewed papers, FIS 2012 had the honor of welcoming two keynote speakers: Stephan Haller of SAP Research in Zürich (Switzerland) and Darius Šilingas of No Magic Europe in Kaunas, Lithuania. Stephan discussed the requirements for the realization of the Internet of Things (IoT). His main point was that for the deployment of the IoT to be manageable, we need advanced tools for the modeling and discovery of processes and services. Darius, in turn, illustrated how the main principle of model-driven development is to raise the level of abstraction in modeling to keep models understandable and maintainable. He stressed that models need to be constrained in size, by employing modularity and decomoposition, to be able to capture usefully the growing complexity of the FI.

August 2012

Witold Abramowicz
John Domingue
Krzysztof Wecel

Table of Contents

AKTB Workshop

BITA Workshop

ESHE Workshop

FSFE Workshop

Future Internet Symposium

Building a Social Recommender System by Harvesting Social Relationships and Trust Scores between Users

Daniel Mican, Loredana Mocean, and Nicolae Tomai

Babes-Bolyai University, Dept. of Business Information Systems, Str. Theodor Mihali 58-60, 400599, Cluj-Napoca, Romania
{Daniel.Mican,Loredana.Mocean,Nicolae.Tomai}@econ.ubbcluj.ro

Abstract. Recommender systems were created to guide the user in a personalized way to interesting resources and to help users cope with the problem of information overload. A system's ability to adapt to the users' needs is based on gathering user-generated collective intelligence. In this paper, we present WSNRS, the system proposed for recommending content within social networks. The main goal of the system is to identify and filter the recently published valuable resources while taking into account the interactions and the relationships the user has within social structures. The interactions are logged and aggregated in order to determine the trust scores between users. Using the scores obtained, one can identify the types of relationship established between users; the scores will then be integrated into an adaptive global model used for recommending resources. Our approach presents several advantages over classic CF-based approaches and content-based recommendations regarding cold start, scalability and serendipitous recommendations. We will illustrate this with a case study that we made using data provided by the implementation of the system in a real online social network.

Keywords: Recommender systems, Social networks, Trust in social networks, Social networks recommender system, Collective intelligence.

1 Introduction

The new technologies and concepts that Web 2.0 brings to web applications and the ever-increasing expansion of the Internet have brought about the emergence of a large number of online communities and social networks based on shared interests. The fast rise in popularity of social media applications has drawn the attention of hundreds of millions of users worldwide. At the same time, it has created new challenges for the developers and researchers dealing with online social applications.

This paper presents the architecture of the social recommender system WSNRS (Wise Social Network Recommender System), the method used for implementing it and a case study. The proposed system logs, aggregates and uses the collective intelligence obtained as a result of the interaction of users with each other and with the content in order to determine the trust scores between users. Using these trust scores, one can identify the most trustworthy users in the social structures of a

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 1–12, 2012.

network. The trust one user has in another is an important piece of information that helps us to improve the recommendation algorithms by efficiently identifying the valuable resources, i.e. the resources that determined the user to get involved. In our system, the user's involvement requires making comments, adding the resource to favorites, rating it, clicking and recommending it on social media websites.

The proposed system fits into the Web 2.0 social applications where users publish, share and interact with the published content. The content is published at an amazing speed and it is very difficult for a user to read and stay up-to-date with all the resources. Therefore, WSNRS aims at identifying the quality of the user-generated content within a social network. If a resource is of low quality, it will receive a low rating. If the resource has a high quality, it will be promoted and recommended to users. Thus, users will stay up-to-date with all the new resources that are relevant and that have a high quality without wasting time.

The proposed recommender system aims to be a guiding service, to provide personalized content and to adapt to the users' needs within a social network. It recommends the most recently published resources taking into account the explicit and implicit relationships that the user has within social structures. The advantage of the proposed system is allowing quality resources to become visible without the risk of losing them for ever in a databank. Thus, the system we propose manages to filter and rank the recently published resources in a specified time period. The system also fulfills the task of moderating resources automatically. This is of ultimate importance because in a system where the number of resources generated by users is very high, manual moderation becomes virtually impossible.

Our approach presents several advantages over classic Collaborative Filtering (CF)-based approaches and content-based recommendation. As the saying goes, birds of a feather flock together and we believe that in a group of friends/sympathizers there are common tastes. In our opinion, this approach does not suffer from the "cold-start" problem due to the fact that the user doesn't have to rate the content in order to receive recommendations. To send recommendations, only the social structure to which the user belongs needs to be considered. Moreover, we don't have to take into account all the elements of the clusters of users and resources, which is a major advantage when considering the scalability problem. We will only consider the sub-clusters of friends and resources published/preferred by them. In order to find a solution to the problem of providing "serendipitous recommendations", a problem which occurs in content-based recommender systems, we will include not only the texts published by friends but also the texts that they rated favorably.

The recently published resources will be recommended within the social structure to which the author belongs. If the resource is liked, it can go through many social structures and go viral in a short time. The purpose of belonging to a social network is keeping in touch with friends, following various celebrities, interacting and meeting new people. This is accomplished by sharing resources within the network. Depending on the shared resources, preferences and interactions, people get to know each other better, discover something new and meet other people.

This paper is structured as follows: Section 2 presents the problem of content recommendation, the approaches and difficulties encountered. Section 3 presents WSNRS, the system proposed for recommending content in social networks. In

Section 4, we will present a case study that will illustrate the way in which the relationships between users that are based on trust can be used in recommending resources. Section 5 contains the related work. In the end, we will present our conclusions and future development directions.

2 Content Recommendation in Social Networks Considering the Collective Intelligence

Recommender systems were created to guide the user in a personalized way to interesting objects and to help users cope with the problem of information overload. A system's ability to adapt to the users' needs depending on how they navigate a website, to predict these needs and to provide a navigation path is particularly important in a web application. A social recommender system can accomplish this just by analyzing the data from collective intelligence extracted from the social network. According to Ramos [11], collective intelligence allows the capturing of the collective behavior of entities that interact with their environment in order to develop functional global patterns. All this data is captured and used to acquire knowledge and to create profiles, reputation models or recommender systems. In order to be efficient, the created models have to adapt to the collective dynamics of the environment in order to deal with possible unforeseen circumstances and changes.

According to [10], there are two approaches for recommending content: content-based approach and collaborative filtering (CF). The content-based recommendation system attempts to recommend resources similar to those a user expressed interest for in the past. The main flaw of this system is its inability to provide "serendipitous recommendations". Serendipitous recommendations are recommendations that are not in the context of users history of visited resources and have the ability to surprise in a positive and pleasant way. The collaborative filtering system takes into account reviews, ratings or explicit voting; its main role is to identify users with preferences similar to those of the current user in order to recommend resources that they prefer. Leaving its advantages aside, the collaborative filtering system faces problems [5], such as "cold start", sparsity, scalability and the malicious ratings. Cold start problem applies to new published resources that anyone in the community has not rated yet.

In social networks, the users are in the centre of the universe. This universe is based on technologies that allow the remodeling of applications with regard to structure, design and usefulness. In recent years, we have witnessed an online social revolution where the users started forming relationships, publishing, aggregating and storing content. These interactions have led to the building of online communities and social networks based on shared interests. Social networks [14] are being made of a finite group of people and the relationships they establish among them. The existence of information on the established relationships is a key feature of social networks. Analyzing the data generated by users within social networks has several practical applications that can be used to develop personalization and recommendation systems, and identification of trends within social systems.

Hogg [4] believes that users with similar preferences and characteristics tend to associate in social networks. This allows the improvement of collaborative filtering systems by considering users' preferences based on social distances and increased

efficiency of the reputation mechanisms. The approach based on information provided by user communities recommends resources by taking into account their friends' preferences. This method is inspired by the proverbs: "Birds of a feather flock together" and "Tell me who your friends are and I'll tell you who you are".

3 Wise Social Network Recommender System (WSNRS)

In this chapter we will briefly describe the formal approach, the architecture and the way to implement WSNRS. On the one hand we will describe the method to calculate trust among users and on the other hand we will present the deduction of the follower's implicit status. In this article we define a follower as a user who is an enthusiastic supporter of another user in regard to his or her ideas or belief.

3.1 The Formal Approach

In our approach any existent relation between two users is represented by a vector that contains all interactions the users had over time.

Definition: A user-user relationship is represented by a vector: $V = (IdU_i \rightarrow IdU_j, LT, NF, NRw, NR, NC, NCk, TSt, T) \in M_{1x9}(R)$. In this formula, l represents the number of existing links among users and v_t the vector corresponding to each relationship for $t=1,l$. In what follows we will describe the vector attributes *V:*

- $IdU_i \rightarrow IdU_j$: represents the existing connection from user *i* to user *j*.
- *LT*: represents the type of the existing connection between user *i* and user *j*, $LT \in \{0,1,2\}$. The first element represents a connection of low intensity, namely below average. The second element represent an explicit connection, and the third one represents a connection of high intensity, above average.
- *NF*: is the number of favorite resources published by the user *j*, $NF \in [1, \infty)$.
- *NRw*: represents the number of positive reviews granted by user *i* to the user *j* or to the resources published by the user *j*, $NRw \in [1, \infty)$.
- *NR*: represents the number of user's recommendations or of the published resources within the social media sites, $NR \in [1, \infty)$.
- *NC*: represents the number of involvements resulted from the comments posted on the user's profile page or on the resource's pages published by the user *j*, $NC \in [1, \infty)$.
- *NCk*: represents the number of clicks to see the user *j* profile or the resources published by him, $NCk \in [1, \infty)$.
- *TSt*: represents the timestamp of last interaction initiated from user *i* to user *j*.
- *T*: represents the calculated trust score offered to user *j* by user *i*, $T \in [0,1]$.

We will use the same formal approach with regard to the relationships that are going to establish between users and content. The vectors containing the interactions among users are stored in tables of the form: *user_user*, whereas those containing the interactions among users and pages will be stored in a table of the form: *page_user*. The structure of the two tables is illustrated in Fig.1.

In order to calculate the trust score T_i for $IdU_i \rightarrow IdU_j$, we have normalized the following attributes of the vector v_i: *NF*, *NRw*, *NR*, *NC*, *NCk,* so that the attributes contain numeric values in the interval [0, 1].

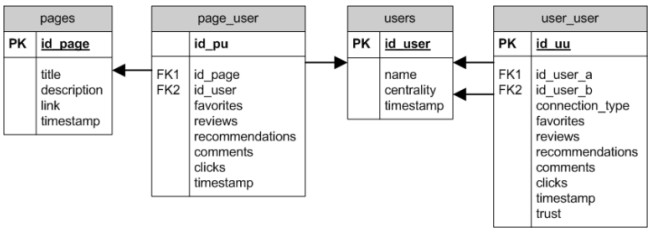


Fig. 1. Database structure for the proposed recommender system

The values have been normalized with the "min-max normalization" [2]. This is a method that performs a linear transformation of input data *v* to *v'* of an attribute *NF,* so that these data can be found in the interval [0, 1]. The formula adapted for the *NF* attribute of the vector *V,* is shown below:

$$v' \frac{v - \min(NF)}{\max(NF) - \min(NF)} * (1 - 0) + 0 \quad (1)$$

The formula for the trust score T_i for $IdU_i \rightarrow IdU_j$, is illustrated below. The trust score is calculated for $\forall v_i$, $i = 1,l$.

$$T_i = \frac{1}{5}\left(\frac{vNF_i}{\max(NF)} + \frac{vNRw_i}{\max(NRw)} + \frac{vNR_i}{\max(NR)} + \frac{vNC_i}{\max(NC)} + \frac{vNCk_i}{\max(NCk)}\right) \quad (2)$$

On the basis of T_i, that is calculated for all $IdU_i \rightarrow IdU_j$ links, all follower implicit links will be deducted. Where $T_i \geq \sum_1^l T_i / l$ all links $IdU_i \rightarrow IdU_j$ will have $LT = 2$. This means that the trust score is sufficiently high to transform the user in follower.

3.2 The Architecture and the Implementation of WSNRS

The main aim of the system is to identify and recommend the valuable resources that have been published recently. In order to achieve this goal, the system takes into account the collective intelligence resulted from the user's interaction within the network. Furthermore, the collective intelligence is collected and quantified with the "Data collection module". The interactions among users are managed with the "User-user interactions management module", whereas those among users and resources are managed with the "User-content interactions management module". The architecture of the system is illustrated in Fig.2.

The user-user interactions management module identifies all interactions that take place among users and among users and resources. All these interactions are managed by this module and then are stored in the *user_user* table. If two users interact for the first time, the system creates a new vector v_i that contains the link between $IdU_i \rightarrow IdU_j$.

However, if the users have interacted in the past the corresponding vector will be updated. If the IdU_i user will add a positive review to IdU_j profile or to a published resource, NRw with a unity will be incremented, $NRw = NRw + 1$. The same happens with the NF, NC and NCk attributes.

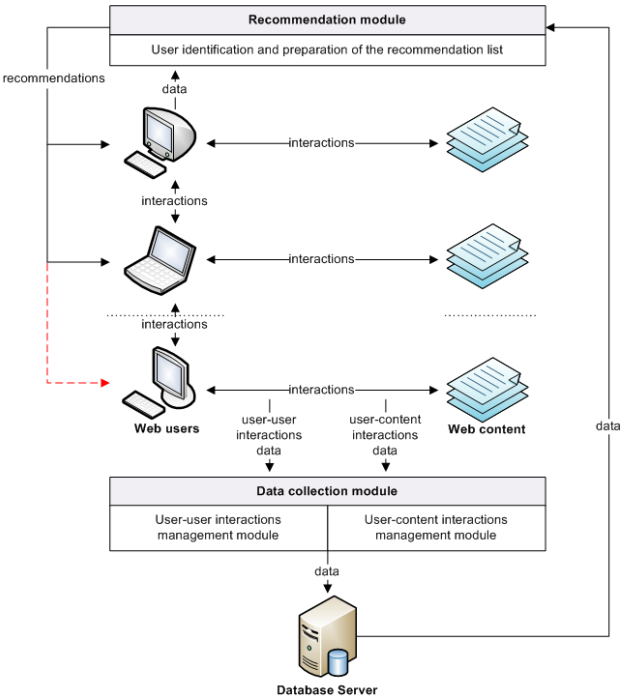


Fig. 2. WSNRS. The proposed architecture

With regard to the *NR's* attribute, the update is more complicated compared to the others. This is due to the fact that a user can give a "like" on the facebook button and recommend the text and after a period of time to withdraw the "like". The system succeeds in capturing not only the "like" action, in which case $NR = NR + 1$, but also the "dislike" action, in which case $NR = NR - 1$. The same is true for the module "User-user interactions management module".

Besides logging and updating the interactions, in this module we have integrated an algorithm that calculates the trust scores T_i, $\forall v_i$, $i = 1,l$, and updates the type of existing connection LT among users. If T_i is over average, results $LT = 2$, meaning that the user is an implicit follower, if not, than $LT = 0$. The time calculation of this algorithm increases with the number of existing links within the social network. Due to this fact it is recommended to run at predetermined periods of time depending on the number of interactions that have been realized in a certain period of time.

The recommendation module is designed to identify the current user on the basis of IdU_i and to prepare for this a recommendation's list. The recommended resources will include not only the texts published by the trusted users but also those that have been

favorably rated. In order to achieve this goal, the algorithm will identify all connections that $IdU_i \rightarrow IdU_j$ has with $LT > 0$. Furthermore, the algorithm will extract user's IdU_j Ids and will return the most recent text published by this users. Likewise, it will also search for the most recent interactions of these users with the published texts and will return the texts list they have interacted with. The final recommendations represent an aggregation of the two lists mentioned above.

4 Case Study

As soon as a user accesses a website and until he/she closes the navigation, he/she performs a series of interactions. In the following section, we will present a case study in order to see how these interactions can be quantified and aggregated to create an adaptive global model. We will begin by presenting the framework used for the case study. Then, we will identify the relationships that users establish and we will determine the level of trust that one user has in another user. And finally, we will illustrate the way in which these relationships based on trust can be used in recommending resources within the network.

In order to achieve this goal, we implemented the proposed recommender system on the portal Intelepciune.ro. With an average number of 300,000 unique visitors for month, this portal is on the top list of Romanian cultural websites, being supported by a large community of users. In this portal, we created a section for a literary circle where any user can publish his own literary works. Due to the fact that Intelepciune.ro was built using Web 2.0 principles, it is possible for users to interact both with the published resources and with other users. Users can comment, rate, add to favorites or recommend any user profile and resource on social media channels.

For the purpose of creating this case study, we collected data provided by the social network that developed in the literary circle section on Intelepciune.ro. During the time period in which we collected data, 511 out of a total of 6,723 registered users established interactions. We quantified a total number of 16,620 direct interactions established among users and indirect interactions established through published resources. The number of interactions was calculated by taking into account the ratings, additions to favorites, recommendations, comments or clicks. After analyzing the interactions, 1,388 links between users were identified. 6.23% of these links are follower relationships expressed explicitly and 18.62% were inferred implicitly based on the proposed algorithm. The relationships established between users also reveal the structure of the social network and of the interest groups respectively. This structure can be subsequently analyzed.

The visual representation of the structure of the social network can be seen in the sociogram illustrated in Fig. 3. This is represented by an oriented graph that reveals the strongest links within the network. Due to the limited space and in order for the graph to be intelligible, we depicted only the first 50 links using arcs. 41 vertices were involved in these links. Thus, one can see the graphic representation of the links, the users involved in these links and the level of trust between the users. Also emphasized are the structures of the established groups that allow the identification of interest groups, leaders and isolated individuals.

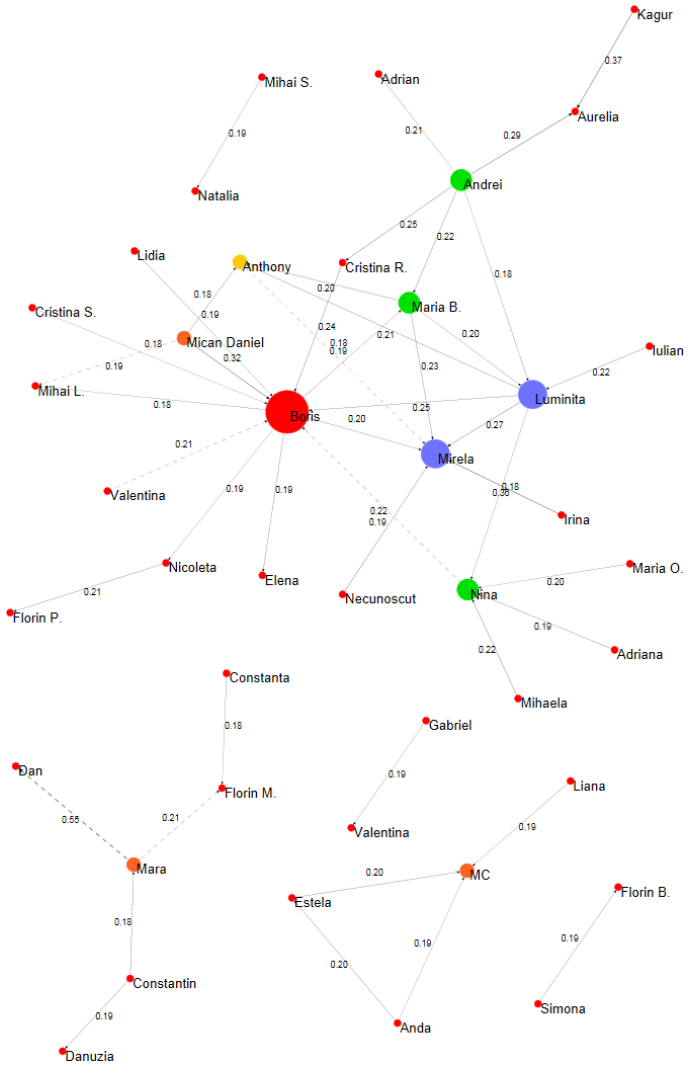


Fig. 3. The interest groups, types of relations and trust scores between users

In the graph, each vertex represents a user and the size of the vertex corresponds to its centrality within the network. The arcs between vertices represent the links that the users established and are labeled with a trust coefficient that was calculated by applying the suggested algorithm. The resulted values were normalized in order to be in the range [0.00, 1.00]. The links were ranked based on how close they were to the best result, which tended to reach value 1. If there is an explicit connection between two vertices, the arc is represented by a continuous arrow and if the connection is implicit, by a broken arrow.

Upon analyzing the graph, we notice that the vertex with the most central position within the network is Boris. This vertex has high levels of trust provided by the

neighbouring vertices. The vertex Mican Daniel has the highest level of explicit trust in the vertex Boris and its value is 0.32. This is followed by the trust level of the vertex Luminita with a value of 0.25 and Cristina R.'s with a value of 0.24 respectively. As to the implicit trust, we can see that the vertex Nina provides a level of 0.22. The vertex Boris is that much more important considering that the neighbouring vertices also receive high centrality values and high trust levels.

Concerning the structure of the social network, we notice that three interest groups emerged at the top. The first group is created around Boris, the central vertex with the leading role. The vertices Mirela, Luminita, Nina and Maria B. are also noticeable within this group. In the second interest group, MC is the central vertex and in the third group, the vertex Mara plays an important role. Moreover, isolated links were identified between the vertices Natalia and Mihai S., Valentina and Gabriel, Simona and Florin B.

What comes as a surprise for us is the fact that the strongest connection within the network is represented by an implicit connection that was discovered using our algorithm. The connection has a trust level of 0.55 and is directed from the vertex Mara to the vertex Dan. This is followed by explicit connections with trust levels that are equal to 0.37, 0.36 and 0.32. The connections are directed from the vertex Kagur to Aurelia, from Luminita to Nina and from Mican Daniel to Boris, respectively.

Below, we will provide an example in order to see how the trust relationships established between users can be used to recommend resources within the network.

Fig. 4. List of recommendations for user Mican Daniel

In order to receive recommendations, a user must have the role of a follower, regardless of the fact that it was expressed explicitly or inferred implicitly. The newest resources will be recommended in the current version of the proposed recommendation algorithm. The recommended resources will be resources that have been published or positively rated by the users that the current user is following. In Fig. 4, one can see the list of recommended resources for the user Mican Daniel.

The list of recommendations contains ten resources arranged after the date when they were published. For the purpose of ensuring the transparency of the recommendation, the author of the resource is specified under the name of the resource along with the type of the relationship and the trust level that the current author has in the author of the resource. Concerning the fifth resource, we can see that it was published by Mihai L. The resource was recommended because the user Mican Daniel is an implicit fan of the user Mihai L. This is represented by the digit 2 enclosed in brackets and the level of trust in Mihai L. We can also see that resource number 5 was positively appreciated by Maria B. and Elena. Mican Daniel is an explicit fan of these users, which is evidenced by enclosing digit 1 in brackets. Resource number 6 was recommended due to the fact that it was liked by Boris and Mican Daniel is an explicit follower of Boris'.

The recommender system proposed was implemented and can be accessed online at the address: http://www.cenaclu.intelepciune.ro. The system can be accessed by creating a new user account or by logging into the account created by us for the purpose of testing the system. The user name and password for this account is WSNRS. The list of recommendations containing additional information on recommendations, the user who published/liked the resource, the type of follower and the trust level can be accessed at: www.cenaclu.intelepciune.ro/wsnrs.php?id=6731.

5 Related Work

As is well known, people usually ask their friends for recommendations when they want to see a film, read a book, listen to a song, find a place to spend their free time or consult any online resource. Therefore, we could assume that users place a higher degree of trust in recommendations made by their friends than in those made by people unknown to them who could share similar preferences. After studying the specialized literature, we reached the conclusion that the above-mentioned hypothesis tends to be true in most cases. The researches published in [12] include a study that drew comparisons between the recommendations made by six popular recommender systems and recommendations made by friends. Results showed that the user's friends constantly provided better recommendations than the analyzed recommender systems.

The use of existing links between users in social network to make recommendations surpasses the classic CF-based approach. This conclusion was reached after an extensive study was made in a social network of over one thousand participants and was published in [1]. Moreover, a collaborative recommender system was created in [6] that takes into account both ratings and social relationships illustrated using a tripartite graph involving users, resources and tags. These experiments also proved that the suggested social system performs better than the CF system based on determining similarity using the Pearson correlation coefficient.

A trust-based recommendation approach was suggested in [3]. This approach can recommend trustworthy agents in a social network. The initial premise was that an agent can be recommended as long as the agent's neighbors display a high level of trust in him/her. The system is modeled by using graphs and by defining similarity measures and determining trust values. To make recommendations, both the structure of the network and the trust values associated with the links are considered. A trust-based

recommendation approach is suggested in [13] as well. In this system, agents use the social network to find resources and the established trust relationships to filter them. The system is self-organizing and takes into account the density of the network, the eclectic nature of the preferences and the knowledge sparseness. Furthermore, a study was conducted to examine how the dynamics of trust among agents influences the performance of the system by comparing it to a frequency-based system. Making recommendations based on what many individuals do was approached in [9]. This approach creates a Decision Data Model that depicts the data used and how it was derived by a large number of decision makers during the decision making process.

The consumers' preferences and interests can be identified on the basis of the knowledge extracted from the interests of their neighbours in the social network. Tad Hogg [4] identified the requirements for the network architecture and the correlation among preferences which contribute to the identification of large groups of users with similar interests. This knowledge improves the usefulness of deriving products from the social network. Based on the information gathered from a few users in the network, the consumers may be interested in these products. Therefore, the final estimations help design combined services for complementary products.

6 Conclusions

In this paper, we have illustrated and described the architecture of WSNRS which stands for the recommender system proposed for social networks. In order to accomplish our goal, we have shown the way in which the system logs and aggregates collective intelligence in order to calculate the trust scores between users. The trust among users is revealed by the information that helps to identify the types of relations established among users and define the social structures. Based on the scores and the types of relations established, we have developed an adaptive global model that enables recommendation and filtering of valuable resources recently published. On the one hand, the recommended resources take into account the relationships that the user has within social structures and on the other hand, they adapt themselves to the collective dynamics of the environment. Likewise, a very important aspect is that the system fulfills the task of moderating resources automatically, which is particularly important in the social network where manual moderation is impossible.

The results of the case study presented in this paper, coming from a real online WSNRS implementation, have revealed the way in which the computed trust scores allows to discover implicit connection between users that was not explicitly expressed and to define the interest groups. Furthermore, we have shown that on the basis of the aggregated trust scores, one can identify group leaders, isolated individuals and the most trustworthy existing users in social structure. Moreover, we have demonstrated that our approach brings a number of advantages over traditional CF-based approaches and content-based recommendation. Recommendations can be realized according to the user's social structure, without having to rate the content. According to our approach, the similarity with other users is revealed by the computed trust scores. This aspect, successfully solves both scalability and "cold start" problems. Due to the fact that the provided recommendations include not only the text published by the friends but also the favorable rated texts, the user will be positively surprised by the resources and receive "serendipitous recommendations".

In future works, we plan to study the way in which trust can be transferred from user to user and from user to resources. Further research with respect to [8] will focus on improving the model by integrating a recommender system with the help of tags. Thus, experts and trusted content identification in different categories of interest will be possible. Likewise, we aim to achieve hybridization through aggregating with the system that was proposed in [7] and that is based on extracting association rules from navigation sessions. Therefore, in future works, we will concentrate on developing a new hybrid system that aims to solve the problems encountered in traditional recommender systems.

References

1. Groh, G., Ehmig, C.: Recommendations in taste related domains: collaborative filtering vs. social filtering. In: GROUP 2007: Proceedings of the 2007 International ACM Conference on Supporting Group Work, pp. 127–136. ACM, New York (2007)
2. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques, 3rd edn. Morgan Kaufmann Publishers Inc., San Francisco (2011)
3. Hang, C.W., Singh, M.P.: Trust-Based Recommendation Based on Graph Similarity. In: The 13th AAMAS Workshop on Trust in Agent Societies, Trust (2010)
4. Hogg, T.: Inferring preference correlations from social networks. J. Electronic Commerce Research and Applications 9(1), 29–37 (2010)
5. Kim, H.-N., Ji, A.-T., Jo, G.-S.: Enhanced Prediction Algorithm for Item-Based Collaborative Filtering Recommendation. In: Bauknecht, K., Pröll, B., Werthner, H. (eds.) EC-Web 2006. LNCS, vol. 4082, pp. 41–50. Springer, Heidelberg (2006)
6. Konstas, I., Stathopoulos, V., Jose, J.M.: On social networks and collaborative recommendation. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 195–202. ACM, New York (2009)
7. Mican, D., Tomai, N.: Association-Rules-Based Recommender System for Personalization in Adaptive Web-Based Applications. In: Daniel, F., Facca, F.M. (eds.) ICWE 2010. LNCS, vol. 6385, pp. 85–90. Springer, Heidelberg (2010)
8. Mican, D., Tomai, N.: Web 2.0 and Collaborative Tagging. In: Proceedings of the 2010 Fifth International Conference on Internet and Web Applications and Services (ICIW 2010), pp. 519–524. IEEE Computer Society, Washington, DC (2010)
9. Petrusel, R., Stanciu, P.L.: Making Recommendations for Decision Processes Based on Aggregated Decision Data Models. In: Abramowicz, W., Kriksciuniene, D., Sakalauskas, V. (eds.) BIS 2012. LNBIP, vol. 117, pp. 272–283. Springer, Heidelberg (2012)
10. Ramezani, M., Bergman, L., Thompson, R., Burke, R., Mobasher, B.: Selecting and Applying Recommendation Technology. In: Proceedings of International Workshop on Recommendation and Collaboration, in Conjunction with IUI 2008, Canaria, Spain (2008)
11. Ramos, V., Fernandes, C., Rosa, A.: Social Cognitive Maps, Swarm Collective Perception and Distributed Search on Dynamic Landscapes. J. of NMNCSE (2005)
12. Sinha, R.R., Swearingen, K.: Comparing recommendations made by online systems and friends. In: DELOS Workshop: Personalisation and Rec. Systems in Digital Libraries (2001)
13. Walter, F.E., Battiston, S., Schweitzer, F.: A model of a trust-based recommendation system on a social network. J. Autonomous Agents and Multi-Ag. Sys. 16(1), 57–74 (2008)
14. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences). Cambridge University Press, New York (1994)

Decision Tree Algorithms: Integration of Domain Knowledge for Data Mining

Auksė Stravinskienė, Saulius Gudas, and Aiste Dabrilaite

Kaunas Faculty of Humanities of Vilnius universitety, Muitinės st. 8,
Kaunas, LT-44280, Lithuania
{aukse.stravinskiene,saulius.gudas}@khf.vu.lt,
aiste.dabrilaite@stud.vu.lt

Abstract. The paper deals with integration of enterprise (domain) knowledge with data mining techniques. The decision tree algorithms CART, ID3, C4.5, CHAID are analyzed in the aspect of relations with enterprise model attributes (i.e. domain knowledge). It is expected, that integration of an enterprise knowledge base in to data mining techniques will improve the data analysis process. The detailed analysis of the modified decision tree algorithm C4.5, integrated with domain knowledge is described. The alignment of business domain attributes used in the decision tree algorithms with the attributes of EM components is defined.

Keywords: Data mining, domain knowledge, decision tree algorithms, modified algorithm C4.5.

1 Introduction

Today the organization's most valuable asset is the ability to save and to use the domain knowledge for enhancing enterprise activities. Intellectual tools must have the ability to use the accumulated knowledge. It is important to the organization's management. There is a need of a good data processing tool that will help the organization to carry out faster decision-making process. Various data mining algorithms, techniques and tools help to organize and classify data and to discover the relationship between them, to make decisions and achieve optimal results.

The knowledge-based theory of the organization considers knowledge as the most strategically significant resource of the organization [1, 2, 3]. We haven't intelligent data mining tools that used knowledge to control or improve data mining process. Organizations require more systematic and formalized methods for knowledge management implementation. Most researchers describe ontologies-based knowledge extraction models. T. Mavroudakis and H. Karanikas [4] proposed ontologies based knowledge extraction from documents, images, and language resources. The main principle is formed semantically rich metadata collections, containing data mining tools are removed from the knowledge sources. Pan Ding, Shen Jun-Yi proposes ontologies based data mining process, which is constantly added to the field of knowledge [5].

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 13–24, 2012.

The idea is integration of domain knowledge (enterprise model – EM) with data mining techniques, to use the analyst's knowledge and improve the data mining process.

The enterprise model (EM) usual includes the components (knowledge items) as follows: **EM components** = {Process, Process attributes, Function, Function attributes, Event, Event attributes, Actor, Actor attributes, Business Rule, Business Rule attributes, Information Flow, Information Flow attributes, Material Flow, Material Flow attributes, Objective, Objective attributes}.

Table 1 shows the most common used algorithms in data mining. The aim of the experiment is to compare features of C.4.5 algorithm, to identify the differences between results and to show a significant algorithm attributes to data mining process.

Table 1. Summary of data mining algorithms

Algorithm	Description
Naive Bayes	Bayesian method based on probability theory can evaluate the probability of training data. This method cannot be used when the data are random or decision based on certain criteria.
Nearest neighbor method)	The nearest neighbor method is used for classification tasks. The principle of this algorithm is that the problem is not solved by grouping items according to the set or the analysis found in the rules.
Regression	Regression algorithm is used for prediction. It predicted a future of real numerical values, identified trends.
Classification	Classification algorithm to divide the data into groups-clusters, which are not known in advance. This algorithm is intended to apply to groups.
Neutral Network	Neural networks are used for forecasting and classification tasks. This non-linear training from the data. The network is composed of several layers: input, output and one or more hidden layers.
Decisions Trees	Decision-tree algorithm is a hierarchical structure, similar to a tree, a data mining method.
Genetic Algorithm	Genetic algorithms are based on the principle of operation in accordance with the biological evolution of the template. They are used to find the approximate problem solution.
Association rules	Association algorithm is used in large databases, repositories to find regularities of objects or events in groups, statistical rules.
Hybrid algorithm	New data mining models of direction - the various hybrid algorithms. This is two or more algorithms for compounds that are used to exploit the best features of each algorithm, data mining process.

It is pending, that integration of an EM (i.e. knowledge base) in to data mining activities will improve the business management and data analysis process. Each of the components of the Enterprise model has attributes (3). This paper aims to show that these attributes can be used for data mining algorithms. Consequently the establishment of a knowledge base to manage the data mining process.

The detailed analysis of the decision tree algorithm C.4.5 integrated with domain knowledge and attributes is described.

2 Overview of the Main Decision Tree Algorithms

These are the most commonly used decision tree algorithms, which help to classify the data, to identify interactions between different groups of variables, to predict organizational perspective or gain the fastest decisions. Decision tree algorithm is a predictive model with a hierarchical structure similar to the tree and currently used in data mining. This algorithm has several advantages [1, 2, 6]:

- Important decisions may be generated on the basis of the experts' experience and the results of the research.
- It uses a "White box" model, i.e. results are easily interpreted and mathematically described.
- Can be combined with other data mining algorithms.
- Is characterized by simplicity, clarity, and imagery.

The Disadvantages of decision tree are as follows: the tree may provide incorrect results, if the constructed tree has too many branches. It is difficult to choose the appropriate selection steps [1, 2, 6].

The components of decision tree algorithms associated with domain knowledge items are as follows: **Decision tree components** = {Root; Root attributes; Node; Nodes attributes; Formulas; Formulas attributes; Decision rules; Decision rules attributes; Decision answers}.

Figure 1 shows common components of decision tree algorithms (i.e. knowledge items - business domain components and attributes). The identifier K denotes attributes of components of decision tree algorithm that are interpreted as specific knowledge items related with enterprise domain.

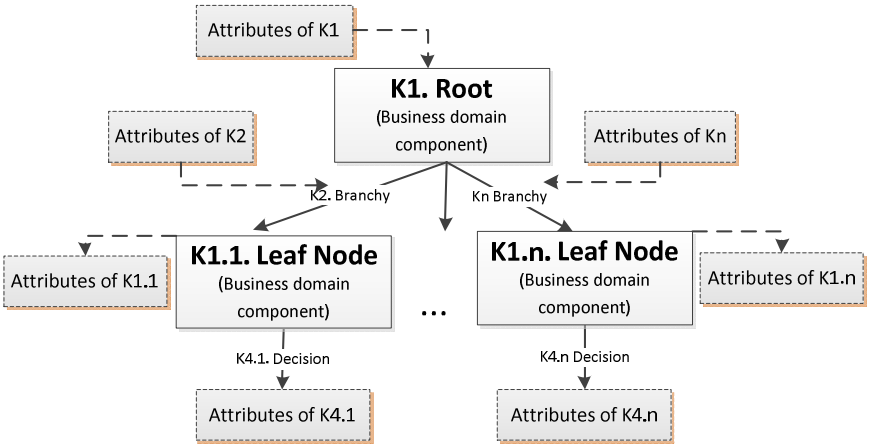


Fig. 1. Knowledge items of decision tree algorithms

The alignment of business domain attributes used in the decision tree algorithms with the attributes of EM components is defined in the Table 4.

2.1 Classification and Regression Tree - CART

CART - the decision trees are known as classification and regression trees. These trees are used to predict the variable "y", which coincident with the parameter vector "f" (p_1, p_2, ..., p_n).

If the variable "y" [7]:

1) Defines the class or category (dependent set of integer or natural numbers), such tree is called a classification tree. The variable "y" can have two values: YES or NO.
2) Defines dependent large numbers of real numbers, this tree is called a regression tree. The regression task can be change the classification task by dividing the variable y in the interval.

Sarmishtha, Saradhi Varma and other authors describe the CART methodology, where the CART algorithm uses a binary recursive partitioning. Each parent node on the tree is split exactly into two child nodes. This is recursive process, because the process can be repeated for each child node as a parent [8].

CART uses the Gini index of diversity as a distribution criterion. RF (C_j, S) denotes the relative frequency of S, which belongs to class C_j. Gini index is defined as [7]:

$$I_{gini}(S) = 1 - \sum_{j=1}^{x} RF(C_{j,}S)^2 \tag{1}$$

The Mean Squared Error, or MSE, is defined as [7]:

$$MSE = E_e\left[\sum_{j=1}^{x}(C_j(e) - P_j(e))^2\right] \tag{2}$$

The components of decision tree algorithm CART associated with presented domain knowledge items are as follows: **CART components** = {Root; Root attributes; Node; Nodes attributes; Formulas; Formulas attributes (Gini index attributes, MSE attributes); Decision rules; Decision rules attributes}(see Table 4).

2.2 Iterative Dichotomiser3 - ID3

ID3 (Iterative Dichotomiser3) is a decision tree induction algorithm, which uses the benefits of information as a quality function for the selected features [6]. Benefits of information are defined as a data set for entropy T before and after the difference when it split to the character of A. InfoGain (A) = E (t) - E (Y | A). The entropy E (T) is defined as [8]:

$$E(T) = \sum_{i=1}^{l}\left(-\frac{|T_{c_i}|}{|T|}.log\frac{|T_{c_i}|}{|T|}\right) \tag{3}$$

The data set of measurements with the objective character S is calculated:

$$Entropy(s) = -\sum_{i=1}^{c} p_i log_2 p_i \tag{4}$$

Information Gain measure, which calculates the entropy decrease can be used:

$$Gain(S, A) = Entropy(S) - \sum_{v \in A} \frac{|S_v|}{|S|} Entropy(S_v) \tag{5}$$

Entropy (S) is the average amount of information that identifies the attribute of the S class. The attribute of benefit is calculated by the formula [8]:

$$Info_A(S) = \sum_{i=1}^{n} \frac{|S_i|}{|S|} Entropy(S_i) \tag{6}$$

The components of decision tree algorithm ID3 associated with domain knowledge items are as follows: **ID3 components** = {Root; Root attributes; Node; Nodes attributes; Formulas; Formulas attributes (Entropy attributes, Gain attributes, Benefit attribute); Decision rules; Decision rules attributes}(see Table 4).

2.3 C4.5

C4.5 is an algorithm that can create a decision tree [6]. These decision trees can be used for classification, and for this reason are often called statistical classification. C4.5 creates decision trees from training data set in the same way as the ID3 algorithm using information entropy concept. Attribute with the highest normalized information gain is the most adequate to choose.

Attribute of gain ratio is calculated by the formula [9]:

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)} \tag{7}$$

The entropy is calculated by the formula [7]:

$$Entropy(S) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{8}$$

Information Gain is calculated by formula [7]:

$$Gain(A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} Entropy(S_i) \tag{9}$$

Information Split is defined as [9]:

$$SplitInfo(A) = -\sum_{i=1}^{n} \frac{|S_i|}{|S|} \log_i \frac{|S_i|}{|S|} \tag{10}$$

The components of decision tree algorithm C.4.5 associated with domain knowledge items are as follows: **C.4.5 components** = {Root; Root attributes; Nodes; Nodes attributes; Formulas; Formulas attributes (Entropy attributes, Gain attributes, SplitInfo attributes; Decision rules; Decision rules attributes}(see Table 4).

2.4 CHi-squared Automatic Interaction Detection - CHAID

CHAID is a type of decision tree algorithms, which can be used for prediction, classification and for detection of interaction between variables. CHAID means Chi-squared automatic interaction detector. Using this technique it is very simple to establish relationships between a dependent variable [6]. CHAID algorithm is a very efficient statistical technique for dividing or tree growth. When used as a test of statistical importance criteria, CHAID evaluates all of the predicted values [10].
Chi-squared value is calculated using the formula:

$$ChiSquared\ value = \sum_{all\ cells} \frac{(Observed\ Cell\ Count - Expected\ Cell\ Count)^2}{Expected\ Cell\ Count} \quad (11)$$

The components of decision tree algorithm CHAID associated with domain knowledge items are as follows: **CHAID components** = {Root; Root attributes; Nodes; Nodes attributes; Formulas; Formulas attributes (Chi-squared attribute); Decision rules; Decision rules attributes}(see Table 4).

3 The Approach to Knowledge-Based Data Mining

The knowledge-based theory describes systematic and formalized methods for knowledge management and this knowledge importance for organization [3]. Most researchers propose ontologies-based knowledge extraction models from documents, images, and language resources. The main idea is to collect semantically rich metadata bases (i.e. knowledge bases), containing with data mining tools and to use these knowledge for managing data mining process.

We selected and analyzed C.4.5 algorithm, which have important knowledge items for data mining process management and have conjunction with Enterprise model. Table 2 shows interaction of knowledge items of EM and the elements of algorithm.

Table 2. Integration of enterprise knowledge with algorithm C4.5

Designation	Components of EM	The components of algorithm
K1, K2	Process, Objective	Root; Root attributes; Node; Branchy; Branches attribute; Node attributes.
K1.1, K1.n	Function	Formulas; Formulas attributes (Entropy attributes; Gain attributes; SplitInfo attributes and etc.)
K3	Rules	Classification Rules; Rules attributes
K4	Result	Pruning attributes

4 Knowledge Extraction of Decision Tree Algorithm

This chapter analyses C.4.5 algorithm and the importance for data mining process. The detailed analysis of the decision tree algorithm C4.5 integration with domain knowledge is described and presented.

The domain knowledge is described by enterprise model (EM) components as follows: **EM components** = {Process, Process attributes, Function, Function attributes, Event, Event attributes, Actor, Actor attributes, Business Rule, Business Rule attributes, Information Flow, Information Flow attributes, Material Flow, Material Flow attributes, Objective, Objective attributes}.

4.1 About Experiment: Integration of Algorithm C4.5 with Domain Knowledge

The experiment objective was to verify the results of the decision tree algorithm C4.5 of different software products, using the same data. The obtained models were compared each other on the most important features of generating tree: tree depth, classification rules, branching of decision tree. The experiment was carried out with three software products: Orange; Pentaho/ Weka; RapidMiner [11, 12, 13]. The experiment was carried out in the following stages: C4.5 algorithm was used; the results were displayed on a decision tree structure, interpreted and compared with each other. The experiment used the bank customer data: id, age, region, married, children, car, save_acct, current_acct, mortgage, pep.

The experiment used 14450 records, where the data was sorted by individual ownership Plan (PEP) records. Algorithm C4.5 was used on experimental data and the results are presented graphically in the Fig. 2 (a, b, c). The data processing steps are described in the Table 3.

Table 3. The data processing steps

Orange	Weka	RapidMiner
1) Scan Data 2) Calculate attribute of Information Gain 3) Set the attributes of splitting 4) Be reviewed the depth of the tree, branching, pruning criteria 5) Progress splitting 6) Plot the graph of a tree	1) Scan Data 2) Calculate attribute of Information Gain 3) Set the attributes of splitting 4) Progress splitting 5) Plot the graph of a decision tree	1) Scan Data 2) Set the depth of the tree, branching, pruning criteria 3) progress splitting 4) Plot the graph of a decision tree

The main difference between the Orange and RapidMiner is that the Orange program calculates the attribute of Information Gain and set attributes of splitting.

All three programs use different calculation techniques. The similarity between the programs is that they all use the same minimum number of leaf and pruning parameters. However, these programs have different settings for partition criteria. Figure 2 (a, b, c) shows the results of the experiment.

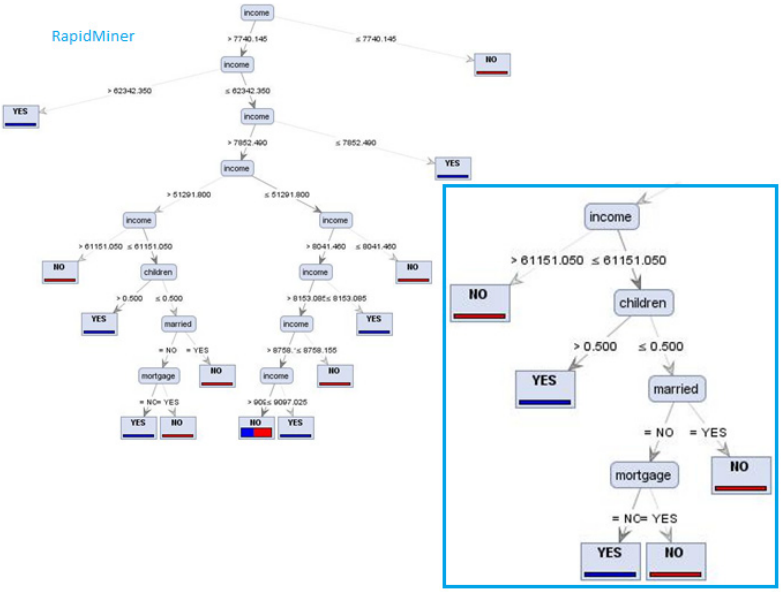


Fig. 2a. Results of RapidMiner tool (the main feature: tree depth)

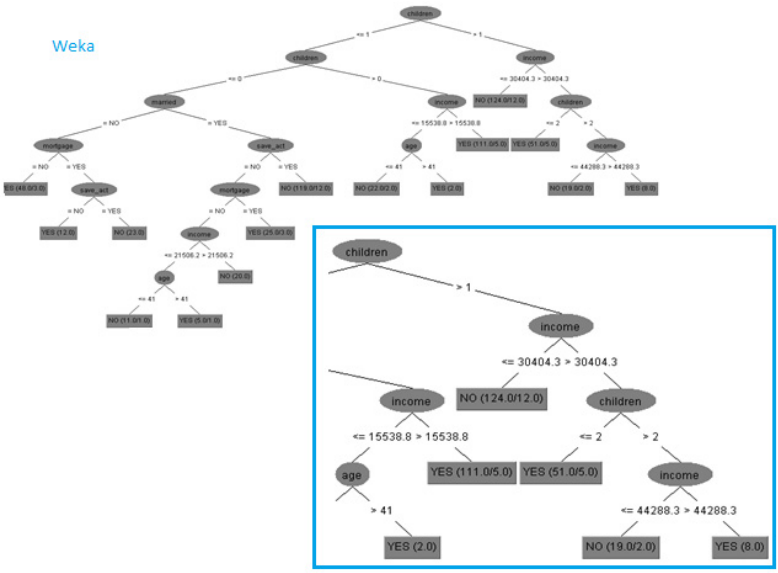


Fig. 2b. Results of Weka tool (the main feature: identified classification rules and branchy tree)

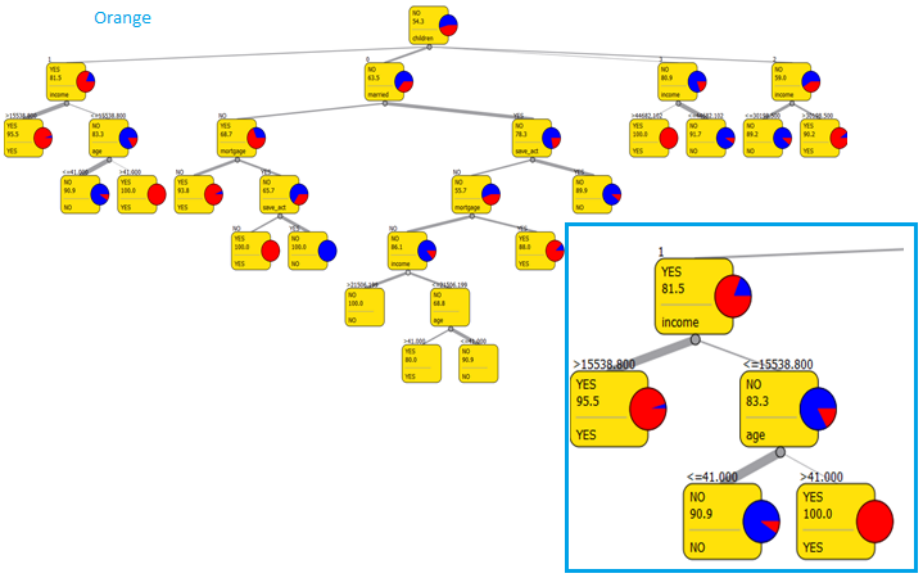


Fig. 2c. Results of Orange tool (the main feature: very large and branchy tree; optimal tree depth)

The results have shown that using the same data and the same C4.5 algorithm, the tool generates different results. RapidMiner is a generated tree with the highest classification depth and provides maximum outputs. The data is grouped into small groups. WEKA and Orange have the advantage: they display the results (leaves) in more detail. RapidMiner results (leaves) are viewed only in one colour (100%), it proves that the bank's data has been classified quite accurately. WEKA and Orange begins classification according to the number of children, and classify the income of RapidMiner. Most data are sorted according to customers' income. It can be assumed that this record has the highest importance for classification. This enables the banker to know customers who purchased the personal Equity Plan.

The experiment results have shown that features of the definite decision tree model (the classification rules, pruning criteria, attributes of splitting) depends on the tool. Let us suppose that these features of decision tree could be controlled by using domain knowledge, so, it could lead up for better results of decision tree generation.

4.2 Relation of the Algorithm C.4.5 with Enterprise Model

The knowledge-based theory of the organization considers knowledge as the most strategically significant resource of the organization [1].

Important role in the knowledge-based view of the organization in sense that information systems can be used to synthesize and expedite large-scale knowledge management [1, 2]. Organizations require more systematic and formalized methods for knowledge management implementation.

S. Gudas and A. Lopata in [3] described Enterprise Model (EM) based on the structure of Elementary management cycle (EMC), this EM is valid for Knowledge-based systems modeling. The main principle of this model's structure is feedback interactions of Enterprise processes and Enterprise management functions.

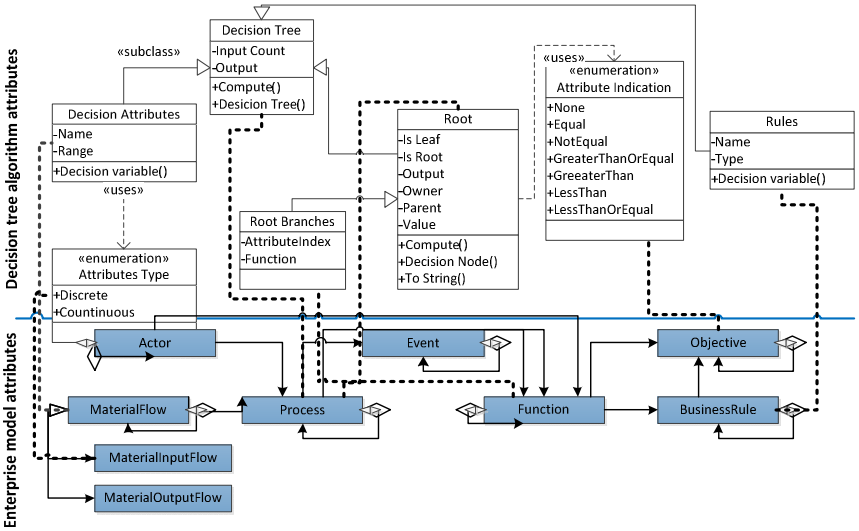


Fig. 3. Enterprise model components relations with decision tree meta-model

Figure 3 presents integrated model of knowledge items connections of Enterprise model, which consists of the following components: process, Function, Goals, Interpretation rules (IN), Realization rules (RE), Decision making rules, Data processing rules (DP) and Interactions (I). Model of decision tree algorithm can be split with Enterprise model, it shows dotted line. This principle model can be used for knowledge base, which implements development of data mining process.

Table 4 shows conjunction of enterprise model and decision tree algorithm attributes.

Table 4. Conjunction of enterprise knowledge attributes with algorithm C4.5 attributes

Knowledge items – components of EM	Enterprise model attributes	The elements of algorithm
Process	Process; Process attributes	Root; Root attributes; Branches; Branches attributes
Function Objective	Function; Function attributes Objective; Objective attributes, DP; DP attributes	Formulas; Nodes; Formulas attributes (Entropy attributes, Gain attributes, SplitInfo attributes, Gain Ratio attributes)
Rules	Decision making rules; Decision making rules attributes	Rules; Rules attributes
Result	IN; RE; IN and RE attributes	Pruning attributes

5 Conclusions

It is expected, that integration of an enterprise knowledge base in to data mining techniques will improve the data analysis process. The approach to integration of attributes of decision tree algorithms and Enterprise model is described. The Enterprise Model is structured and explicitly retained in knowledge base that can be used to control and improve data mining process. In this paper the features of decision tree algorithms CART, ID3, C4.5, CHAID were analyzed. The analysis shows that parameters of algorithms could be related with enterprise model components and attributes (i.e. domain knowledge items). The alignment of business domain attributes used in the decision tree algorithms with the attributes of EM components is defined in the Table 4. Enterprise model (EM) components (Process, Function, Rules, Objectives and etc.) are conjugated with decision tree algorithm components (Root, Root attributes, Node, Node attributes and etc.).

The detailed analysis of the decision tree algorithm C4.5, integrated with domain knowledge (with components of Enterprise model) is described, relations of Enterprise model components and decision tree algorithms is verified.

Integration of an enterprise knowledge base in to data mining techniques can give additional functionality for experts to manage data mining process and to save important operational experience, related with definite business domain. Such data mining system enhanced by Enterprise knowledge components can give additional functionality for experts.

References

1. Larose, D.T.: Data mining methods and models, 340 p. John Wiley & Sons, Inc., Hoboken (2006) ISBN-13 978-0-471-66656-1
2. Phillips, J., Buchanan, B.G.: Ontology-Guided Knowledge Discovery in Databases. In: Proceedings of the 1st International Conference on Knowledge Capture, pp. 123–130 (2001)
3. Gudas, S., Lopata, A., Skersys, T.: Approach to enterprise modeling for information systems engineering. Informatica 16(2), 175–192 (2005) ISSN 0868-4952
4. Mavroudakis, T., Karanikas, H.: Intelligent Search via Ontology Driven Metadata Analysis. In: 10th Intl. Protégé Conference, Budapest, Hungary, July 15-18 (2007)
5. Pan, D., Jun-Yi, S.: Ontology service-based architecture for continuous knowledge discovery. In: Proceedings of 2005 International Conference on Machine Learning and Cybernetics, vol. 4, pp. 2155–2160. IEE Press (2005)
6. Kohavi, R., Quinlan, J.R.: Decision-tree discovery. In: Klosgen, W., Zytkow, J.M. (eds.) Handbook of Data Mining and Knowledge Discovery, ch. 16.1.3, pp. 267–276. Oxford University Press (2002)
7. Das, S., Saradhi Varma, G.P., Naidu, G.R., Murthy, A.V.D.N.: Data Mining Predictive Technique: CART. Research Cell: An International Journal of Engineering Sciences (2011) (Issue September 2011) ISSN: 2229-6913

8. Kaur, P., Sukhwinder, S., Sushil, G.: Harmanpreet. Limited Search, ID3, CHAID, C4.5, Improved C4.5 and OVA Decision Tree Algorithms to design Decision Support System. In: CP1324, International Conference on Methods and Models in Science and Technology (ICM2ST 2010). American Institute of Physics, 4 p (2010), 978-0-7354-0879-1/10
9. Subrata, P., Utpala, N.C., Pramanik, B.K., Huda, N.: A comparative study of bagging, boosting and C4. 5: The recent improvements in decision tree learning algorithm. Asian Journal of Information Technology 9(6), 300–306, 7 (2010) ISSN: 1682-3915
10. Fowdar, J., Bandar, Z., Crockett, K.: Inducing Fuzzy Decision Trees in Non-Deterministic Domains using CHAID. American Association for Artificial Intelligence, 6 p (2002)
11. RapidMiner Report for the Future.pid-I GmbH (2012), http://rapid-i.com/content/view/181/196/
12. Pentaho Data Mining (Weka). Pentaho Corporation (2005-2011), http://weka.pentaho.com/
13. Orange, Features. Orange (2012), http://orange.biolab.si/features/

Open Innovation in Health Service Value Networks: A Methodology for the Innovation of Ambient Assisted Living Platforms and Services

Davor Meersman[1] and Pieter De Leenheer[2]

[1] Curtin University
Kent Street, Bentley 6102, Australia
d.meersman@curtin.edu.au
[2] VU University Amsterdam
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
p.g.m.de.leenheer@vu.nl

Abstract. In this paper we introduce a methodology for collaborative innovation in the ambient assisted living domain using service value networks. We look to solve problems associated with heterogeneity, domain knowledge formalisation, low acceptance, integration and immaturity in the ambient assisted living domain. Our methodology consists of domain elicitation and modelling, needs elicitation and laddering, atomisation, recombination, and deployment and monitoring.

Keywords: ambient assisted living, healthcare systems, innovation, service value networks, service platforms.

1 Introduction and Motivation

The population of the European Union is undergoing significant demographic changes that have various implications on the nature of services and innovations in the current and future extended European economic space. By 2060, the European population older than 65 years old is projected to be more than 30 percent [1]. The associated cost of consequentially increased care is expected to be a significant burden on European economies [2]. This effect is further exacerbated by the old age dependency ratio (i.e. the population older than 65 divided by the working age population supporting them) is expected to rise from 25% to 53% by 2060 [1], meaning that for every old person there will be maximally two people of working age that can support that person, compared to four people now.

Assisted living solutions underpinned by ambient intelligence technology have been identified as a viable option to mitigate the impact of the associated cost of the demographic changes faced by Europe. This trend is apparent also from the European Commission's comprehensive Ambient Assisted Living (AAL) Joint Programme, which has funded 50 projects to date with a budget of EUR 600 million in calls focusing on the prevention and management of chronic conditions, the advancement of social interaction, and on the advancement of elderly people's independence and

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 25–36, 2012.

participation in the 'self-serve society' [3]. AAL technologies can help provide autonomy to elderly and disabled people, allow them to live at home individually for longer, raise their quality of life and to relieve some of the economic burden on public health care systems in the process.

The content of this paper is organised as follows: Section 2 discusses the home care systems and AAL technologies. Section 3 outlines the challenges in AAL. Section 4 discusses innovation in the context of healthcare platforms and services. Section 5 introduces Service Value Networks as a vehicle for collective intelligence and co-creation. Section 6 outlines the innovation methodology, including elicitation and modelling, atomisation, recombination, and deployment. Section 7 evaluates the methodology in light of the AAL challenges. Section 8, finally, draws some conclusions and projects future work.

2 Home Care Systems and AAL

Systems focusing on supporting people with special needs in their home environment are called Home Care Systems (HCS). Technologies underpinning home care have various labels. 'Assisted Living' refers to devices and services that help people stay at home longer. 'Assistive Technologies' refer to devices that aid with daily living of patients. 'Telehealth' and 'Telecare' refer to remotely monitoring and supporting patients. 'Smart Home' refers to home automation and monitoring via sensor networks [4]. The home care system domain is coarsely categorised in [2] as follows:

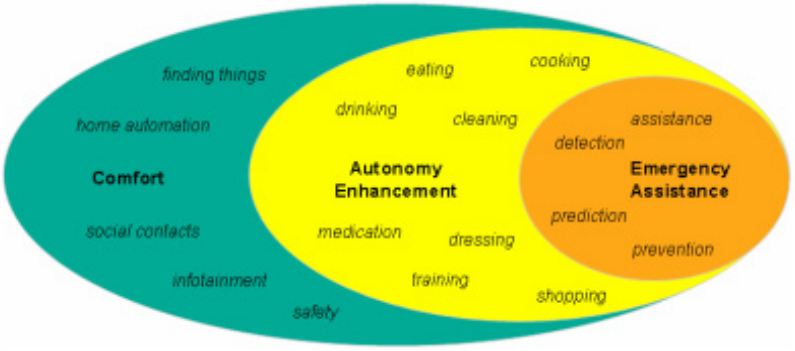


Fig. 1. The home care system domain [2]

What are defined in Fig. 1 as comfort, autonomy enhancement and emergency assistance services, can be seen as a rudimentary categorisation of patient needs, serviced by service providers. In our methodology, we use laddering techniques to derive specific needs from these general patient needs (described in Section 6). In the HCS domain, there are 2 main actors: Service Recipients (patients) and Service Providers (including physicians, home care givers and relatives) [5].

For now it is important to note that whilst the above categorisation is useful in itself as a description of the domain, it has no real function in the context of

developing innovative home care services. Innovators in the domain are generally already well endowed with the intricacies of the domain, and a categorisation has the limitations that processes belong to a single category. For instance, in the context of remote Diabetes Type 2 monitoring and management, the process `Medication` might satisfy the `EmergencyAssistance` need via a `ContinuousGlucoseMonitoringSystem` administering `Insulin` to the `Patient`, rather than satisfying `AutonomyEnhancement` as is proposed in the model.

Today's commercially available technologies in the AAL domain are products such as necklaces with emergency buttons, fall sensors in mobile phones with notification services, vital data monitoring plasters, wireless blood pressure and blood glucose meters [2, 5], and a myriad of other technologies that are increasingly being integrated in a smart object-based Internet Of Things with corresponding wireless standards such as ZigBee Pro [6] and 6LoWPAN [7].

3 Challenges in AAL

The ambient assisted living domain is relatively young and in the early stages of its development. A number of technological AAL challenges have been identified in [2]:

1. Adaptivity: systems need to monitor their environment and adapt themselves constantly.
2. Natural interactions: systems need to provide interfaces for users with varying needs.
3. Heterogeneity: systems are closed, standalone, and provided by different suppliers with diverging knowledge and technologies.
4. Domain knowledge formalisation: domain knowledge that is difficult to formalise needs to be transformed for processing.
5. Elderly stakeholders: the main stakeholders of AAL have generally low degrees of computer literacy and variable degrees of mental clarity, alertness and memory function, creating interface constraints.
6. Low acceptance: systems that are marketed as solely assisting with health problems have low acceptance rates because of the social stigma associated with them.
7. Integration of available technologies: AAL systems and services are characterised by heterogeneity, which offers integration challenges.
8. Immaturity: although it is generally expected that AAL will be a huge market, there is only limited knowledge about what the products will look like, what their economic viability will be, who will provide them, how they will integrate, etc.

In our work we focus on tackling 5 out of 8 of the above challenges for the AAL domain, namely we look to solve problems associated with *heterogeneity, domain*

knowledge formalisation, low acceptance, integration and *immaturity*. This is done in collaboration with a large telecommunications provider who provides the supporting smart home communication infrastructure on top of which the AAL devices and services exist. The smart home platform hardware consists of a back-end and network infrastructure, and a home gateway that controls a wireless sensor network, network communications and the delivery of a range of services including but not limited to: health, security and smart energy. The provision of services of these three domains over congruent endpoint infrastructure reveals that many services are overlapping and there is no need for disjoint approaches that result in a proliferation of heterogeneous systems and services that are potentially not economically viable or interesting on their own (e.g. solutions in the long tail of the spectrum). As a result of the agnostic product atomisation process in our methodology, different service providers can (collaboratively) innovate in adjacent domains. This blurs unnecessary, conceptual and artificial boundaries between domains.

4 Healthcare Platform and Service Innovation

There are two dimensions to the pace and nature of innovation on a HCS platform. One dimension is the evolution of the platform itself. As the platform grows and transforms, new capabilities are added to support new types of services. The other dimension is the innovation of services. By identifying, composing and developing new services, emerging customer needs are to be met with current platform capabilities. The platform innovation life cycle is typically slower because it involves hardware engineering, high investment costs and the need for a relatively stable service environment, whereas service innovation has a more dynamic nature and can thus respond to emerging customer needs faster [8]. However, all services are dependent on the platform over which they are delivered. Consider, e.g., the release of the iPhone 4 platform, which through its new 3-axis gyroscope functionality allowed the iPhone developer network to build new 3-axis gyroscope-utilising services that had heretofore not been possible. Or closer to home in the HCS domain, e.g. the introduction of fall detection sensors has enabled service providers to deliver emergency assistance to patients. Therefore innovation in a service-based economy will have a dual character of platform and service innovation, with the platform determining the provided services, and the services (or lack thereof) influencing the shaping of the platform.

A platform provider needs to take into account service dynamics and the evolving technological landscape, and anticipate accordingly. Related work on health care service customisation and personalisation [9, 10] is valuable and useful, but does not generally include platform constraints and evolution when devising new services. For this reason, whilst new services do arise from personalisation and customisation, it cannot be fully considered service innovation.

5 Use of Service Value Networks as Collective Intelligence for Service and Platform Innovation

The potential range of services in the HCS domain is huge. This has led to a proliferation, making it difficult to innovate in the space as such. Each provider supplies part of the solution, but suppliers do not co-innovate systematically. In our approach, we use Service Value Networks (SVN) as an ecosystem for collective intelligence, co-creation and open innovation. A service value network is a flexible and dynamic web of enterprises and final customers who reciprocally establish relationships with each other for delivering an added-value service to a final customer (see Razo-Zapata et al. [11]; Hamilton [12]; Allee [13]; and Lovelock and Wirtz [14].

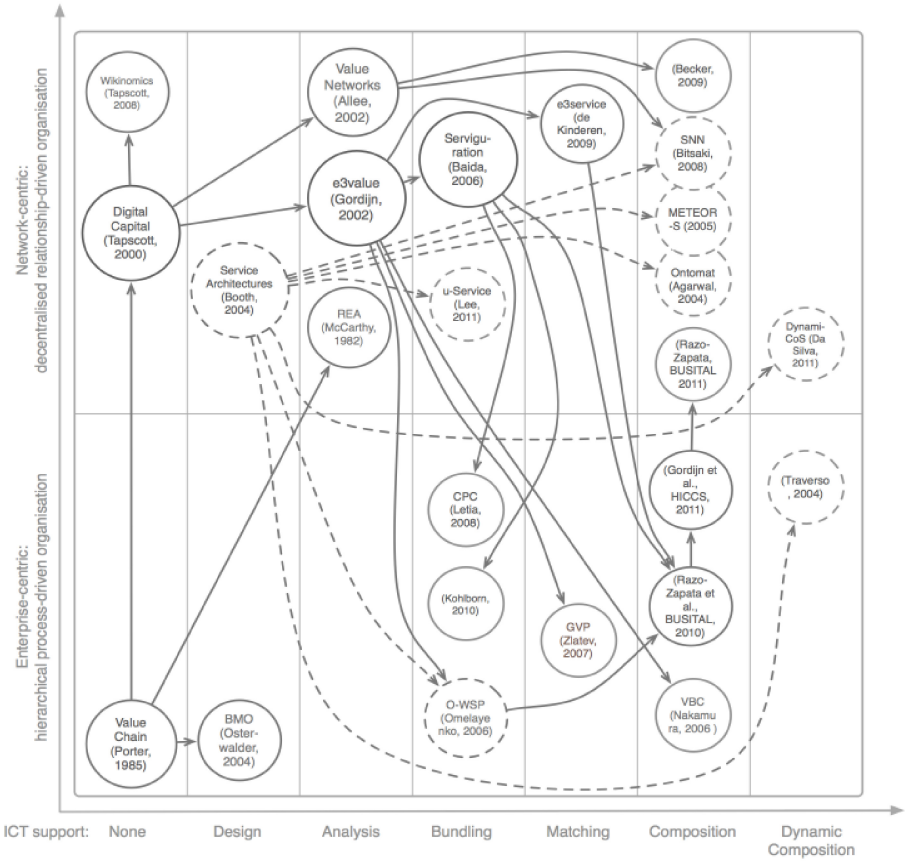


Fig. 2. Service Value Network approaches [11]

Whereas Value chains were sequences of activities that add each value to a production process, in SVNs the value is co-created in a networked setting. The key issue of SVN technology research is to automatically match customers with competencies on the market. In other words, SVN composition aims to bundle relevant competences in a network such that they deliver value in such a way that they answer a customer need. Suppliers, as well as customers can deliver these competencies. This value co-creation happens usually in different tiers: at B2C side, service providers make sure they offer relevant competences in fulfilment of a customer need; at B2B side service enablers make sure the technology space can make these services possible [15].

The foundation for SVN that is adopted in our approach is the well-established e3value [16, 17]. It provides ontologies to analyse and model perspectives of customers and providers on service needs. Inspired by service marketing and management theory, the conceptualisation of services focuses on value aspects, rather than merely computer-technical aspects as found in most service-oriented computing paradigms. As depicted in Fig. 2, e3value has been evolving since 2002 taking up different related service network approaches.

It shows how the innovation space in which different SVN approaches have emerged from business research [18, 19] and influenced each other. Solid circles represent business-oriented approaches taking into account value aspects of services, whereas dotted circles stand for process-oriented approaches hence focusing more on software aspects of services. The axes of the space indicate how these approaches co-evolve with changing business practice trends and increasing demand for ICT support: vertically, there is the economic context, which is evolving from a hierarchical process-driven organisation to a decentralised and relationship- driven organisation. Horizontally, there is the support of ICT in the different activities of SNs in these organisations.

The direction in which we aim to advance current state of the art lies in dynamic composition of service value networks in decentralised business environments. This corresponds to the upper-right corner of the SVN innovation space in the figure. These trends are confirmed in business research literature by, i.a., Tapscott [20], Van Heck and Vervest [21], and Chesbrough [22].

6 Methodology

Domain Elicitation and Modelling

The initial step is the analysis of industry-sourced information provided by industry representatives. This encompasses manuals, design documents, project deliverables, personas, training manuals, expert interviews, etc. Based on the initial domain knowledge, we develop a high-level domain model with a low granularity. After breaking up the domain in these main building blocks, we add general concepts and relations without delving into the specifics too much.

Adhesive	is a	subsumes	Property
Bathing	is a	subsumes	Process
AutonomyEnhancement	providedby	provides	Training
Infotainment	is a	subsumes	Process
AutonomyEnhancement	providedby	provides	Medication
ServiceProvider	performs	isperformedby	Process
Process	is a	subsumes	T
GlucoseSensor	usedin	uses	Prevention
Comfort	providedby	provides	SocialContacts
Actor	is a	subsumes	T
EmergencyAssistance	providedby	provides	Detection
ContinuousGlucoseMonitoringSystem	haspart	partof	Transmitter
Drinking	is a	subsumes	Process
Needs	is a	subsumes	Quality
AutonomyEnhancement	providedby	provides	Drinking
EmergencyAssistance	providedby	provides	Assistance
Sensor	is a	subsumes	Object

Fig. 3. Examples of lexons

In the domain modelling phase, we formalise the interview annotations as lexons (see Fig. 3), according to the DOGMA approach. A lexon represents a binary fact-type and is formally described as a 5-tuple (γ , concept A, role, co –role, concept B) where γ is the context (γ not shown in Fig. 3) [23]. The context is the HCS domain in our case. Role and co-role represent the relation and inverse relation of the respective concepts. The facts are added to the ontology until the representation for that part is complete, based on the provided information. Assumptions from the initial ontology are trumped by domain expert evaluations.

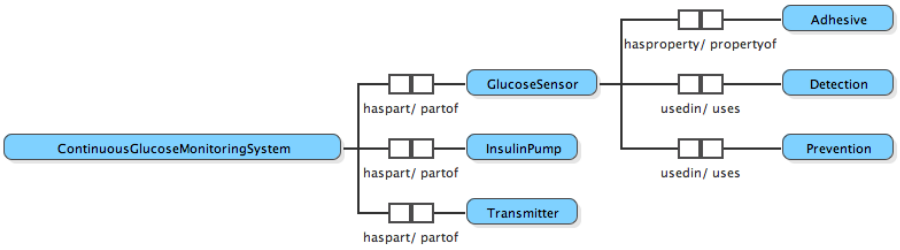


Fig. 4. Example of a domain knowledge pattern

Subsumption relations are modelled separately from other relations. The highest level of the subsumption hierarchy features the following classes [24]:

- Actor: any actor using objects in any process in the domain
- Object: products and components used in processes by actors and objects

- Process: any process in the domain, executed by an actor with the use of objects
- Quality: concepts that define how, to what extent, when etc. something happens; properties and functions of objects, actors and processes. The product application context is also included in the model, as the function of the product has an influence on the reasoning behind feature introduction. By formalising the entire domain, one can for example gain valuable insight in how certain functions that are performed by users today can be taken over by technology tomorrow.

Needs Elicitation and Laddering
Laddering has been widely used in marketing to represent how customers link specific product properties to high-level values [25]. In our case, by making use of the domain model, patient needs as stated by the domain model are refined into so-called functional consequences or FCs. For instance, a patient need such as "As an elderly diabetic, how can I enhance my autonomy?" can be refined into the following FCs: `Assistance`, `Prediction`, `Prevention`, and `Detection`. These processes use the following objects in the context of diabetes: `ContinuousGlucoseMonitoringSystem`, `BloodGlucoseMonitor`, and `InsulinPump`. These processes and objects better describe a patient need in terms of specific requirements [25, 26, 27].

Atomisation
We extract abstract properties and functions from objects, services and components in a domain. Properties and functions are related and one can rebuild objects and services based on their properties and functions.

Because the existence of properties in a domain is driven by the function they perform, i.e. the fulfilment of initial requirements and in some cases posterior cost considerations, the inclusion of functions and their linkage to products, product components and product properties is crucial in the context of ideation. Why is something there?

Recombination
Based on previous work on 'directed variation' [28] focusing primarily on product engineering, we recombine property/function clusters to create new services or products (i.e. platform components in the case of the HCS domain) based on elicited user needs. Product/service innovation takes places when a new property is found for an existing function, or when a new function is found for an existing property (see Fig. 5). Common examples of this principle in innovation are the Swiffer, the billion dollar brand household cleaner (property 'static electricity' instead of 'suction' for function 'cleaning') [28], or Jack Daniels Smoking Pellets, composed out of chopped up whiskey barrels (function 'burning' instead of 'containing' for wooden barrel properties).

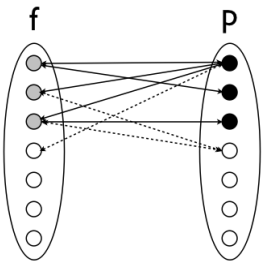


Fig. 5. Schematic representation of the recombination of properties p and functions f

In the HCS domain, patient needs are linked to property/function clusters. Processes (laddered needs) are linked to objects (property/function clusters). When new needs arise, the SVN evaluates them in terms of how they are linked to their part of the property/function cluster. It is important to identify a candidate service space that can fulfil the anticipated needs. For example, `Detection` and `Prevention` are linked to `FallDetection + Seismometer` and its properties. By recombining the properties, a certain type of `Accelerometer` can be a better solution for certain processes (for example in case of absence of possibility of seismometers, or when their properties cannot perform their function). By the principles of atomisation, the SVN can agnostically design solutions, and even providers outside of the traditional HCS scope can offer contributions.

Deployment
SVN design results in a set of quasi-equally interesting compositions of products/services that answer to the valued requirements. We verify their market suitability and derive a ranking by deploying them for limited and separate audiences. We then deploy the best-scoring SVN designs in a commercial setting and monitor/adapt them. This makes sure the innovation is sustaining under potentially disruptive technical innovations, and anticipates and co-evolves with changing business needs and technical opportunities.

7 Evaluation

Our approach for innovation in the HCS domain focuses on tackling following challenges for the AAL domain: heterogeneity, domain knowledge formalisation, low acceptance, integration and immaturity.

The use and SVNs in innovation promises to alleviate some of the burdens associated with heterogeneity: by collaboratively innovating across company boundaries, domain limits, and service/platform divisions.

Our approach of combined high-level domain formalisation and agnostic atomisation offers opportunities for domain formalisation in an innovation context. We do recognise that in the context of medical services and science, the intricacies of the domain remain.

The agnostic nature of the atomisation process and the trans-domain context of our industry smart home use case (health, security, energy) offers opportunities for breaking open a perhaps currently too narrowly defined HCS domain.

By having a single, standards-compliant platform provider, integrated component providers and a service provider network, some of the integration challenges are solved at the requirements and subsequent innovation stage.

Finally, perhaps a generalised domain model with extractable function/property clusters and laddered customer needs, can shed light on the reality of the HCS domain.

8 Future Outlook and Conclusions

In future research, we will start using sensor data in our innovation process. The data-heavy nature of the sensor-driven AAL domain will provide opportunities for detecting and eliciting emerging patient needs patterns use these in innovation processes.

The smart home context of our AAL industry use case will allow service value networks from various domains to contribute to a more integrated view of AAL of which the boundaries blur into other domains.

Another area of research is the distinction between services and platform in light of atomisation, and how we can interpret new or unmet emerging service needs as patterns for platform innovation.

A final aspect for future work is the evaluation of the methodology by examing the process outcomes: during deployment, the innovation is deployed and its success measured, which also determines the success of the current methodology.

The principles and methodology introduced in this paper provide a novel framework for innovation in the ambient assisted living domain. Specifically, they contribute solutions to the problem scope of heterogeneity, domain knowledge formalisation, low acceptance, integration, and immaturity of AAL through domain elicitation and modelling, needs elicitation and laddering, atomisation, recombination, and deployment and monitoring.

References

1. Eurostat Press Office. Population Projections 2008-2060, http://europa.eu/rapid/pressReleasesAction.do?reference=STAT/08/119 (accessed on March 26, 2012)
2. Kleinberger, T., Becker, M., Ras, E., Holzinger, A., Müller, P.: Ambient Intelligence in Assisted Living: Enable Elderly People to Handle Future Interfaces. In: Stephanidis, C. (ed.) UAHCI 2007 (Part II). LNCS, vol. 4555, pp. 103–112. Springer, Heidelberg (2007)
3. Ambient Assisted Living Joint Programme. Catalogue of Projects (2011), http://www.aal-europe.eu/projects/AALCatalogueV3.pdf (accessed on March 26, 2012)

4. Turner, K.J., Docherty, L.S., Wang, F., Campbell, G.A.: Managing home care networks. In: Bestak, R., George, L., Zaborovsky, V.S., Dini, C. (eds.) ICN 2009, pp. 354–359. IEEE Computer Society, Los Alamitos (2009)
5. Dohr, A., Modre-Opsrian, R., Drobics, M., Hayn, D., Schreier, G.: The Internet of Things for Ambient Assisted Living. In: 2010 Seventh International Conference on Information Technology: New Generations (ITNG), April 12-14, pp. 804–809 (2010)
6. Wheeler, A.: Commercial Applications of Wireless Sensor Networks Using ZigBee. IEEE Communications Magazine 45(4), 70–77 (2007)
7. Mulligan, G., Group, L.W.: The 6LoWPAN Architecture. In: 4th Workshop on Embedded Networked Sensor, Cork, Ireland (2007)
8. Eisenmann, T., Parker, G., Van Alstyne, M.: Platform Networks: Core Concepts. MIT Sloan Technical Report (2007), http://ebusiness.mit.edu/research/papers/232_VanAlstyne_NW_as_Platform.pdf (accessed January 10, 2012)
9. de Blok, C., Luijkx, K., Meijboom, B., Schols, J.: Modular care and service packages for independently living elderly. International Journal of Operations & Production Management 30(1), 75–97 (2010)
10. Zarifi Eslami, M., Zarghami, A., Sapkota, B., van Sinderen, M.: Service tailoring: Towards personalized homecare systems. In: Proceedings of the 4th International Workshop on Architectures, Concepts and Technologies for Service Oriented Computing, ACT4SOC 2010, pp. 109–121. SciTePress, Athens (2010)
11. Razo-Zapata, I., De Leenheer, P., Gordijn, J., Akkermans, H.: Service Network Approaches. In: Barros, A., Oberle, D. (eds.) Handbook of Service Description: USDL and its Methods, pp. 45–74. Springer (2012)
12. Hamilton, J.: Service value networks: Value, performance and strategy for the services industry. Journal of Systems Science and Systems Engineering 13(4), 469–489 (2004)
13. Allee, V.: A value network approach for modeling and measuring intangibles. In: Transparent Enterprise Conference (2002)
14. Lovelock, C.H., Wirtz, J.: Services Marketing: People, Technology, Strategy, 7th edn. Pearson Higher Education (2010)
15. Basole, R.C., Rouse, W.B.: Complexity of service value networks: Conceptualization and empirical investigation. IBM Systems J. 47(1) (2008)
16. Gordijn, J., Akkermans, H.: Value-based Requirements Engineering: Exploring Innovative e-Commerce Ideas. Requirements Engineering Journal 16(4), 114–134 (2003)
17. Akkermans, J.M., et al.: Value Webs: Using Ontologies to Bundle Real-World Services. IEEE Intelligent Systems 19(4), 57–66 (2004)
18. Porter, M.: Competitive Advantage. Free Press, New York (1985)
19. Tapscott, D., Ticoll, D., Lowy, A.: Digital Capital - Harnessing the Power of Business Webs. Nicholas Brealy Publishing, London (2000)
20. Tapscott, D., Williams, A.D.: Wikinomics: How Mass Collaboration Changes Everything, Portfolio (2006)
21. Heck, E.V., Vervest, P.: Smart business networks: how the network wins. Communications of the ACM 50(6), 28–37
22. Chesbrough, H.: Open Services Innovation: Rethinking Your Business to Grow and Compete in a New Era (2011)
23. Trog, D., Tang, Y., Meersman, R.: Towards Ontological Commitments with Ω-RIDL Markup Language. In: Paschke, A., Biletskiy, Y. (eds.) RuleML 2007. LNCS, vol. 4824, pp. 92–106. Springer, Heidelberg (2007)

24. Meersman, D., Dillon, T.S.: The Open Innovation Paradigm and the Semantic Web: An Ontology for Distributed Product Innovation. In: Proceedings of OTM Workshops (2010)
25. de Kinderen, S.: Needs-driven service bundling in a multi-supplier setting: The computational e3service approach. PhD thesis, Vrije Universiteit Amsterdam (2010)
26. Razo-Zapata, I.S., De Leenheer, P., Gordijn, J., Akkermans, H.: Dynamic cluster-based service bundling: A value-oriented framework. In: IEEE 13th Conference on Commerce and Enterprise Computing (2011)
27. Razo-Zapata, I.S., De Leenheer, P., Gordijn, J., Akkermans, H.: Service value networks for competency-driven educational services: A case study. In: 6th International BUSITAL Workshop (2011)
28. Dewulf, S.: Directed variation of properties for new or improved function product DNA – A base for connect and develop. Procedia Engineering 9, 646–652 (2011)
29. Ticoll, D., Scott, R.: Collaborate and Innovate: a new world of sourcing. PriceWaterhouseCoopers (2007)

Fuzzy Expert System for Virtual Team Collaboration and Work Evaluation

Sandra Strigūnaitė and Dalia Kriksciuniene

Departament of Informatics, Vilnius University, Muitines 8, 44280 Kaunas, Lithuania
{sandra.strigunaite,dalia.kriksciuniene}@vukhf.lt

Abstract. The purpose of the article is to present a new approach of efficient and comprehensive virtual team collaboration data analysis. The original conceptual structure of the fuzzy expert system model includes hierarchical parameter structure and application of fuzzy rules that allow team performance analysis in a way that is similar to expert thinking. The parameter values used for virtual work evaluation are noted by experts and also gained from virtual collaboration logs by applying computational methods. The research results provide evidence for the feasibility of using the proposed system model as the enhancement of virtual collaboration environment for real time teamwork evaluation.

Keywords: Fuzzy expert system, virtual work evaluation, collaboration data analysis, hierarchical fuzzy system.

1 Introduction

The purpose of the article is to present fuzzy expert system for virtual teamwork performance evaluation. The aim of investigations is to analyse if the designed system is able to evaluate team work performed in virtual environment, provide assistance to the process of expert evaluation and present results similar to expert observations.

Work and labour management is more and more integrated into virtual environment with the help of software solutions specialized for team collaboration activities. Looking from the work effectiveness perspective virtual work is expected to be more advanced comparing to traditional work. Software developers offer plenty of tools designed to watch and supervise human resources in virtual space. Technological possibilities allow gathering different data for continuous observing activities of project participants during their all working time. However the problems of trust, control, productivity and quality are more relevant for virtual work.

Management of each project processes or work progress still needs human input. It is noted by project managers during the experimental research presented in this article, that in the environment of virtual work, the structured processes and established rules become a priority, as human leadership influence, traditionally based on observing teamwork processes, becomes more difficult to implement. Leaders evaluate execution of virtual work based on mainly subjective assessments about

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 37–43, 2012.

efforts of individual team members and their cooperation, by summarizing information provided by the software systems designed for virtual work and generating insights based on their own assumptions [9,10]. We explore if it is possible to design model for generating insights matching the assessment of project leader or expert based on the data extracted from virtual cooperation systems?

The following second part of the article deals with the the necessity of designing the system for assisting project leader, which could solve evaluation complexity of virtual team performance. In the third part the novel model of fuzzy expert system and its parameter structure are described. In the fourth part the experimental verification and results of applying the suggested fuzzy expert system are presented. The article is summarized in the conclusions section.

2 Complexity of Virtual Teamwork Evaluation

Comparing to traditional project management the concepts and principles of virtual work management differ not only due to the different modes of communication in virtual and regular workspaces. The difficulties which encumber evaluation of virtual work progress are related to undefined impact of leadership on team work and growing demand for tools fostering honesty of team members and trust in each other [1, 2]. Communication is fundamental for any kind of project team but essentially important within virtual team members [3] with no less important factor of trust [4].

The project leader of the team working in virtual environment loses possibility of direct observation and control of team activities, instead, he can evaluate team performance and affect its work only by virtual communication. In the scientific literature there are few research works for virtual work evaluation based on of data collected by teamwork software solutions. The investigations and solutions for virtual data extraction and their sequence visualization are presented in [5,6]. These research works show that if semantics of proper data visualisation is determined, it is possible to simplify evaluation of virtual work.

Human interaction management (HIM) theory introduced by Broninski defines different perspective of human work modelling and suggests specific notation for visualisation of team communication processes, emphasising the importance of roles of the participants and their interactions [7].

The possibility of virtual teamwork project performance evaluation by deriving interaction statistics-based variables and finding their causal relationships by applying balanced scorecard approach is researched in [8].

The feasibility of introduction of interaction–based variables and fuzzy rules for revealing their interrelationships for comprehensive evaluation of project team performance is researched in [9,10]. Communication information captured in project environment cannot be directly applied for evaluation of team members and defining project status, but the combined fuzzy analysis of the derived variables can predict project outcomes with sufficient precision.

3 Structure and Features of Fuzzy Expert System

Fuzzy expert system for virtual team collaboration and work progress evaluation is designed to follow the processes of human thinking and expert evaluation. It aims to evaluate progress of virtual work by computational model able of deriving evaluations which generally require high level integration of expert observations, analysis and experience. In the model the high level performance evaluations resembling expert comprehension are derived from the low level team collaboration data instances captured by the information systems and characterizing all activities of participants.

Suggested system includes method for indirect evaluation of three main characteristics (team, task and interaction) by combining various quantitative data and qualitative indicators both acquired from the system logs and derived by experts.

The conceptual model of suggested fuzzy expert system is presented In Fig. 1.

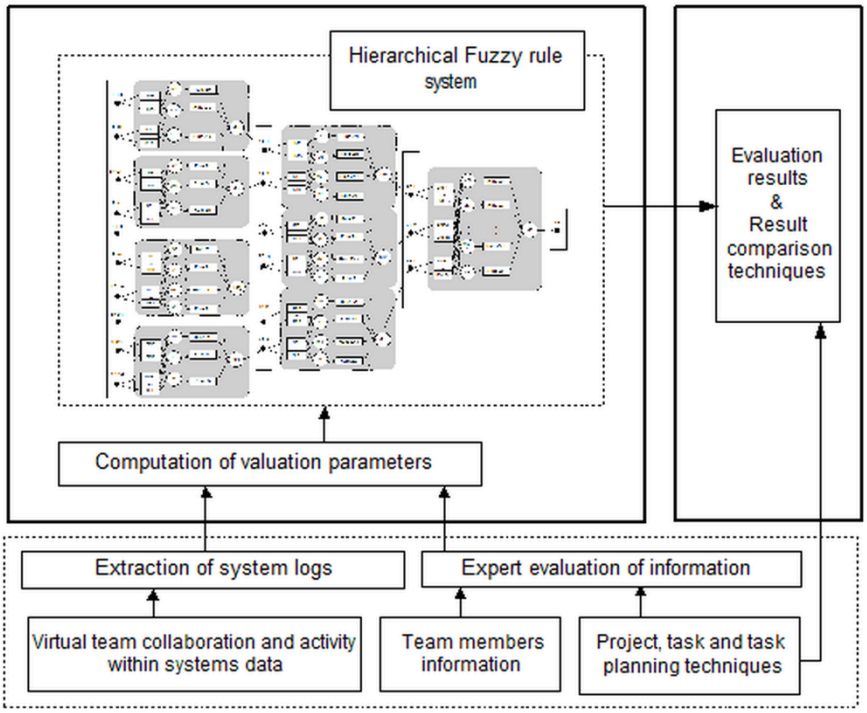


Fig. 1. The conceptual model of fuzzy expert system for virtual team collaboration and work progress evaluation

Model includes blocks indicating data flows and their accumulation actions. The main part of model is hierarchical fuzzy rule system that is based on applying fuzzy rules for hierarchical structure of evaluation parameters. Hierarchical fuzzy rule system serves for accumulating expert knowledge and experience information.

All parameters for fuzzy expert system can be selected in a way best describing evaluations of virtual team, task and collaboration activities by team members that

experts assign intuitively or by experience. The parameter set can vary in different enterprises and project settings, however it is implied that the suggested structure of valuation parameters has three hierarchical levels, arranged in three groups representing one of the evaluation objects: team, task or interaction.

As none of these variables can be directly measured by the information system used for team collaboration, they are evaluated by deriving their values from observing and measuring indirect characteristics. The principle applied for parameter value accumulation is that higher level parameter values are derived by aggregating lower level parameter values. Various operators, such as mean, min or max can be used for aggregation purposes. The final evaluation result is gained by accumulating all level parameters successively - starting from first level till the final result can be computed.

The hierarchical aggregation principle allows to solve two problems. Firstly, it allows to deal with the problem of missing variables in the lower hierarchical levels, as they can be replaced by variables of the higher level. Secondly, the fuzzy rules applied for hierarchical variable structure enable to encapsulate large number of variables and to avoid exponential increase of number of rules if applied for unstructured variable sets of same size.

4 Results of Model Verification Experiment

The aim of the model verification experiment is to test if the suggested fuzzy expert system with the hierarchical parameter structure can be applied for evaluation of task implementation progress.

The data of the experimental research was derived from project team, which members communicated in the virtual space during the most stages of the project. The project for creating electronic commerce solution was conducted at Vilnius university and performed by master level students of management information systems study program.

The analysed database consisted of 834 records of interactions among 9 members of two teams, who worked in the virtual environment *ComindWork* [11]. The number of interaction records were different (first team 558 and second team 167). Performing of 60 cases of project tasks by 9 members of the two project teams were analysed by two experts in order to define the intermediate results and final status of project results.

All parameters taken into consideration as experimental data are shown in Table 2. Parameters are arranged in one of three hierarchical levels and belong to one of the following groups: Team, Task or Interaction. Fuzzy expert system can provide sufficiently reliable results, only if it is constructed by using information that measures the progress of project task and by applying logical sequence of information processing, which is used by experts or other project assessors.

The project task outcomes were evaluated by consensus of the experts and by applying the suggested fuzzy hierarchical system model and compared by their performance in recognizing level of success of each task fulfilment leading to the final project result.

Table 1. Hierarchical parameter structure used in experiment

	First level (L1)	*Second level (L2)*	*Third level (L3)*	*Result*
Team (A1)			(C1) Team	Situation (C)
Task (A2)	(C21) Task intelligence level		(C2) Task	
	(C221) Phase	(C22) Task difficulty		
	(C222) Result clarity			
Interaction (A3)	(C31) Punctuality level		(C3) Interaction	
	(C321) Meetings	(C32) Activity level		
	(C322) Questioning			
	(C323) Information			
	(C324) Work			

The most significant part of the experiment is to select proper valuation characteristics of the membership functions (MF) of the parameters and to extract knowledge used to construct rule logics in order to fit to the investigated project situations and analysed data. The experimental data valuation parameter settings are presented in Table 3. Settings for A1 and A2 group parameters are defined by expert evaluations, and A3 group parameters are derived by computational data processing.

Table 2. Valuation parameter settings

Group	*Level*	*Parameter*	*MF num*	*Value range*	*Membership function (MF) values*
(A1)	(L3)	(C3)	2	[0 1]	Weak / Strong
(A2)	(L1)	(C221)	3	[1 10]	Start / Middle / End
(A2)	(L1)	(C222)	2	[1 10]	Clear / Confused
(A2)	(L1)-(L2)	(C21)	2	[1 10]	Routine / Effort driven
(A2)	(L2)	(C22)	3	[0 10]	Low / Middle / High
(A2)	(L3)	(C2)	3	[0 10]	Low / Middle / High
(A3)	(L1)-(L2)	(C31)	2	[0 1]	Late / On schedule
(A3)	(L1)	(C321)	2	[0 1]	Low / High
(A3)	(L1)	(C322)	2	[0 1]	Low / High
(A3)	(L1)	(C323)	2	[0 1]	Low / High
(A3)	(L1)	(C324)	2	[0 1]	Lack / Excess
(A3)	(L2)	(C32)	6	[0 1]	Very low/ Low with sharing/ Middle with low sharing / Middle with sharing / High with low sharing / Very high
(A3)	(L3)	(C3)	3	[0 1]	Bad / Correctable / Well

The validation result statistics related to evaluation of cases where fuzzy expert system succeeded or failed to identify task performance is presented in Table 4. Each team performed 30 tasks, which were evaluated by the project experts. From table 4 we can see that the suggested model had better performance in identifying task performance of the team, characterized as “strong”. The bigger number of the interaction data of group A3 increased precision of performance evaluation by the fuzzy hierarchical system, therefore its performance is highly similar to the task assessments provided by the project experts.

Table 3. Summary of experimental results

Team type	*Tasks*	*Result type*	*Case count*
Strong	30	Succeed to identify	23
Strong		Failed to identify	7
Weak	30	Succeed to identify	12
Weak		Failed to identify	18

Validation experiment shows that fuzzy expert system with parameter settings, presented in Table 3, and data aggregation methods, presented in Section 3, can generate task progress evaluation results which can assist the expert evaluation and follow their performance with high degree of reliability. The availability of communication data extracted from virtual communication systems tend to increase performance of the suggested model.

5 Conclusion and Further Works

Suggested fuzzy expert system model is designed for using by project managers or virtual team leaders to provide assistance in evaluating effectiveness of virtual teamwork collaboration and task implementation. The main challenges of evaluation of project processes performed by virtual teams are strongly based on subjective insights of the experts and the vague understanding of various observation-based characteristics and derived indicators preventing from consistent and quantitative evaluation of team performance and project status. The suggested technique is based on designing hierarchical parameter structure combining log information stored in virtual teamwork environments and expert evaluations which is further used for application of hierarchical fuzzy rule-based expert system able of providing evaluations of virtual teamwork processes.

The advantage of the suggested expert system is the possibility to provide objective evaluations by using variables extracted from virtual collaboration systems as fuzzy rules inputs. The system is different from the existing solutions because of the hierarchical fuzzy rule system and the method designed for calculating virtual teamwork collaboration situation evaluations. The hierarchical structure of valuation parameter helps to solve one of the main problems– the exponential increase of number of fuzzy rules by including more input variables.

Experiment results of virtual work valuation model, based on intellectual and fuzzy logics justify feasibility of applying the suggested model for project leaders for generating their insights and judge project status. Further investigations of the model are envisioned in the area of processing information extracted from virtual teamwork environments.

Acknowledgments. This work was carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme. This Programme is supported by the Marie Curie Co-funding of Regional, National and International Programmes (COFUND) of the European Commission. We would like to thank everyone who helped during experimental data collection and model verification experiment implementation.

References

1. Balthazard, P., Waldman, D., Howell, J., Atwater, L.: Shared Leadership and Group Interaction Styles. In: Proceedings of the 37th Hawaii International Conference on System Sciences, p. 10. IEEE Press (2004)
2. Konradt, U., Hoch, J.: A Work Roles and Leadership Functions of Managers in Virtual Teams. International Journal of E-Collaboration, 16–35 (2007)
3. Monge, P., DeSanctis, G.: Communication Processes for Virtual Organizations. Journal of Computer-Mediated Communication 3 (2006)
4. Jarvenpaa, S.L., Leidner, D.K.: Communication and trust in global virtual teams. Organization Science 10(6), 791–815 (1999)
5. Babic, F., Wagner, J., Jadlovska, S., Lesko, P.: A logging mechanism for acquisition of real data from different collaborative systems for analytical purposes. In: 8th International Symposium on Applied Machine Intelligence and Informatics, pp. 109–112. IEEE Press (2010)
6. Paralic, J., Richter, C., Babic, F., Wagner, J., Racek, M.: Mirroring of Knowledge Practices based on User-defined Patterns. Journal of Universal Computer Science 17(10), 1475–1491 (2011)
7. Harrison-Broninski, K.: Human Interactions: The Heart and Soul of Business Process Management. Meghan-Kiffer Press (2005)
8. Kriksciuniene, D., Sakalauskas, V.: Quality management and assessment for web-based virtual teamwork. In: Bastiaens, T., et al. (eds.) Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, pp. 1043–1052 (2009)
9. Kriksciuniene, D., Strigunaite, S.: Virtual Team Tasks Performance Evaluation Based on Multi-level Fuzzy Comprehensive Method. In: Nygard, K.E., Lorenz, P. (eds.) Future Computing 2011, pp. 20–25. Think Mind (2011)
10. Kriksciuniene, D., Strigunaite, S.: Multi-level Fuzzy Rules-Based Analysis of Virtual Team Performance. In: Skersys, T., Butleris, R., Nemuraite, L., Suomi, R. (eds.) I3E 2011. IFIP AICT, vol. 353, pp. 305–318. Springer, Heidelberg (2011)
11. Comindwork, manage projects online. 2007-2012 Comindwork, Online Project Management Software, `http://www.comindwork.com/`

Case-Based Business-to-Business Integration Model for SME – The Impact of ERP Implementation

Justyna Patalas-Maliszewska[1] and Irene Krebs[2]

[1] University of Zielona Góra, Poland
j.patalas@iizp.uz.zgora.pl
[2] Brandenburg University of Technology Cottbus, Germany
krebs@iit.tu-cottbus.de

Abstract. Managing small and medium sizes enterprises (SMEs) and building their competitiveness is not an easy task. Such enterprises not only differ in their organizational potential (managerial, intellectual, financial), but also in their operational power in relation to their surroundings, their susceptibility to external influences and their reactions to the needs of the market etc. This paper presents concept of the Business-to-Business Model Integration for SME.The authors suggest that the assessment of an ERP implementation depends on a firm's intention to use the ERP system to achieve strategic goals. We attempt to answer one of the fundamental research questions: "What are the benefits which have to be considered during the post-implementation phase an ERP system?" Using a real case study the authors put forward a potential possibility to create B2B Integration Model in accordance with the assessment of the post-implementation phase of an ERP system.

Keywords: Business to Business Model, an ERP system, SMEs.

1 Introduction

The fact that ERP (Enterprise Resource Planning) systems are in demand as a support tool in management is proven by their dynamic development and a significant growth in their sales [1]. Since all companies, including small and medium-size enterprises, can buy the same software, it can be clearly argued that the software itself is not the sole source of competitive advantage [4]. It follows from the theory of competitive strategy that, if an organization chooses to implement an ERP, the organization will seek ways to derive strategic value from the ERP [9]. Many organizations follow a change of management plan in which the strategy is changed before any complementary changes in people, processes, and technologies are implemented [7]. An ERP system opens new opportunities for strengthening a distinctive strategic position [14]. For an SME (small and medium-size enterprise) to operate competitively in an international and local market, it is necessary for it to deal with its tasks in an efficient and practical way; this can be achieved by the systematic analysis of the undertaken tasks in relation to the company's organizational and economical ability for their completion.

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 44–48, 2012.

Business-to-business (B2B) integration refers to all business activities that have to do with the electronic exchange of business documents between the companies [2]. B2B integration extends Electronic Data Interchange (EDI) by emphasizing that these business documents are exchanged as electronic messages following public business processes, i.e. business processes between the companies [2]. The purpose of B2B integration is to automate business interactions, i.e. the exchange of business documents in the public business process. In order to harmonize the meanings for terms, the modes of operations, and the messaging interfaces for B2B integration, the e-business frameworks specify the business documents, business processes, and messaging [12].

The post-implementation phase of an ERP system is an understudied research topic [5], [16]. The authors discuss the strategic impact of an ERP system implementation as the ability to create B2B integration model of a small service enterprise. This article addresses the following research questions: What is the post-implementation phase of an ERP system in an SME and why has it become an issue? What is a strategic impact - a concept which has to be considered during the implementation an ERP system? Some of these issues will be illustrated based on a real case of a service firm in Poland whose managers made effective use of an ERP system.

We suggest that the assessment of the implementation of an ERP system depends on a firm's intention to use the ERP system to achieve strategic goals. Enterprises functioning in the market economy have to implement changes in the systems of organization and management that they use [13]. This paper examines the B2B Integration Model for Service Small Enterprises as the result of the long-term effects of ERP system revisions, the information is based on a real case study. The next section presents an example of the implementation of Comarch CDN XL (an ERP) as a case study, we next show the possibility of building the Business-to-Business Integration Model for Service Small Enterprises as the result of the impact of ERP implementation. Finally, the paper discusses further research and presents conclusions.

2 Illustrative Example

In order to illustrate the possibility of assessing the strategic impact of the implementation of an ERP system, let us consider an SME that deals with providing services in the form of projects for both organizations and individual customers. The main areas of the company correspond to the following functions: the sale, the supply, the order scheduling, the service, the accounting, human resources management, export/import transactions. This company completed, in 2008, a project to implement an ERP system and the following modules were implemented: Contracts, Sales, Accounting, Admin, Mobile sales.

The strategic goal of an ERP system is the transformational capability of the organization to meet the new needs of its customers and the needs of any new customers. This also includes strategic partnering with customers – carried out between a network of about 300 stores. Also the following issues have been defined in this network: (1) a lack of communication between the network partners and suppliers, (2) an inability to transmit orders to suppliers for the purchase in a uniform manner, (3) an inability to download electronic invoices in a uniform manner, (4) an inability to send data on promotions, newsletters and current prices, (5) a lack of opportunities to exchange information among its members such as through the ability to create surveys, newsletters, (6) a lack of reporting on the performance of the network.

Extending the functionality of the ERP system for Business to Business solutions can help the organization to solve these problems and correct potential future problems with customers. The implemented ERP system can be extended with additional functionality that will enable the creation a Business to Business system. In accordance with the strategic goals: extending the functionality of the ERP system for Business to Business solutions allows for the following business processes to be defined: the sales process - improving the exchange of commercial information within the network - between members and the network operator. Comprehensive data exchange: the exchange of purchase orders, sales invoices exchange, registration of invoices; the promotional process - generating and distributing data on promotions, newsletters, current prices within the network; customer relationship management - improving the exchange of commercial information within the network - between members and the network operator. Comprehensive data exchange: exchange regarding customer satisfaction surveys, newsletters; the process of information management and reporting - data exchange on how to optimize business processes and partners, all through the mechanisms of effective management information provided by multilevel analysis.

The task of the new system will be to automate logistics processes within the company's network: (1) automatically send orders to suppliers for a purchase of a unified manner; (2) the automatic retrieval of electronic invoices in a uniform way; (3) the automatic routing of data on promotions, newsletters, current prices; (4) the automatic exchange of information among its members, for example, through the ability to create surveys, newsletters; (5) reporting on the performance of the network.

Based on the interview and literature, benefits from B2B in Table 1 were included in the questionnaire. For each of benefit in the questionnaire, we asked the respondent to answer the following question: How do you see the chosen benefit from B2B compared to the situation without B2B?

Table 1. The benefits from B2B for SME

Benefits from B2B (based on the literature) [3], [6], [8], [10], [11], [15]	**Benefits from B2B (based on the interview in a small enterprise)**
Reduce manual work	Improvement of business processes
Improve speed of business interactions	Quick response to changes in the company and within the network
Reduce errors in business interactions	Management of information directed to managers, suppliers and customers
Reduce data transmission costs	Reduce costs and increase the effectiveness of reporting and analysis of the network
Improve customer/supplier responsiveness	Use of knowledge about customers and their preferences
	Control the execution of business strategy and network

Based on this real case, the authors put forward a potential possibility to create B2B Integration Model in accordance with the assessment of the post-implementation phase of an ERP system.

3 Business-to-Business Integration Model for SME

B2B system will integrate multiple systems and programs used by members of the network to exchange data - making it easier to work inside the network. Allows you to transmit orders to suppliers for purchase and download electronic invoices in a uniform way - for providers in terms of computer network will be treated as a single client, it will be a tool of communication between partners and suppliers. Applicants - as the network operator - will enable reporting on the effectiveness and profitability of the network. So, the structure of B2B Integration model for SME is given (see Fig. 1).

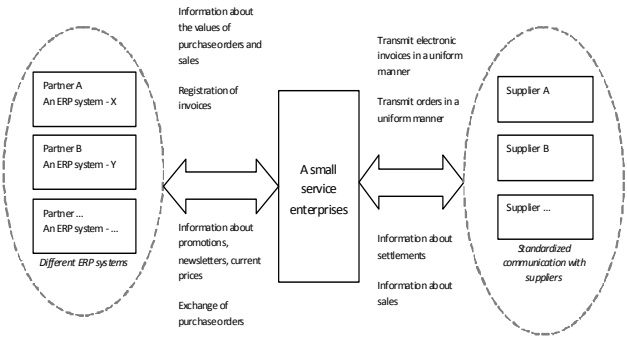


Fig. 1. The structure of the B2B Integration Model for SME

B2B system will be created in the following technologies: **(1)** database layer: Microsoft SQL Server 2008**, (2)** business layer: Microsoft Visual Studio**,** NET Framework 3.5 SP1**,** C # programming language**, (3)** user interface layer: Microsoft Visual Studio, NET Framework 3.5 SP1**,** the programming language C #, XAML**,** NET Framework ASP**,** XML, AJAX, MVC, Dynamic Data

SME sector enterprises have limited access to collaboration with research and development aimed at searching for the implementation and dissemination of innovation. Using the proposed B2B Integration Model may provide opportunities for being innovative. A small enterprise has access, via B2B connections, to other companies and can base their production capacity on the knowledge that different firms have and want to share innovations and are able to create them. This provides a powerful combination of the effects of several cooperating companies.

4 Concluding Remarks

The strategic decision, as far as ERP system implementation is concerned, is based on the assessment of its potential advantages, resulting from the operation of such a system in the company. Additional value for SMEs can be defined by an effective

implementation of an ERP system. SMEs with integrated IT systems also gain a further advantage in relation to the whole company because it enables them to have an insight into every aspect of its operation with a precise and correct evaluation of the company's financial situation.

The paper addressed the topic of evaluation of an ERP implementation in SME. After describing the real case study the B2B Integration Model for SME was identified to motivate future research. The next step of the author's research will be more detailed analysis of the problem domain. After developing initial mechanisms for B2B Model Integration creation, the goal is to extend these mechanisms and improve the prototype as the result of these further requirements.

References

1. Al-Mashari, M., Al-Mudimigh, A., Zairi, M.: Enterprise resource planning: a taxonomy of critical factors. European Journal of Operational Research 146 (2003)
2. Bussler, C.: B2B Integration: Concepts and Architecture. Springer, Berlin (2003)
3. Chwelos, P., Benbasat, I., Dexter, A.S.: Empirical test of an EDI adoption model. Information Systems Research 12(3), 304–321 (2001)
4. Davenport, T.: Putting the Enterprise in the Enterprise System. Harvard Business Review 76(4) (1998)
5. Gattiker, T.F., Goodhue, D.L.: What Happens After ERP Implementation: Understanding the Impact of Interdependence and Differentiation on Plant-level Outcomes. MIS Quarterly 29(3), 559–585 (2005)
6. Iacovou, C.L., Benbasat, I., Dexter, A.S.: Electronic data interchange and small organizations: adoption and impact of technology. MIS Quarterly 19(4), 465–485 (1995)
7. Lassila, K.S., Brancheau, J.C.: Adoption and utilization of commercial software packages: exploring utilization equilibria, transitions, triggers, and tracks. Journal of MIS 16(2) (1999)
8. Mackay, D., Rosier, M.: Measuring organizational benefits of EDI diffusion: a case of the Australian automotive industry. International Journal of Physical Distribution & Logistics Management 26(10), 60–78 (1996)
9. Mata, F.J., Fuerst, W.L., Barney, J.B.: Information technology and sustained competitive advantage: A resource-based analysis. MIS Quarterly 19(4) (1995)
10. Mukhopadhyay, T., Kekre, S., Kalathur, S.: Business value of information technology: a study of electronic data interchange. MIS Quarterly 19(2), 137–156 (1995)
11. Ngai, E.W.T., Gunasekaran, A.: Implementation of EDI in Hong Kong: an empirical analysis. Industrial Management & Data Systems 104(1), 88–100 (2004)
12. Nurmilaakso, J.M., Kotinurmi, P.: A review of XML-based supply-chain integration. Production Planning & Control 15(6), 608–621 (2004)
13. Patalas-Maliszewska, J., Krebs, I.: Model of innovation transfer is small and medium enterprises (SME). In: Proceedings of the 5th International Conference on Advances in Production Engineering, APE 2010, Warsaw University of Technology. Faculty of Production Engineering (2010)
14. Poston, R., Grabski, S.: Financial impacts of enterprise resource planning implementations. International Journal of Accounting Information Systems 2, 271–294 (2001)
15. Scala, S., McGrath, R.: Advantages and disadvantages of electronic data interchange: an industry perspective. Information & Management 25(2), 85–91 (1993)
16. Staehr, L., Shanks, G., Seddon, P.B.: Understanding the business benefits of enterprise resource planning. In: Eighth Americas Conference on Information Systems, pp. 899–906 (2002)

Adaptive User Interface Personalization in ERP Systems

Zdenek Eichler[1] and Martin Dostál[2]

[1] Faculty of Informatics, Masaryk University Brno, Botanická 68a
60200 BRNO, Czech Republic
zdenek.eichler@mail.muni.cz
[2] Wolkerova 647/4
77900 OLOMOUC, Czech Republic
martindostal@me.com

Abstract. Enterprise Resource Planning (ERP) systems had grown into complex applications integrating various processes and large amounts of data in detail. Such complexity negatively impacts usability owing to providing single static user interface that should meet users with diverse roles, tasks and requirements. In this paper we introduce a concept of adaptive user interface personalization in the SAP ERP system. We address two important aspects of personalization; (1) personalization to the user's prominent tasks, (2) personalization of individual on-screen task forms using adaptive progressive disclosure with ephemeral visualization. In the next we describe a mock-up of our adaptive system in detail and results of brief questionnaire study.

Keywords: Adaptive User Interfaces, ERP Systems, Usability.

1 Introduction

Today Enterprise Resource Planning (ERP) systems are often large-scale complex systems covering a huge number of different tasks performed across entire organization [1, 18]. Among other aims, ERP systems struggle for an increase of the operational efficiency and information flow within the organization [2]. However, ERP systems had been criticized for poor usability over a long period [21]. The reasons for usability issues can be identified in several causes; (1) ERP systems are complex systems covering a huge number of tasks. (2) There is a lack of usability metrics and methodologies that address ERP specifics [11, 20]. (3) ERP's user interfaces often do not follow Human Interface Guidelines and principles of good user interface design appropriately. (4) A static, "one-size-fitts-all", interface is shared by users with diverse roles, tasks, requirements and responsibilities. Typically, users use only a small fraction of the provided functionality while the unused functionality represent a burden for efficient use of the system. Although there are no studies available on analyzing the differences between ERP systems users, these had been analyzed in the area of word processing tasks in detail [3, 4, 16].

Poorly organized interfaces are obviously a particular cause for users's annoyance, disorientation, confusion and cutting of productivity. Moreover, substantial differences across users can be identified even in performing a single task. The tasks in ERP user

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 49–60, 2012.

interfaces are typically represented by complex on-screen forms composed of tenths of user controls. However, users of different roles usually interact with a small subset of user controls available within the form. These differences in requirements can be addressed by user interface personalization.

User interface personalization is a process in which the user interface is adjusted to individual user's needs. Personalization may be either user- or system-driven. It is represented by *adaptive* and *adaptable* user interfaces. Adaptive interfaces dynamically and automatically adjust the interface in order to conform to the user's needs and usage habits. Such systems provide a mechanism to understand the user and build-up personalized user interface. Again, adaptable interfaces provide customization mechanisms which rely on the user to apply those mechanisms to do the adaptation. For instance, users can customize a menu, a toolbar or assign custom keystrokes. The advantage of user interface personalization is that the user is provided with a user interface that is customized to individual requirements. On the other side, personalization has also disadvantages. It has been reported that the user-driven adaptation is usually too complex for the "average user" [15, 17]. Various studies [10, 12–14, 19] reported users' dissatisfaction by adaptive user interfaces when usability design principles are violated. Also several comparative studies between adaptive, adaptable and static user interfaces [7,9,15,23] showed that users preferred a static or an adaptable user interface over the adaptive one.

To date, adaptive interfaces have been studied mostly on research mock-ups (these applications serve for interaction research, no real functionality is provided), or either word processors or spreadsheets. However, the idea of adaptive personalization of ERP systems is not new. It has already been mentioned by Singh et. al [20–22], for instance. In this paper we introduce a concept of adaptive user interface personalization in the SAP ERP system that (1) personalizes the user interface to tasks frequently and/or recently used by the user, and (2) personalizes on-screen forms to frequently and/or recently used parts of the form on the basis of the so-called ephemeral visualization [8]. The concept is built upon our existing system called "Boulevard" which personalizes user interface of the OpenOffice.org Writer text processor. We introduce Boulevard briefly in the following section.

2 Boulevard

Boulevard, our system for adaptive user interface personalization (depicted on Fig. 1), is represented as a panel container window containing the personalized user interface of the application — currently the OpenOffice.org Writer word processor. In contrast to the adaptive user interfaces these days, it does not modify the static part of the interface destructively (e.g. menu or toolbar contents or structure) so that a user is able to clearly understand which part of the user interface is static and which part is dynamic (adaptive). Seemingly, this approach could reduce user disorientation and improve the understanding of the user interface. It also provides a choice to use or not to use the adaptive part of the user interface. Furthermore, since the panel container centralizes the most preferred functionality of the application, the user's effort required to find a command in complex menus or toolbars is reduced. In contrast to pull down menus,

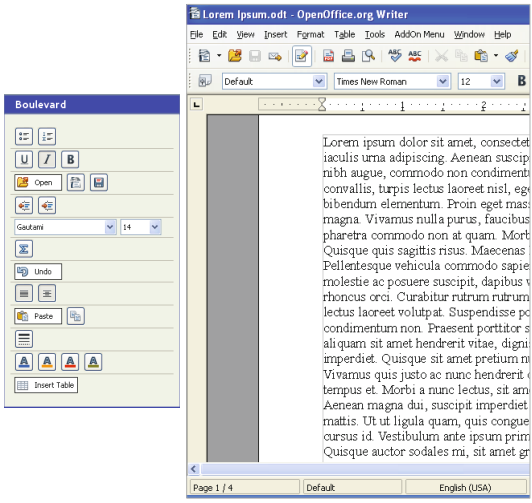


Fig. 1. Boulevard in OpenOffice Writer word processor

the functionality provided by our adaptive panel container is directly available without the need to open and navigate within a menu — this feature is sometimes referred as "one-click" approach.

Our system addresses various aspects of personalization, such as personalization to user preferred commands, preferred user interface representations and most frequently used parameter values applied to user commands. The system automatically tracks the user activity in the Writer application and considers the frequently and/or recently used user commands (see Section 2.1 for more detail) as the prominent for the user. Boulevard is driven internally by a forward-chaining expert system, more information about the expert system implementation can be found in [5]. Boulevard provides the following features to support the user:

Discovering prominent functionality — we propose innovative algorithms for computing the user command prominence. See section 2.1 for details.

The "Sweeping-back" feature — provides an intelligent organization of similar user commands. Besides the frequency and/or recency of usage, our system organizes user commands in the panel according to their semantical relationships using the grouping at rows. For example, "Bold", "Italic" and "Underline" represent a group of font style related commands, or "Align-left", "Align-right", "Center" and "Justify" command represent the text alignment group.

Adaptive representation — performs the personalization on user interface usage style. In addition to the user commands usage, our system also tracks *interaction styles* used to issuing user commands. Most user commands can be applied using various interaction styles, typically using a menu, a toolbar button, a pop-up menu, a keystroke or a dialog. Our system presents user commands in the panel container visually, according to the most frequently used interaction style. The most frequently used interaction style on a command is considered as the most preferred

representation. Presumably, users associate user commands with their visual representation, e.g. “Print” as a toolbar icon with a printer symbol or “New” as a menu item in an upper part of “File” menu category. It enables a visual association with a user command representation in the panel container corresponding to the interaction style preferred by the user.

Recommending parameter values for user commands — we extend the personalization procedure by recommending the most frequently used values to apply to user commands, e.g., most frequently used colors or font sizes. Actually, such parameters are offered to the user as individual commands applied directly with the desired value, e.g. red text color button, blue text color button, etc. See the next-to-last row of the Boulevard depicted on Fig. 1.

Adaptive disclosure — the panel container allows to contain, apart from the menu and toolbar items, also dialogs depicted in Boulevard as forms. However, only the frequently used parts of dialogs are presented. In fact, it is an adaptive variant of the well established interaction technique called *progressive disclosure*, which sequences complex dialogs or screens to several parts (i.e. from basic to advanced features) that can be disclosed progressively by the user in order to manage the visual clutter and reduce overwhelming number of presented features to the user. The principal difference from the common progressive disclosure is that the our variant is controlled adaptively.

2.1 Discovering the Prominent Functionality

Here we describe how the prominence of a user command is observed from the user's interaction with the application. Basically, we consider the most *frequently* used commands the most prominent for the user. It well reflects long-time usage patterns. In addition to tracking the frequency of user commands usage, we also track *recently* used commands in order to personalize to short-time usage patterns. The motivation for tracking the recently used commands can be expressed by the following example: a user obtains a new task which requires the usage of not yet frequently enough used functionality of the application. Nevertheless, such functionality may also be useful to be included in the panel container in order to support the user's work on the new task. Using the tracking of user command usage frequency only, such functionality would not appear in the panel container until it is used frequently enough. However, by utilization of the recently used commands tracking in conjunction with tracking of the frequency of usage, such functionality can quickly appear in the panel container. Again, if the user stops using such recently used functionality, it will disappear quickly from the panel container.

Using Formula 1 the user command prominence is computed. In this formula, $|x|$ expresses the count of user command x activations, T expresses the total count of all commands activations. This part of the formula represents the *frequency of usage* factor. The rest of the formula is related to the *recency of usage* factor. Each user command may appear in the queue of size q at multiple positions. We put the set P_x which contains the positions of occurrence of user command x in the queue. The topmost position is 1 (refers to the most recently used user command) and the lowermost position is q.

The $w \in < 0, 1 >$ parameter represents the weight of *relative frequency of usage* factor (long-time usage patterns), while $1 - w$ represent the weight of *recently used user commands factor* (short-time usage patterns). The result of the formula is a value which we term as the *rank* of a user command x. $\mathrm{Rank(x)} \in < 0, 1 >$ expresses the prominence of a particular user command x for the user.

$$\mathrm{rank(x)} = w\frac{|x|}{T} + (1 - w)\frac{\sum_{p_i \in P_x}(q - p_i + 1)}{\sum_{i=1}^{q} i} \tag{1}$$

2.2 Proof-of-Concept Usability Study

The goal of the study was to verify the basic concepts behind our system. We used both quantitative and qualitative measures focused on task time and error rate analysis measured on selecting user commands using three different interaction styles: toolbar, menu and Boulevard. The study provided promising results because the utilization of Boulevard was found comparable to toolbars in quantitative measurements (although no statistically significant difference was found) and faster than menus. More importantly, Boulevard was well received by users. According to qualitative measures, it was rated as a better interaction style than toolbar and menus by most of the participants. Furthermore, most participants reported that they would like to use Boulevard. More details can be found in [5, 6].

3 Adaptive Personalization of ERP System

We introduce an adaptive user interface that inherits Boulevard benefits and principles. Boulevard integrated in ERP system should be little different from the Boulevard in word processor owing to different interaction techniques and nature of ERP systems and enterprise environment.

Adaptive user interfaces seem promising to be used in ERP systems for several reasons. From the economic point of view, a properly personalized user interface could reduce task times and thus increase productivity. Furthermore, a precise customization of ERP individual developers is very costly and must be performed repeatedly to achieve good results. That is why we introduce the adaptive user interface which can perform continuous, fine-grained adaptation without additional costs for customization. Below we present principles and features of Boulevard modified and designed for the ERP system environment. Not all features of Boulevard's word processor implementation are utilized, because screens in ERP systems are typically more complex than most dialogs in word processors, thus some features seem not suitable for ERP systems and some must be adapted from the word processing environment to ERP.

3.1 Nondestructive Approach

Boulevard separates the user interface into the adaptive (dynamic) part and the static part. The adaptive part of the interface does not affect the static part in contrast to destructive approach used in many adaptive user interfaces. Usage of an add-on panel,

where the functionality is duplicated and the original interface remains in its original state, was found beneficial in our study [5,6]. It has been recognized that interfaces destructively changing under the user's hand are considered annoying and causes excesses in task times in cases when adaptivity does not supports the user correctly. We consider such property especially important for enterprise environment, for the following reasons: (1) when the prediction fails to meet users' requirements, they can easily continue their work using the original static part of the user interface without an excess in task time. (2) A user support (e.g. on telephone or by manuals) is much more easy when both advisor and user see exactly the same user interface, which is hard to accomplish in destructively adapted user interfaces.

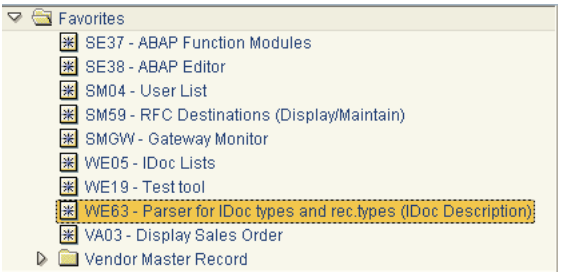


Fig. 2. SAP ERP Easy Access

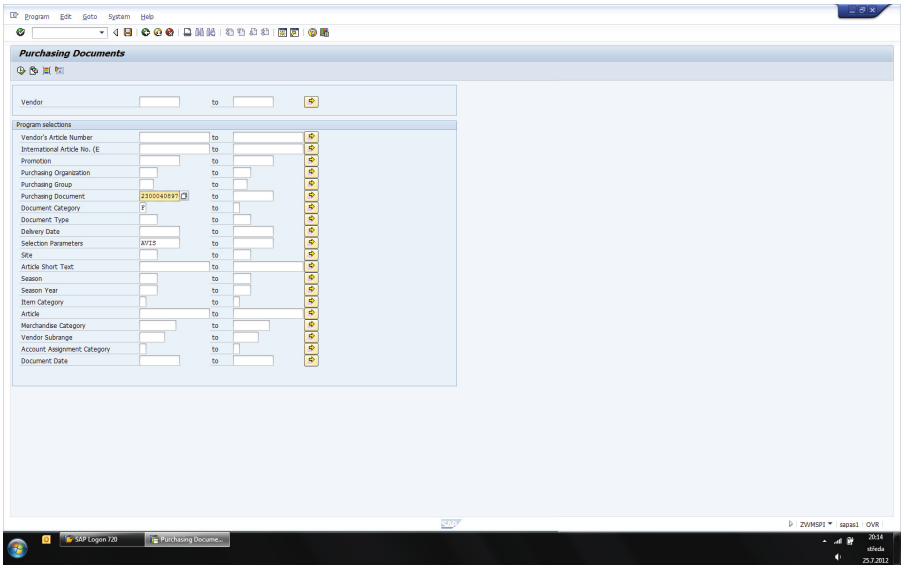


Fig. 3. Original SAP ERP user interface

3.2 List of Prominent Functionality

In the ERP environment user commands are represented by the so-called *transactions* (i.e. data manipulation tasks) and *reports*. Todays ERP systems often utilize a customizable list (in SAP ERP called *SAP Easy Access*) of user defined commands. Such a list we depict on Fig. 2. The list of prominent functionality must be maintained manually and repeatedly by the user, while performing such a customization is quite a complex task. It has been recognized that the user-driven adaptation is often too complex for the user [15, 17]. Furthermore, as users' work habits change, they need to recustomize the user interface again.

In Boulevard, the list of frequently and/or recently performed user commands is maintained automatically, which saves user's time and energy required to maintain customization manually. Also skills required to perform such personalization by user are not necessary. Prominent commands are easily accessible using direct, "one-click" activation. The algorithm for discovering prominent commands is basically the same as in Boulevard for the word processing environment. In order to support the readability and predictability of the Boulevard behavior, the user interface changes are supported by animations.

3.3 Recommending Predicted Parameter Values

This feature assists the user in performing user commands (transactions) repeatedly with some constant invariant parameters (fields values). Boulevard ensures automatic fill-in of frequently and/or recently used values in particular fields. This feature prevents the user from entering the same values in particular form fields repeatedly.

3.4 Adaptive Disclosure with Ephemeral Visualization

This feature should reduce the visual search time and accelerate interaction in complex screens (transactions) while maintaining spatial consistency. Adaptive disclosure is processed in three stages: (1) discovering the prominent functionality, (2) visualization of the prominent functionality and (3) adaptive focus traversal in the on-screen forms.

Discovering the Prominent Fields. Prominent fields are discovered using the same approach as discovering prominent user commands as it was discussed in Section 2.1.

Ephemeral Visualization. Ephemeral visualization introduced by Findlater et.al [8] reduces the visual search time in complex screens while maintaining spatial consistency. Using ephemeral visualization, predicted prominent fields become visible immediately, while the rest of fields (those that were not predicted as prominent) will fade-in gradually, but stops at 75% of visibility, so that the user is able to easily distinguish between prominent and non-prominent fields, even after animation completes. Ephemeral visualization preserves original positions of the fields, which conforms the above-mentioned nondestructive principle and maintains users' orientation in the form by preventing moving or reordering of form fields.

When prediction works perfectly, the user perceives only fields he/she use without being disturbed by constantly unused fields. But when the prediction fails, the rest of fields (the not predicted ones) will get more visible in time gradually. Ephemeral-based menus was found faster than static menus when accuracy of predicted items is high and also not significantly slower when it is low [8]. This property is very beneficial for enterprise application where the excesses in task times are unwanted. The duration of ephemeral visualization (gradual fade-in) is also adaptive — Boulevard tracks how long does it take a user to fill-in the particular form.

Adaptive Focus Traversal. The user is navigated in fill-in form more effectively through predicted fields. During the focus traversal, the non-predicted fields are skipped automatically. Such optimized interaction optimizes the time needed to navigate through fields. When needed, the user is able to skip into a non-predicted field using the `CTRL + TAB` keystroke respecting the original field order on screen strictly. When such not-predicted field gets focus, the visibility of the field changes from 75% to 100% immediately.

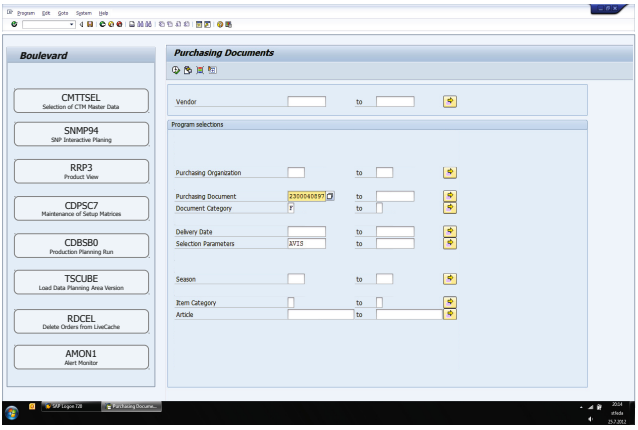
4 Boulevard Mockup for the SAP ERP

In this section we present a mockup of Boulevard for the SAP ERP which is a well-known widely used ERP system, depicted on Fig. 3. We depict a mockup of Boulevard for the SAP ERP on Fig. 4(c) . The adaptive part on the left contains the predicted prominent (frequently and/or recently used) user commands (transactions). Boulevard provide the so-called "one-click" availability of prominent user commands. The *recommending prominent parameter values* ensures the automatic fill-in of corresponding fields (*purchasing document* and *selection parameters*) by the frequently used identical values.

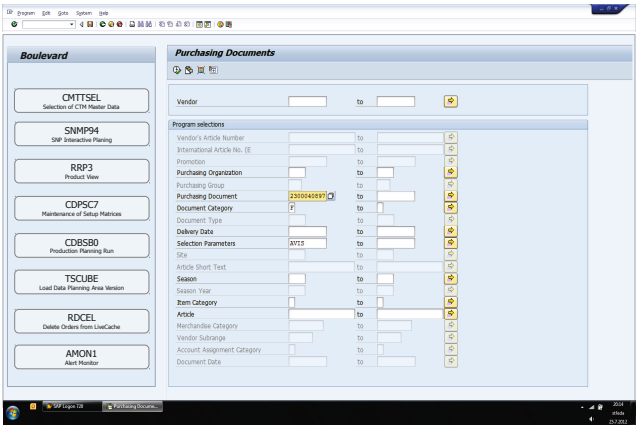
Fig. 4 depicts how the adaptive disclosure with ephemeral visualization helps the user to operate with a complex user interface with overwhelming number of user controls. We depict three stages of ephemeral visualization which describes the progression of visualization through time, where the prominent fields are displayed immediately (subfigure a), while the rest of fields gradually fade-in (subfigure b), and stops at 75% of visibility (subfigure c). In this example, the prominent fields are: *Purchasing Organization, Purchasing Document, Document Category, Delivery Date, Selection Parameters, Season, Item Category* and *Article*.

4.1 Real Implementation of Boulevard in SAP ERP

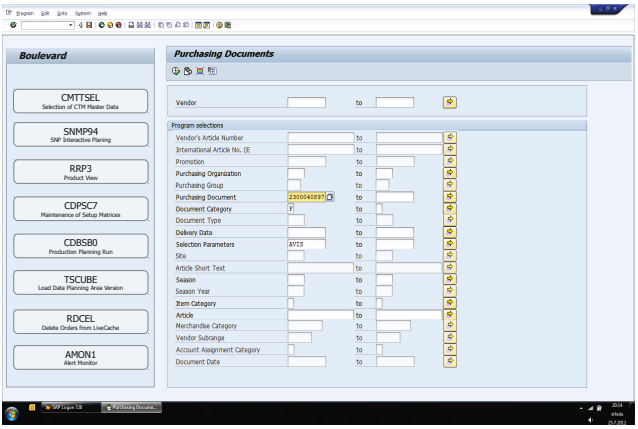
Here we describe briefly how the above mentioned concept could be be implemented. Since source code of SAPGUI application is not publicly available, the SAPGUI cannot be easily extended with Boulevard. We performed a feasibility study that proved that the Boulevard implementation would be possible. Fortunately, we discovered several feasible ways to develop such adaptive SAPGUI with integrated Boulevard adaptive user interface. One of the possible approaches is to use *SAP GUI Scripting API*, which utilizes the original SAPGUI application and through the provided API a two way communication with SAP is possible. In this approach, there is a need of having *SAPGUI*



(a) Phase 1 — visualization start



(b) Phase 2 — visualization in progress



(c) Phase 3 — visualization completed

Fig. 4. Gradual progress of fields fade-in visualization

installed on the client side. Other approach could be to use *SAPGUI FOR HTML*, where the HTML code can be relatively easily analyzed and high-level user interface description can be extracted (with some HTML limitations, e.g. double-click does not trigger anything). The last solution we found, was to use *SAP Automation GUI Interfaces*, which seems as a more complex approach, however, it provides probably a good way to perform since it provides a direct connection to the SAP and does not require SAPGUI installation on the client side.

The above mentioned approaches discussed briefly how the core of the client can be implemented. However, there is also need to maintain Boulevards adaptive data — collected data of how the user operates the application. Such data represent a core information for expert system and are unique for every particular user. The data must be maintained on the server side in order to preserve long-term continuos adaptation. This can be covered by using of a file server or by a native integration of Boulevard into the SAP ERP, where the user data can be stored directly in SAP database.

4.2 Brief Questionnaire Study

We conducted a brief questionnaire study in which six professional SAP ERP users evaluated the above introduced user interface. We created a web page for that purpose. The page contained in addition to brief description and images of the interface also an animated demo where the ephemeral visualization was demonstrated. We used online-based questionnaires (LimeSurvey) to determine subjects' opinion to the proposed user interface. The study was conducted on six participants (all males), aged from 28 to 42 (Mdn=34 years), subjects were not paid for participation. Most subjects had high experience using computers, subjects' experience with SAP ERP system differs from 1 to 14 years (Mdn=6). Particularly, there were two regular SAP ERP users, four users on SAP support level and one developer.

Subjects were asked to evaluate the following statements on the five-point Likert scale (1 = strongly disagree, 5 = strongly agree):

- I found adaptive user interfaces suitable for ERP applications (Mode=4, Mdn=4).
- I do understand clearly the basic principles behind Boulevard (Mode=4, Mdn=4).
- Boulevard is intuitive and predictable (Mode=4, Mdn=4).
- Boulevard is interesting (Mode=5, Mdn=4.5).
- Overall, I like Boulevard (Mode=4, Mdn=4).
- I would like to try Boulevard (Mode = 5, Mdn=5).

Although this brief study produced promising results, it does have some drawbacks. First, we realize that results of study conducted on only six participants may be inaccurate. Second, the study include only subjective, not quantitative measures. However, the preliminary results are encouraging for the further development.

5 Conclusion and Future Work

Since ERP systems are complex applications with complex user interface, it is believed that adaptive user interfaces can help users to operate more effectively and thus

increase productivity which is a crucial factor in enterprise environment. We presented the current adaptive user interface Boulevard implementation in a word processor and the design of the Boulevard integrated in the ERP system. We found the key features of Boulevard suitable for an ERP system, mostly because of the nondestructive adaptation principle that seems appropriate for enterprise applications.

Brief questionnaire study showed promising response from SAP ERP users. We believe that a real implementation for the SAP ERP will be developed soon and the concept presented in this paper will be verified on a prototype by a field user study (both quantitative and qualitative measures). However, there are other features that could also be considered, namely:

Sweeping Back — which is successfully used in Boulevard in word processor. To date, we did not find a way how to beneficially apply this approach to ERP Boulevard. In complex ERP systems, users possibly perceive semantical relationships between transactions differently from each other. Utilization of this approach is under consideration, since "intelligent" layout of items in Boulevard decreases time needed to find and use a desired command.

Workflow Adaptation — which would trace and identify repeated patterns (called episodes) in the user's behavior and aggregate them into "compound" user commands (transactions).

Adaptable Personalization — in enterprise environment it can be sometimes advantageous to predefine Boulevard content and behavior, where the further adaptation can be disabled partially or completely. Predefining Boulevard content should be much less expensive in contrast to customizing the system to a particular user.

References

1. Boudreau, M.-C.: Learning to use erp technology: A causal model. In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS 2003) - Track 8, vol. 8, IEEE Computer Society, Washington, DC (2003)
2. Chou, S.-W., Chang, Y.-C.: The implementation factors that influence the erp (enterprise resource planning) benefits. Decis. Support Syst. 46(1), 149–157 (2008)
3. Dostál, M.: An analysis of usage patterns in utilization of interaction styles. In: HCI International 2011 – Posters' Extended Abstracts, pp. 128–132 (2011)
4. Dostál, M.: On the Differences in Usage of Word Processing Applications. In: Stephanidis, C. (ed.) Posters, HCII 2011, Part I. CCIS, vol. 173, pp. 133–137. Springer, Heidelberg (2011)
5. Dostál, M., Eichler, Z.: Fine-Grained Adaptive User Interface for Personalization of a Word Processor: Principles and a Preliminary Study. In: Stephanidis, C. (ed.) Posters, HCII 2011, Part I. CCIS, vol. 173, pp. 496–500. Springer, Heidelberg (2011)
6. Dostál, M., Eichler, Z.: A research framework for performing user studies and rapid prototyping of intelligent user interfaces under the openoffice.org suite. In: Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems, EICS 2011, pp. 153–156. ACM, New York (2011)
7. Findlater, L., McGrenere, J.: A comparison of static, adaptive, and adaptable menus. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2004, pp. 89–96. ACM, New York (2004)

8. Findlater, L., Moffatt, K., McGrenere, J., Dawson, J.: Ephemeral adaptation: the use of gradual onset to improve menu selection performance. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, pp. 1655–1664. ACM, New York (2009)
9. Gajos, K.Z., Czerwinski, M., Tan, D.S., Weld, D.S.: Exploring the design space for adaptive graphical user interfaces. In: Proceedings of the Working Conference on Advanced Visual Interfaces, AVI 2006, pp. 201–208. ACM, New York (2006)
10. Höök, K.: Designing and evaluating intelligent user interfaces. In: Proceedings of the 4th International Conference on Intelligent User Interfaces, IUI 1999, pp. 5–6. ACM, New York (1999)
11. Hornbaek, K.: Current practice in measuring usability: Challenges to usability studies and research. Int. J. Hum.-Comput. Stud. 64(2), 79–102 (2006)
12. Jameson, A.: Adaptive interfaces and agents. In: The Human-Computer Interaction Handbook, pp. 305–330. L. Erlbaum Associates Inc., Hillsdale (2003)
13. Keeble, R.J., Macredie, R.D.: Assistant agents for the world wide web intelligent interface design challenges. Interacting with Computers 12(4), 357–381 (2000)
14. Kühme, T.: A user-centered approach to adaptive interfaces. In: Proceedings of the 1st International Conference on Intelligent User Interfaces, IUI 1993, pp. 243–245. ACM, New York (1993)
15. McGrenere, J., Baecker, R.M., Booth, K.S.: An evaluation of a multiple interface design solution for bloated software. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves, CHI 2002, pp. 164–170. ACM, New York (2002)
16. McGrenere, J., Moore, G.: Are we all in the same "bloat"? In: Graphics Interface, pp. 187–196 (2000)
17. Oppermann, R., Simm, H.: Adaptability: user-initiated individualization. In: Adaptive User Support, pp. 14–66. L. Erlbaum Associates Inc., Hillsdale (1994)
18. Ragowsky, A., Gefen, D.: What makes the competitive contribution of erp strategic. SIGMIS Database 39(2), 33–49 (2008)
19. Shneiderman, B.: Direct manipulation for comprehensible, predictable and controllable user interfaces. In: Proceedings of the 2nd International Conference on Intelligent User Interfaces, IUI 1997, pp. 33–39. ACM, New York (1997)
20. Singh, A., Wesson, J.: Evaluation criteria for assessing the usability of erp systems. In: Proceedings of the 2009 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists, SAICSIT 2009, pp. 87–95. ACM, New York (2009)
21. Singh, A., Wesson, J.: Improving the usability of erp systems through the application of adaptive user interfaces. In: ICEIS (4), pp. 208–214 (2009)
22. Singh, A., Wesson, J.: The design of adaptive interfaces for enterprise resource planning systems. In: ICEIS (4), pp. 281–286 (2011)
23. Tsandilas, T., Schraefel, M.C.: An empirical assessment of adaptation techniques. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2005, pp. 2009–2012. ACM, New York (2005)

Weekly Quantitative Analysis and Trend Trading in Futures Market

Saulius Masteika[1,2], Kestutis Driaunys[1], and Vera Moskaliova[1]

[1] Vilnius University, Faculty of Humanities, Muitines 8, LT-44280 Kaunas, Lithuania, EU
[2] Vilnius Gediminas Technical University, Faculty of Business Management, Sauletekio al. 11, LT-10223, Vilnius, Lithuania, EU
saulius.masteika@vgtu.lt, kestutis.driaunys@khf.vu.lt, vera.moskaliova@vukhf.lt

Abstract. The main task of this paper is to examine a trend trading strategy in futures market based on quantitative and chart pattern analysis. A contribution of this paper lies in a modified chart pattern related to fractal formation and its application to up-trend trading. Experimental research and time series analysis is on a weekly basis. Trading strategy was tested with the most active futures using historical weekly data records from NYMEX, ECBOT, GLOBEX, ICE and CFE exchanges (2001- 2012). The trend trading strategy has refuted the efficient market hypothesis and has given better returns if compared to overall market index (CRB). The proposed strategy can be attractive for futures traders or hedge funds and be applied as a decision support tool in market analysis.

1 Introduction

During the last decade expansion of electronic trading, application of computational analysis and trading algorithms has attracted attention of many market participants and researchers. The interest in automated trading, high frequency trading and time series analysis shows an increasing belief that new quantitative techniques can be helpful for decision making processes in financial markets [13], [4], [2]. Because of the growing demand for hedging businesses trading volumes from stock and currency markets are increasingly moving to derivative markets [1]. The need for trading algorithms and quantitative analysis in futures markets are in great demand.

A well-known key concept is being often discussed when talking about financial markets: the Efficient Market Hypothesis (EMH). EMH states that the current market price reflects all the information available. So in essence prices reflect all publicly available information and analysis of historical information is unfruitful. Therefore according to EMH it is impossible to get better returns in the market than the market index itself (e.g. CRB index for commodities). There are a lot of well-known publications that partly confirm this hypothesis [17], [6]. However there are also published reports with opposite statements, that EMH is far from the truth [18], [5], [15]. The fact that some high frequency traders or hedge funds can consistently beat the market is a good indication that the EMH may not be just right. Usually the information about profitable algorithms is kept secret and not published by

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 61–68, 2012.

practitioners. Despite the shortage of hard evidences about the profitability of trading algorithms there are some research papers claiming that the application of momentum trend trading and chart patterns can give some useful information to market participants [11], [16], [3], [19]. We are trying to reinforce these statements in this paper. A contribution of this paper lies in a modified chart pattern related to fractal formation and chaos theory and its application to short term up-trend trading strategy. The paper presents the details of making computational analysis and trading algorithm in futures market.

2 The Basic Concept of the Proposed Trading Strategy

The proposed trading strategy is based on up-trend following system and continuous bar chart pattern. Bar chart pattern is an indicator of technical analysis which is mostly applied for short term investments [7].

The proposed continuous chart pattern is related to a fractal pattern and chaos theory, broadly discussed by Benoit B. Mandelbrot and Ph.D. Bill M.Williams [19], [8].

Chaos theory in financial markets is based on assumption that it is impossible to encounter all new incoming information and in order to succeed in the markets you need to want what the market wants or in other words- be in the flow with the trend. And fractal is suggested by B.M.Williams as indicator of the trend momentum. The fractal formation is built from minimum two preceding and two following bars with lower highs. In a buy fractal, we are interested only in the bars' high [19]. A buying signal is generated when the highest price of a fractal is hit. The reasons preventing from development trading algorithms on typical fractals relate to a big variety of formation, signal quality measurement and clear rules for closing positions. Inaccuracies make fractal as a lagging, subjective and complicated indicator. Maybe therefore typical fractal patterns are often discussed in conjunction with other forms of technical analysis, like moving averages, Elliott Waves analysis or MACD indicator [20]. In general, typical fractals are difficult to determine precisely and can be a good decision support tool, but not the basic indicator itself. Therefore a decision to search for modified, more determined and at the same time still conformable with trend trading and chaos theory chart pattern was made.

2.1 Chart Pattern for Trading Strategy

The proposed short term chart pattern is composed of only 3 consecutive chart bars. The proposed pattern uses a modified truncated fractal of three bars to generate a buy signal and so increases prediction sharpness because of only the latest data used for decision making. A fractal occurs when there is a pattern with the highest high in the middle and one lower high on each side, as it can be seen in some examples in Fig.1.

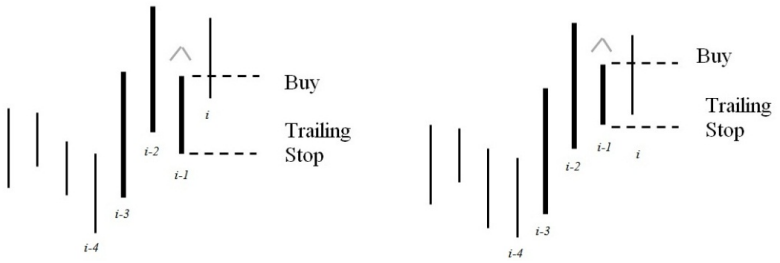


Fig. 1. Examples of the proposed chart pattern

The trading strategy opens a long position when the current price (*i*) tops the previous bar's (*i-1)* highest price. An open price is set as the bar's (*i-1)* highest level plus a tick size. Tick size is a minimum possible price increment (more details in the forthcoming Table 1). If current price bar (*i*) opens higher than (*i-1*) and forms a gap, a buy signal is identified as an opening price. Deciding the appropriate time to close the position is just as important as determining the best time to open it. Our strategy suggests that a trailing stop order should be set as the lowest level of the bar (*i-1*) price minus a tick size. The trailing stop price is adjusted and moved to the next bar lowest price each time the new bar is formed. It lets profits run and cut losses in response to the market price changes. If there is a sudden crash in the market and the next bar opens with a gap down, sell price is set as an open price of this following bar. It should also be noted that the opening and closing of the position cannot be fulfilled on the time frame of the same bar.

2.2 Signal Strength of the Trading Pattern

Most of the time back-test analysis is implemented only with historical data from the preselected particular contract. This approach can be applied but it is static, subjective and lacks flexibility to all the time changing instruments and their behavior in the markets. The research results can often be predetermined in advance by taking only a particular contract or the most suitable time period for analysis. In order to avoid such subjectivity when doing quantitative analysis the quality of signals and ranking techniques of contracts must be considered [9].

The proposed trading strategy is looking for futures contracts, which during two day's period increased in a price the most, and the lowest price of the last bar is higher than the one a month ago, i.e. the lowest price of the period *i-1* is greater than the lowest price of the period *i-20*. The strength of the signal and its rank for the trade is measured using the following equation:

$$R = \frac{C_{i-2} \times 100}{C_{i-4}} - 100 \tag{1}$$

In the equation (1) *R* represents the rank of a particular contract and *C*- the closing price of the period *i*. The biggest increase in a price of a particular contract during the period between *i-4* and *i-2* also means the best quality and the highest rank. If the

chart pattern is formed with several contracts at the time, the one with the highest rank is chosen for a trade.

3 Experimental Investigations

The efficiency of the proposed trend trading strategy was tested using historical data collected from the automated electronic futures exchanges. The weekly time series were collected from exchanges like GLOBEX, NYMEX, ECBOT, CFE, ICE and ICE-NYBOT from August 10, 2001 till February 3, 2012. The most active and liquid contracts from each sector of futures market were included in the tests. Most active futures were taken from these sectors: Energies, Metals, Grains, Financials (Interest rates), Indices, Currencies, Softs and Meats. Table 1 shows the list of futures included in experimental research.

Table 1. Futures and their specifications for back test analysis

Symbol	Name	Tick size	Full Point value	OIMR	Commis-sions
CL	WTI Crude Oil	0.01	1000	8775	$2.32
NG	Henry Hub Natural Gas	0.001	10000	3713	$2.32
HO	Heating Oil #2	0.0001	42000	9534	$2.32
GC	Gold	0.1	100	6075	$2.32
SI	Silver	0.1	50	25920	$2.32
HG	Copper	0.05	250	5852	$2.32
PL	Platinum	0.1	50	4375	$2.32
ZC	Corn	0.25	50	3260	$2.68
ZS	Soybean	0.25	50	4388	$2.68
ZW	Wheat	0.25	50	4374	$2.68
ZL	Soybean oil	0.01	600	1688	$2.68
ZM	Soybean meal	0.1	100	2363	$2.68
GE	Eurodollar	0.0025	2500	608	$2.06
ZN	10 Year Treasury Note	0.015625	1000	2049	$1.43
ZB	30 Year Treasury Note	0.015625	1000	3078	$1.43
ZT	2 Year Treasury Note	0.0078125	2000	675	$1.43
ES	E-mini S&P500	0.25	50	5000	$2.01
NQ	E-mini Nasdaq 100	0.25	20	3500	$2.01
YM	DJIA Index Mini	1	5	5000	$2.01
TF	Russell 2000 mini	0.05	100	4375	$2.01
VIX	Volatility Index	0.05	1000	4000	$1.87
EUR	Euro FX	0.0001	125000	5616	$2.47
JPY	Japanese Yen	0.0001	125000	3375	$2.47
GBP	British Pound	0.0001	62500	2147	$2.47
AUD	Australian Dollar	0.0001	100000	3213	$2.47

Table 1. (*Continued*)

CAD	Canadian Dollar	0.0001	100000	2484	$2.47
SB	Sugar #11	0.01	1120	3557	$2.62
CT	Cotton #2	0.01	500	8400	$2.62
OJ	Orange Juice	0.05	150	1798	$2.62
KC	Coffee	0.05	375	7787	$2.62
LB	Lumber	0.1	110	2745	$2,76
LE	Live Cattle	0.025	400	2106	$2.76
HE	Lean Hogs	0.025	400	2160	$2.76
GF	Feeder Cattle	0.025	500	3119	$2.76
DA	Class III Milk	0.01	2000	675	$2,76

In Table 1 the futures listed are the most liquid and all together take the biggest part of trading activity in the North American futures market. Basic specifications necessary for calculating risk, profitability and position size when back-testing strategies are also presented in Table 1. 'Full point value' and Overnight Initial Margin Requirements (OIMR) were used to calculate the returns and the size of position. Tick size was applied defining exact price for opening and closing positions. Currency and commissions show the costs per contract to make a trade. The successive equation presents the profit/loss calculation used in the analysis:

$$P/L = N \times ((P_C - P_O) \times FPV_i - 2 \times Comm) - Int \tag{2}$$

In the equation (2) P_c and P_o represent closing and opening prices of a particular trade. *FPV* shows the Full Point value of a particular contract *i* (see: Table 1). The variable *Comm* shows the value of commissions per trade. The variable *Comm* is multiplied by 2 because there is a commission charge both for opening and closing position. The variable N represents the number of contracts and *Int* shows the value of interest paid on margin.

The following equation is applied doing experimental investigation when evaluating the size of a single trade:

$$PS = \frac{Ie \times \frac{Rr}{100} - 2 \times Comm}{OIMR_i} \tag{3}$$

In the equation (3) the calculation of position size - *PS* is presented. The variable *Ie* shows the size of initial capital and *Rr* (Risk ratio)- the percentage size of risk taken if compared with total equity. *Rr* depends on the size of initial equity and also on the forwardness toward risk taking. It is common in the industry that larger capitalized futures traders tend to risk rarely more than 1-2 % per trade of total capital [14]. Therefore in our research Rr coefficient equals to 1.5%.

Backward adjusted data series were used for analysis. Backward adjusted data uses the actual prices of the most recent contract with a backward correction of price discontinuities for successive earlier active delivery months [10].

4 Experimental Results

The strategy was back tested and calculations were carried out applying software of technical computing- MatLab [12]. Considering that trading costs consist of commissions and also possible trade execution slippage, the size of a slippage was 8 ticks on every trade, which is quite a big size knowing that most of the time the spread between bid and ask is not greater than 2 ticks.

The experimental results of the strategy are presented in the following figure.

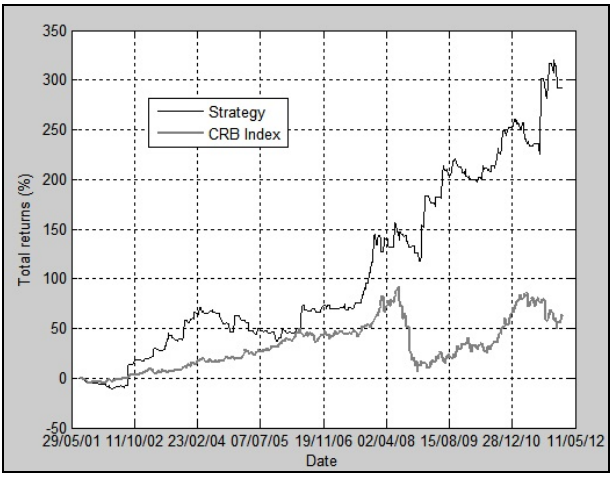


Fig. 2. Total returns using trading strategy

Fig. 2 shows the dynamics of total returns when using the proposed strategy. The Fig.2 shows that the proposed trading system had outperformed the benchmark (Commodity Research Bureau (CRB) Index ICE(NYFE) Weekly) and the total return at the end of testing period was nearly 300%. Fig.2. also shows that despite huge fluctuations in CRB Index the trading strategy gives rather stable results. Financial crisis in 2008 and the market crash, when CRB Index was tumbling nearly 50%, didn't impact a lot the overall performance of the strategy. These results confirm the capability of the chart pattern in different market conditions.

The following figure shows additional results of the research related to comparison of weekly returns.

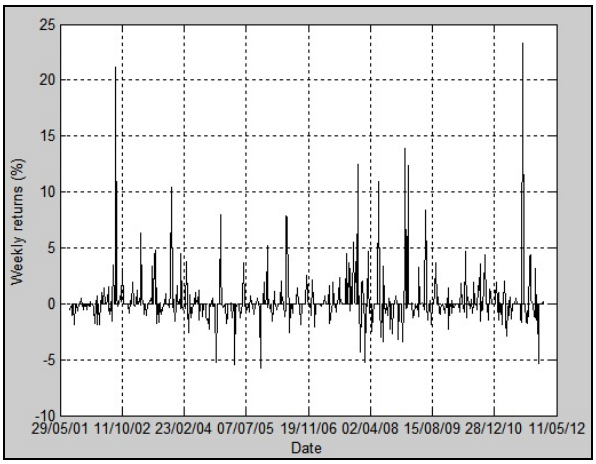


Fig. 3. Weekly returns

As we can see in Fig.3 weekly returns also show rather a good proportion, e.g. the biggest drawbacks are only up to 5%, while some profitable weeks generate over 20% returns. To sum up, the results confirm the capability of the trading strategy and can be implemented in futures market.

5 Conclusions

The paper reviewed the proposed trading strategy in futures market based on chart pattern. The research was carried out on weekly historical data of the most active futures contracts in the USA markets. The results clearly contradict the statements of Efficient Market Hypothesis and have given significantly better returns if compared to CRB index. The research has also shown rather stable results considering global economic fluctuations over the last decade, e.g. nearly 50% CRB Index decrease in 2008. Characteristics of contracts necessary for computational and technical analysis and the equations related to profitability and position size calculations trading futures were also discussed in this paper.

The conclusion can be made that the proposed strategy gives some promising results and can be attractive for hedge funds when managing the risks associated with price fluctuations or futures market participants who intend trading short term strategies, especially when volatility in markets increases. Despite good preliminary results more extensive tests with the proposed methodology should be done (different markets, longer time frames, downtrend and short selling opportunities) to evaluate more precisely the whole domain of the possible applications of this strategy.

Acknowledgment. This research as Fellowship is being initiated by the European Union Structural Funds project ”Postdoctoral Fellowship Implementation in Lithuania” within the framework of the Measure for Enhancing Mobility of Scholars and Other Researchers and the Promotion of Student Research (VP1-3.1-ŠMM-01) of the Program of Human Resources Development Action Plan.

References

1. Will, A.: Record Volume 2010 (Annual Volume Survey). Futures Industry, 12–29 (March 2011)
2. Ehlers, J.F.: Cybernetic analysis for stocks and futures: Cutting-Edge DSP Technology to Improve Your trading, p. 256. Wiley (2004)
3. Friesen, G.C., Weller, P.A., Dunham, L.M.: Price trends and patterns in technical analysis: a theoretical and empirical examination. Journal of Banking & Finance 33(6), 1089–1100 (2009)
4. Izumi, K., Toriumi, F., Matsui, H.: Evaluation of automated- trading strategies using an artificial market. Neurocomputing 72(16-18), 3469–3476 (2009)
5. Jarret, J.E., Schilling, J.: Daily variation and predicting stock market returns for the Frankfurter Borse (stock market). Journal of Business Economics and Management 9(3), 189–198 (2008)
6. Kan, D., Anreosso-O'Callaghan, B.: Examination of the efficient market hypothesis-the case of post-crisis Asia Pacific countries. Journal of Asian Economics 18(2), 294–313 (2007)
7. Kirkpatrick II, C.D., Dahlquist, J.R.: Technical analysis: The Complete Resource for Financial Market Technicians, 2nd edn., pp. 362–386. FT Press (2010)
8. Mandelbrot, B.B., Gomory, R.E., Cootner, P.H., Fama, E.F., Morris, W.S., Taylor, H.M.: Fractals and Scaling in Finance: Discontinuity, Concentration, Risk, p. 561. Springer (2010)
9. Masteika, S.: Short Term Trading Strategy and Trend Trading in NASDAQ Biotechnology Stock Market. Transformations in Business & Economics 9(2(20)), 170–178 (2010)
10. Masteika, S., Rutkauskas, A.V., Alexander Janes, A.: Continuous futures data series for back testing and technical analysis. In: Conference Proceedings, 3rd International Conference on Financial Theory and Engineering, vol. 29, pp. 265–269. IACSIT Press (February 2012)
11. Miffre, J., Rallis, G.: Momentum strategies in commodity futures markets. Journal of Banking & Finance 31(6), 1863–1886 (2007)
12. Pratap, R.: Getting started with MatLab: a quick introduction for scientists and engineers, p. 288. Oxford University Press (2009)
13. Schmidt, G., Mohr, E., Kersh, M.: Experimental Analysis of an Online Trading Algorithm. Electronic notes in Discrete Mathematics (36), 519–526 (2010)
14. Schwager Jack, D.: The New Market Wizards: Conversations with America's top traders. Marketplace Books (First Collins Business edition 2005), p. 614 (2008)
15. Simutis, R., Masteika, S.: A hybrid stock trading strategy based on fundamental and technical analysis. In: Proceedings Databases and Information Systems: Seventh International Baltic Conference, Baltic DB&IS, pp. 113–121 (2006)
16. Szakmary, A.C., Shen, Q., Sharma, S.C.: Trend-following trading strategies in commodity futures: A re-examination. Journal of Banking & Finance 34(2), 409–426 (2010)
17. Uri Noel, D., Jones Jonathan, D.: Are the markets for financial assets efficient?: Evidence for the USA, 1974-1988. Economic Modeling 7(4), 388–394 (1990)
18. Warwick, B.: Event Trading: Profiting from Economic Reports and Short-Term Market Inefficiencies, 275 p. McGraw-Hill (1996)
19. Williams, B.M., Williams, J.G.: Trading chaos: maximize profits with proven technical techniques, 2nd edn., p. 228. John Wiley & Sons, Inc. (2004)
20. Williams, B.M.: New Trading Dimensions: How to Profit from Chaos in Stocks, Bonds, and Commodities, p. 260. John Wiley&Sons, Inc. (1998)

Dynamic Simulation of Pension Funds' Portfolio

Marius Liutvinavicius and Virgilijus Sakalauskas

Vilnius University, Lithuania
marius.liutvinavicius@gmail.com,
virgilijus.sakalauskas@vukhf.lt

Abstract. The article investigates the dynamics of pension fund portfolio by using adaptive Powersim simulation models. Many countries use systems of investment to pension funds for ensuring older peoples' financial stability and encourage their participation by partial tax relief. There is a need to create advanced tools that could help the unprofessional investors, who make up the majority of pension fund customers, to make well informed decisions and reduce the potential risk. Financial companies use a variety of spreadsheets based on parameters describing investors and markets, which remain constant during all calculation process. Due to high volatility of the financial markets, personal investment power and change in tax relief system the actual portfolio dynamics highly deviates from the initial forecast in the initial stage of the investment. Description of the tool, which was created with the Powersim software package, has been presented in the article. The model, proposed by the authors, introduces adaptive and dynamic variables used in the portfolio simulation. Volatile values of return rate, contributions and fees are used in calculations. The results of research show the advantages of such method over traditional spreadsheets.

Keywords: pension funds, volatile return rate, process simulation.

1 Introduction

Many countries have a lot of problems while trying to ensure older people's health and financial stability. This is related to the aging population, lack of social guarantees and lack of measures to ensure them.

The European Commission (2010) in [1] emphasizes a tendency to shift from one-pillar to multi-pillar pension system. This change has resulted in a majority of EU countries' choice to reduce the current pension funding from the state social insurance system and to focus on additional private pension funds.

In 2004 private pension funds started to operate in Lithuania. Currently, Lithuania has an old-age pension system of three pillars. People can choose from nine additional voluntary pension scheme (third pillar) pension funds. But there are several problems and shortcomings: low financial literacy, failure to choose optimal pension fund, emotion-based migration between funds. In this paper we are going to describe the simulation tool that could help understand potential risks and make better decisions.

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 69–80, 2012.

Both Lithuanian and foreign researchers study investment instruments. Research works mainly focus on investigation of social systems, and portfolio optimization. Summarizing we can state that researchers try to find the ways how to make simulation of real life processes more accurate. The ability to evaluate different scenarios by using advanced technologies could prevent from real problems. For example A. Rutkauskas in [2] uses the imitation technologies technique for an adequate portfolio development and management.

The [3] states, that due to the instability of the world economy, increased longevity, higher healthcare and long-term care costs, an adequate retirement income level is a daunting challenge today. The author uses the dynamic stochastic simulation to identify the optimal portfolio allocations. T. Ermolieva in [4] analyzes optimization-based approaches for the social security simulation model under demographic and economic uncertainties. The design of optimal robust strategies is achieved by an adaptive simulation-based optimization procedure defined by non-smooth risk functions.

Risk management is one of the most important issues. The [5] suggests the decision support model that describes sustainable management of pension-funds in the strategic planning of the available asset and liability policy instruments. The main characteristic of the approach is that the relevant risk-drivers are modeled by scenarios, rather than by probability distributions. The author investigates the scenario generation methodology, and how the scenarios could be used by pension-fund managers to simulate and improve asset/liability strategies until the strategy is identified as good by all who carry responsibility for the pension-fund. The tool, specially designed to simulate investing to the third pillar pension funds, has been presented in the paper.

Researchers also analyze the methods for evaluating investment funds. In addition to determination of traditional indices [6] propose to incorporate into evaluation the identification of key socioeconomic factors, qualitative analysis and complex quantitative assessment of the factors on the basis of the given models. Usually, constant values of salary and contribution to funds are used in spreadsheets. But these values are highly dependent on economic changes and investor opportunities, so in our model these values vary inconstantly.

Financial institutions also try to create methods and tools for decision support and better risk management. They use a variety of spreadsheets to calculate how much money a person can accumulate. It is important to use such tools which can simulate scenarios as closely as possible to real processes. New technologies allow this.

The main variables used in spreadsheets are the following: duration of accumulation process, average annual return rate, contribution amount, contribution and asset fees. The problem is that the values of these variables remain constant during all calculation process.

In the previous research [7] the authors found that in order to accurately simulate the accumulation of high-risk investments, a volatile return rate should be used. Especially when there are large variations in fund's unit value and contributions are paid irregularly. There may be a situation when, for example, during the first 5 years fund's annual return rate is 6%, but contributions to the fund are small. Meanwhile,

over the next two years large sums of money may be paid, but the fund will suffer a sharp 20% drop in units' value. Then, the declared average annual growth of fund's unit value will not match the actual return of investments. In order to investigate such cases, the model should be able to generate a random variation of unit's value, instead of using constant return rate.

The approach, which was implemented by using the Powersim software package, has been described in the article. The model, created by the authors in such environment, tries precisely to imitate the operations of pension funds. Volatile values of return rate, contribution amount and fees have been used in calculations. Simulation conditions can be adapted to specific investors' requirements by changing values of variables describing salary, duration of accumulation and other factors. Various scenarios can be simulated with the proposed tool in order to reduce risk and make a better decision.

We hope that this methodology can give more accurate results than traditional calculation methods.

The Lithuanian pension fund system has been described in the second section of the article. Simulation environment and methodology have been presented in the third section. In the fourth section two case studies have been made by using the proposed simulation technique.

2 Third Pillar Pension Funds in Lithuania

Lithuania is among those countries (like Bulgaria, Estonia, Latvia, Hungary, Poland, Romania, Slovakia, Sweden), which recently have reshaped their statutory systems to include a tier of mandatory funded, private pension schemes and financed these by shifting parts of the overall pension contribution away from the pay-as-you-go scheme. In most of these countries significant parts of the future adequacy of pensions is set to be based on these schemes which are expected to contribute to poverty avoidance as well as adequate income replacement [8].

Lithuanian pension system reform was performed in 2004. Currently, Lithuania has a three-pillar pension system [9]:

- The first pillar - public pension. It will be composed from future employees' taxes paid to SoDra (State Social Insurance Fund Board).
- The second pillar - working people accumulate part of their contributions to SoDra in their private pension accounts.
- The third pillar - people voluntarily invest their own money. The state encourages this through tax relief. People can take their money back at anytime, but in this case they should withdraw from tax relief.

In June 30, 2011, Lithuania had 9 additional voluntary pension scheme (third pillar) pension funds managed by 5 companies. At the same time the third-pillar pension funds assets were 98.4 million Lt. The number of participants amounted to 25.4 thousand (see [10]).

Asset management companies offer various spreadsheets that can help to calculate the expected amount of money to be accumulated. Table 1 shows the variables used in such spreadsheets and their default values.

It should be noted that all spreadsheets use similar methods. In some of them initial variables are entered directly (e.g. accumulation time). In others they are calculated by the spreadsheet using additional variables (e.g. date of birth, age up to which contributions are paid).

Table 1. Variables used in spreadsheets

Institution/ input variable	SEB[1]	MP Pension Funds[2]	DNB[3]
Annual return rate	8%	7% (unavailable to change)	7%
Contribution	100 Lt	spreadsheet offers it's own suggestion according to income	100 Lt
Duration of investment	-	-	20 years
Periodicity of contributions	1 month	1 month (unavailable to change)	1 month
Contribution fee	3%	3% (unavailable to change)	3% (unavailable to change)
Date of birth	1970	-	-
Age at the end of investment process	55 years	-	-
Income	10000 Lt	+	-
Age	-	+	-
Gender	-	+	-

Traditional spreadsheets use the same value of contributions for the entire accumulation period. The expected average return rate and tax rates remain constant too. These assumptions can lead to inaccurate calculation results.

3 Simulation Environment and Methodology

This section presents a new methodology that we suggest using for imitating a real process of investment to the third-pillar pension fund. It's based on system dynamics. The model contains variables that affect each other in time.

[1] http://fin.seb.lt/vbfin/calc/pensionCalculator.fw?lang=lt
[2] http://www.mppf.lt/main.php/id/87/lang/1
[3] http://www.dnb.lt/lt/skaiciuokles-privatiems-klientams/papildomos-pensijos-skaiciuokle/

As in traditional spreadsheets the investor specifies the investment period, the salary and the part of it that should be transferred to fund. The default frequency of contributions is one month, but this parameter can be changed. The investor can also specify the rates of fees. The tool simulates the changes of portfolio value depending on these factors and another important variable – funds' unit price. Unit prices are loaded from the file and may correspond to real historical data or be randomly generated.

Only the period of investment remains constant during all simulation. Funds' unit value, amount of contributions and rates of fees randomly fluctuate during the simulation. The proposed tool lets determine not only the initial values of contributions and fees, but also the limits they might change in the future. For example, you can specify that a monthly payment should fluctuate within 25% compared to the baseline, and the rates of fees should vary between 1-2.5% and change every 3 years.

The generalized model in Figure 1 describes the components of this process.

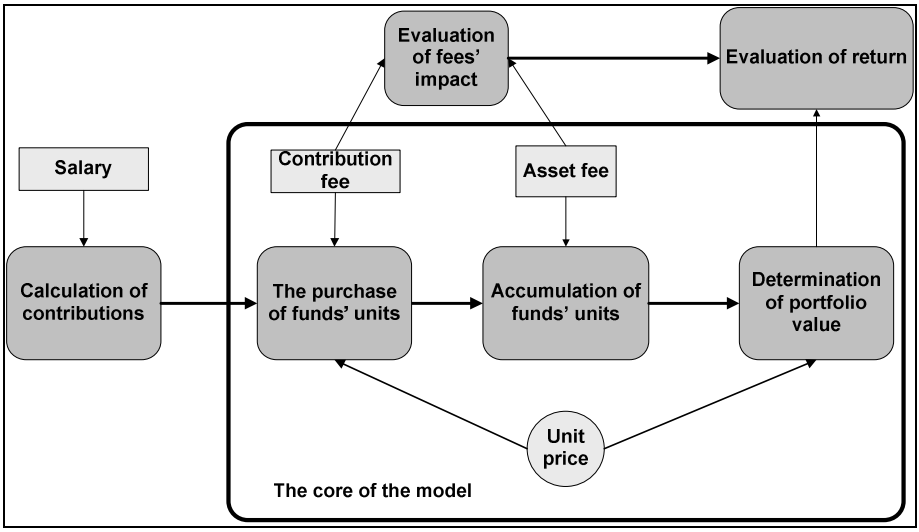


Fig. 1. The model of accumulation process

First of all, one module calculates the monthly contributions to the fund as part of investors' salary. The contribution fee is deducted from each payment. The remaining sum of money is used to buy the units of fund for the current units' price. The units are accumulated in personal account. The portfolio value depends on the fund's unit value. Each month asset fee is paid to asset Management Company. Finally, the impact of fees is investigated and the final return of all accumulation process is determined.

The model has been implemented using the Powersim software environment. The model is specially designed to simulate the accumulation in accordance to laws and

procedures applicable in Lithuania. Generalized model in Figure 1 is explained in detail in the next 4 figures that are made in the Powersim environment.

Simulation begins with these steps:

- Each month a certain part of salary is transferred to the fund: the initial payment (Lt) = salary (Lt) * contribution rate (%).
- Contribution fee is deducted from the initial payment: monthly contribution after contribution fee (Lt) = initial payment (Lt) - initial payment (Lt) * contributions fee (%).
- Each month the transferred sum of money is used to buy the funds' units for the current day price (units purchased per month = contribution per month / unit price).

The realization of funds' units purchase is presented in Figure 2.

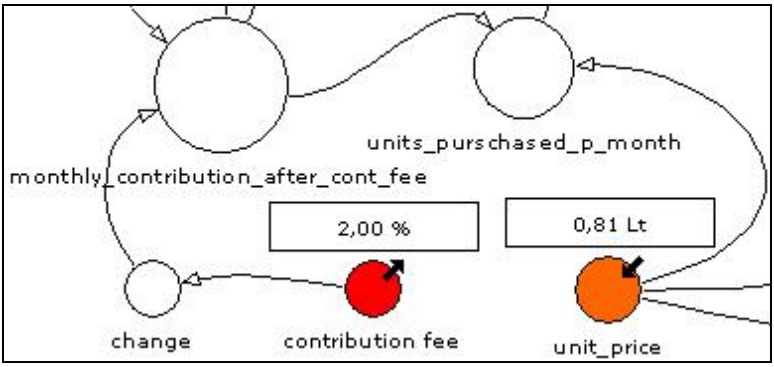


Fig. 2. The purchase of funds' units

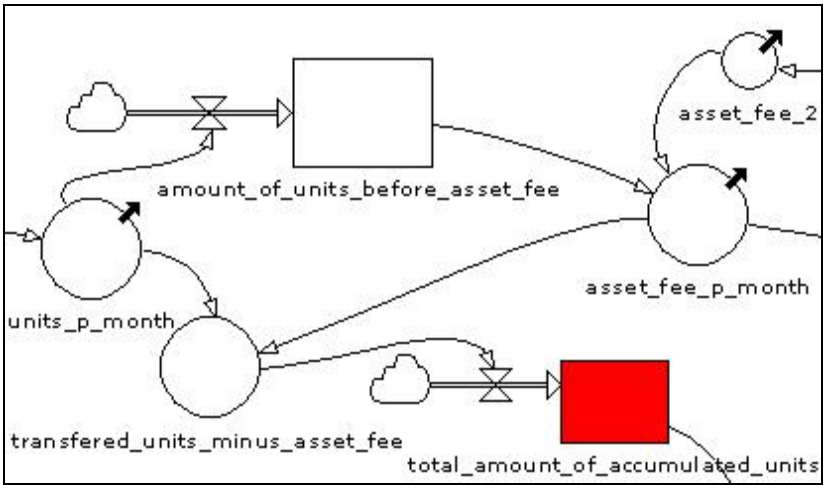


Fig. 3. The accumulation of funds' units

Figure 3 presents the part of model that realizes accumulation of funds' units. New units are added to the previously purchased units. Each month asset fee is deducted from total portfolio value by reducing the amount of fund's units. The total amount of accumulated units increases each month.

The actual portfolio value can be determined at any time. This is done by multiplying the number of units by the current unit price. Unit price is one of the key input variables in the model. It allows simulating the changes of investment returns. Simulation tool can use real historical data as well as randomly generated funds' unit values. Traditional spreadsheets use constant annual average return rate and give only final results. With this tool it is possible to analyze the changes of portfolio value during all investment period. The use of changing units' value realizes volatile return on investment rate. This part of model is presented in Figure 4.

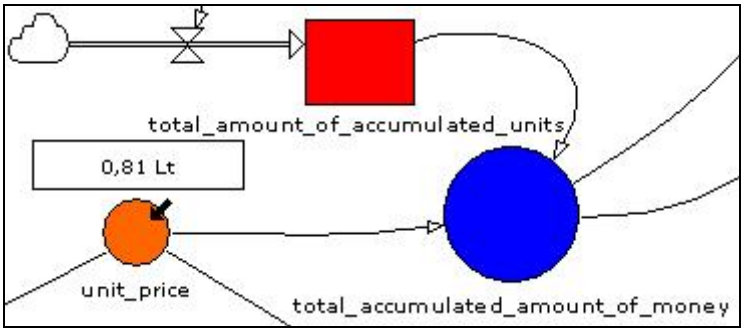


Fig. 4. The determination of portfolio value

Return of investment is calculated by deducting all contributions from the total accumulated amount of money (see Fig.5).

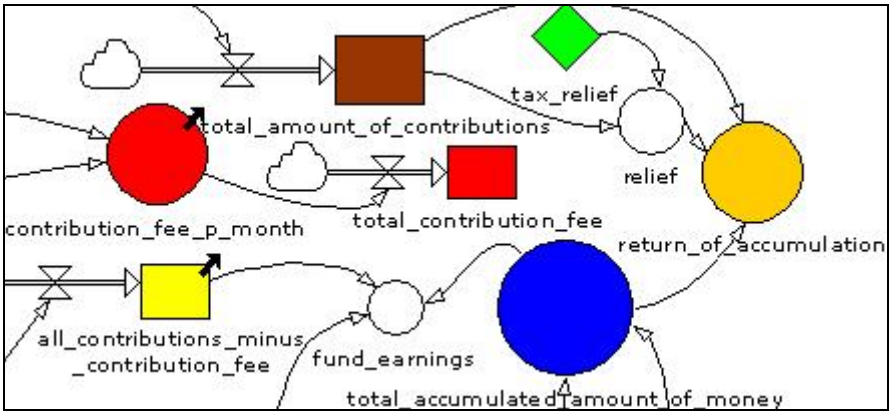


Fig. 5. Evaluation of fees' impact and investment return

It is very important to assess the impact of fees. Unprofessional investors often pay more attention to possible return rates rather than actual rates of fees. But there is common situation when total amount of fees exceeds investment returns and investor suffers losses.

Investing in the third pillar pension fund is similar to investing in life insurance funds and even traditional funds. Only structure of fees and ability to take advantage of tax relief differ in these cases. Basically, our simulator can be adapted to any kind of funds.

4 Case Simulations

Two cases of typical problems that investors might encounter will be presented in this section. Simulation tool is used to analyze and display portfolio value changes in different situations.

In the first case we encounter with irregular growth of portfolio units' value. There is some key input variables used for simulation: duration - 21 years, initial salary - 4,000 Lt, part of the salary paid to the fund - 10%. Contribution rate changes every year and varies randomly within 20%. The initial contribution fee - 2%, it varies for the entire period between 2% and 2.5%. Asset fee is 2%. We suppose that all random variables have normal distribution with varied mean and standard deviation.

For simulation process is very important truthfully imitate the funds' unit value variation over time. Figure 6 shows changes of funds' unit value that are generated by normal distribution. If average return rate is used in calculations, then unit value graph is linear. It does not reflect reality. In our case, the changes of units' value fairly reproduce the real process and uses normal randomness.

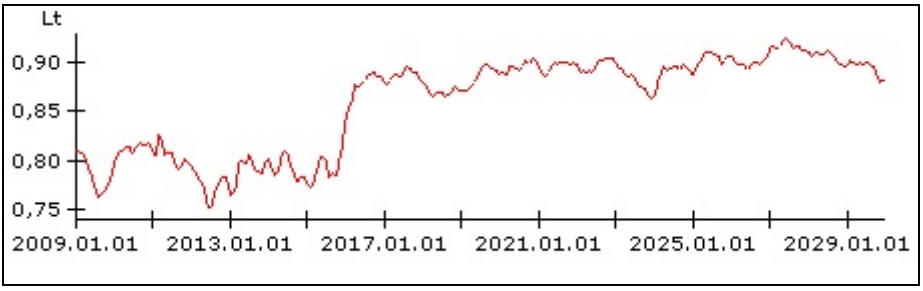


Fig. 6. The change of fund's unit value

Figure 7 shows the evolution of the total portfolios value when volatile values of return rate, contributions and fees are used for simulation. Instability of variables leads to irregular changes of portfolio value.

Calculations show that, despite the fact of increased fund's unit value by 10% during all period, the total accumulated amount of money is lower than the total amount of contributions. Although 122162 Lt was transferred to the fund, the final portfolio value is only 96.847 Lt. This loss occurs due to irregular change of unit value and unequal payments. Growth of value was higher in the first half of the period, but then stopped. Meanwhile, the asset fee has been paid from the increasing total portfolio value. Even if 18.324 Lt is got as tax refund, the final loss would be 6991 Lt.

Figure 7 illustrates such situation, but this is only one possible scenario. By changing various input parameters it is possible to obtain different nonlinear variations of portfolio value.

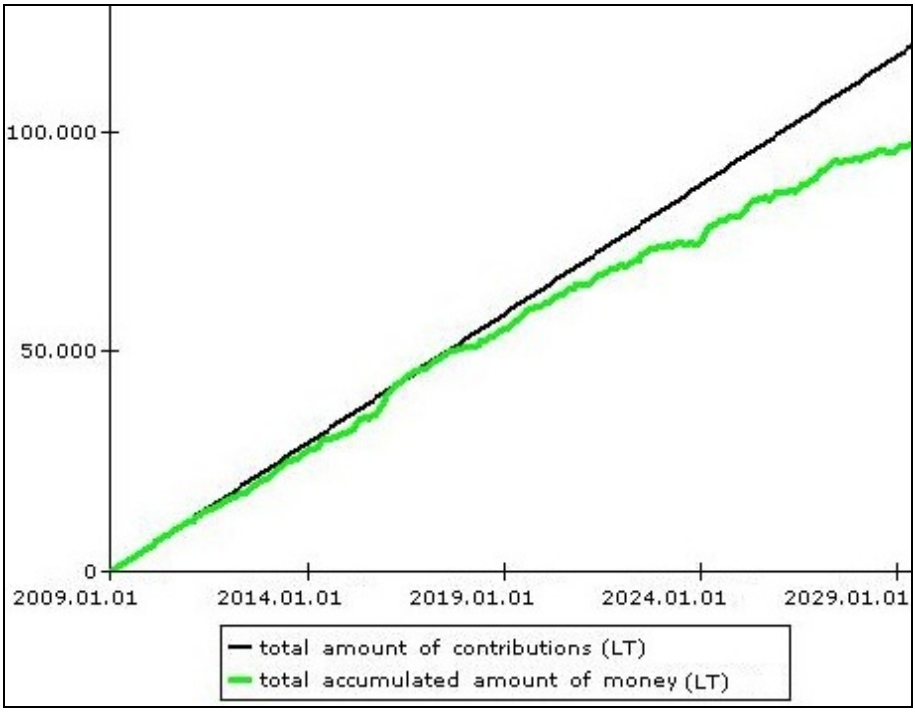


Fig. 7. Simulation results

In the second case we tried to investigate investors' behavior in crisis time. For this purpose we used funds' unit value variation adopted from real historical data of SPDR S&P 500 exchange traded fund from August 2006 to February 2012. This data is used to simulate accumulation process during the last financial crisis (Fig. 8).

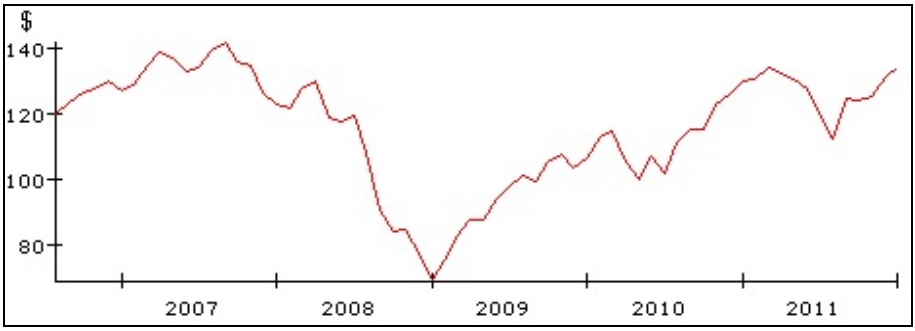


Fig. 8. The change of SPDR S&P 500 fund's unit value 2006-2011

All input values are the same as in the previous case, only duration and change of unit value differ. The simulation results are shown in Figure 9. In situation A contributions remain unchanged during all period of time while in situation B investor reduces monthly contributions from 400 Lt to 50 Lt at the peak of crisis.

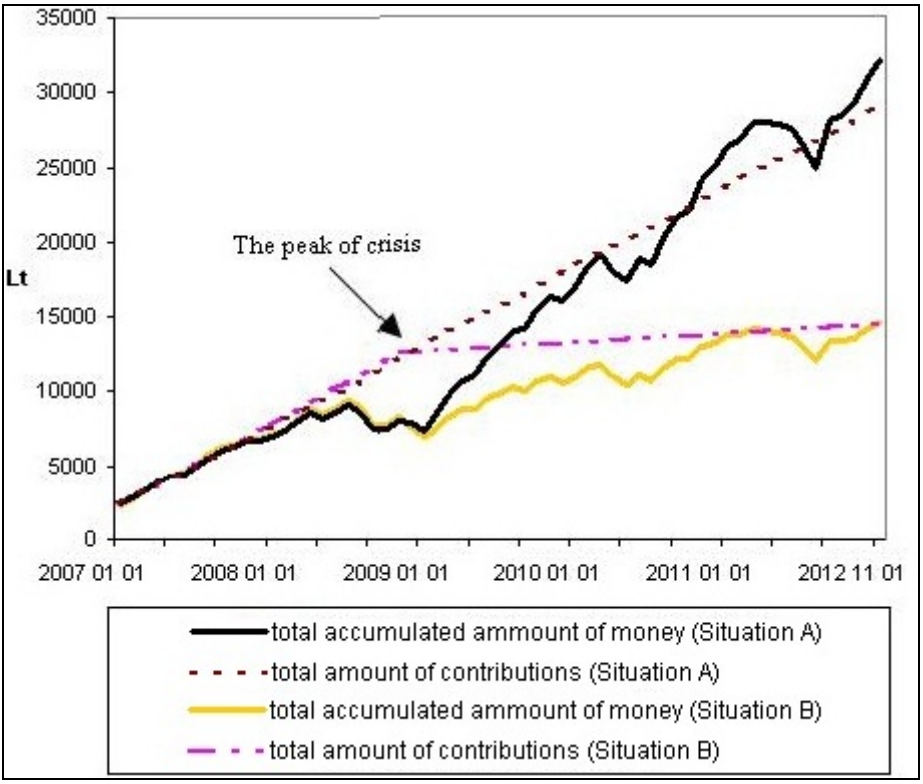


Fig. 9. Simulation results

This figure illustrates one of the most common situations, when portfolio value falls sharply during the peak of crisis. The other studies show that most unprofessional investors sell their financial asset exactly at this time and suffer losses. The results of our simulation let us conclude that if unchanged contributions had been paid during all period of time, then portfolio value would have reached the value of total amount of contributions at the end of 2010 and even exceed it in 2011 (Situation A in Fig. 9). It's important to understand that at the peak of crisis units are bought at a lower price. When units' value increases, it not only compensates the previous losses of the investments made before crisis, but also makes gains for investments made during and after it. If monthly payments had been reduced to 50 Lt at the peak of crisis (Situation B in Fig. 9), then investor wouldn't have suffered losses, but the final portfolio value would have been half the size of portfolio value in situation A.

These two examples show how the proposed simulation technique allows investigating different situations and monitoring the changes of portfolio value depending on various factors. Better understanding of various factors could reduce the risk of investment.

In order to better evaluate the advantages and disadvantages of the model more case studies should be done. The model should be tested in different conditions (markets, periods, etc.). It should be also evaluated by unprofessional investors, because they are the target users of proposed tool.

5 Conclusions

The implementation of new three-pillar pension system in Lithuania brought in new needs for creating optimal investment portfolio and using advantage of tax allowance.

As the majority customers of the pension fund investment system lack financial literacy and computational skills for making information-based decisions and managing personal finances, there is a great need for professional and intelligent simulation models and tools.

The existing tools used by the pension fund companies are based on creating spreadsheets which use static parameters of customers' demographic information, their payment capacity, taxes and expected contribution rate. These tools have high marketing impact for acquiring new customers, but they become highly imprecise in the later periods of investment to personal pension portfolio.

The designed model serves as an innovative tool for simulation of pension fund investment strategies. The dynamic variables of volatile return rate, personal contributions capacity, risk of changes in taxes, government allowances and fees are applied instead of static parameters.

An important advantage of the described simulator is the ability to see not only the final portfolio value, but also all its development. This provides a better understanding of the accumulation process itself. The suggested simulation tool can assist unprofessional investors to raise their level of financial literacy and to better assess the potential risks.

The experimental research was performed by applying the Powersim software. Variables were generated as random normal distribution. The results present high influence of the introduced variables and their dynamic characteristics in the highly volatile markets and instability of tax system. The simulation results based on the characteristics of Lithuanian market showed that the spreadsheets using static parameters presented highly overestimated return value, preventing customers from timely and optimal decision of changing investment strategy or the type of pension fund.

The model relatively well imitates the real process of investment to the third pillar pension funds. But there is a need to make the simulation tool as much user-friendly as possible, so that unprofessional investors could make appropriate decisions according to the simulation results.

References

1. European commission: Green paper towards adequate, sustainable and safe European pension systems, Brussels (2010)
2. Rutkauskas, A.V.: Adequate investment portfolio anatomy and decisions applying imitative technologies (2006) ISSN 1392–1258
3. Klimaviciene, R.: Determination of asset allocation for a sustainable retirement portfolio using a dynamic stochastic simulation. Social Sciences Studies 3(7), 59–79 (2010) ISSN 2029–2244
4. Ermolieva, T.: Simulation-based optimization of social security systems under uncertainty. International Institute for Applied Systems Analysis, Schlossplatz (2004)
5. Boendera, G.: A hybrid simulation/optimisation scenario model for asset/liability management. European Journal of Operational Research 99, 126–135 (1997)
6. Joksienė, I., Zvirblis, A.: Principles of evaluation of the impact of economic and social factors on investment funds. Current Issues of Business and Law 6(2), 335–348 (2011) ISSN 1822-9530 print / ISSN 2029-574X online
7. Liutvinavicius, M., Sakalauskas, V.: Reasearch of factors, affecting pension funds effieciency. Social Technologies 1(2), 328–343 (2011) ISSN 2029-7564
8. Private pension schemes: their role in adequate and sustainable pensions. European Commission. Directorate-General for Employment, Social Affairs and Equal Opportunities (2010) ISBN 978-92-79-14657-2
9. Andriukaitis, D., Čeponytė, Z., Jurevičienė, D., Kriščiukaitytė, K., Vaičiulis, M., Vaitkunskienė, R., Varanauskienė, J.: Finansinių paslaugų vadovas. Vilnius, Lietuvos vartotojų institutas (2009) ISBN 978-9955-783-03-9
10. Lietuvos Respublikos Vertybinių popierių komisija.: Valdymo įmonių valdomų papildomo savanoriško pensijų kaupimo pensijų fondų veiklos 2011 m. II ketvirčio apžvalga (2011)

Information Technology Skills and Competencies – A Case for Professional Accountants

Ku Maisurah Ku Bahador and Abrar Haider

School of Computer and Information Science, University of South Australia, 5095, Mawson Lakes, South Australia, Australia
kuykb001@mymail.unisa.edu.au, abrar.haider@unisa.edu.au

Abstract. Growing utilisation and value profile of information technologies have also forced professional accountants to upgrade their skills set so as perform their job better. As a result, there is an increasing concern about the level of competencies that accountants possess in the use of information technologies, and whether they are prepared to meet the challenges of contemporary business environment. However, the skills acquired by professional accountants are still below the minimum level drawn by the accounting professional bodies such as the International Federation of Accountants and the American Institute of Certified Public Accountants. This study focuses on information technology based competencies framework for professional accountants, specifically those operating in small to medium sized accounting practices. This study makes a significant contribution to knowledge and practice by providing theoretical base for developing information technology related competencies for knowledge workers in general and professional accountants in particular.

Keywords: Information technology, competencies, accountants.

1 Introduction

As technologies evolve, corporate world has embraced technology at a fast rate. As a result, almost entire businesses have been automated. The result of this automation has been profound on job designs of the employees of these businesses. For example, terms like 'knowledge worker' have been consistently used to describe contemporary business workers, since information technologies (IT) have become an integral part of their routine work. The role of accountants in modern business organisations is no exception, which has been transformed from manual to computerised accounting [1]. In fact, IT has changed the way data is collected, processed, stored, and aggregated for preparation of accounting and finance related information required by the management to control and manage business activities [2].

IT competencies are imperative for accountants to perform their tasks [3] and constitute IT skills, IT experience, management skills (in particular project management) and conceptual skills. These skills on one hand aid the routine business activities related to accountants' work, and on the other hand help them create an

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 81–87, 2012.

environment where these technologies operate at their optimum level for the strategic internal and external advantage of the business. However, many parties, including accounting practitioners have expressed their concern about the level of IT-related skills and competencies acquired by accountants [4, 5]. Although the demand for accounting professional arose through the need to achieve skills in IT, the skills acquired by professional accountants are still below the minimum level drawn by the accounting professional bodies such as the International Federation of Accountants (IFAC) and the American Institute of Certified Public Accountants (AICPA) [4, 5, 6 and 7]. In addition, accountants are unclear about the exact IT-related skills and competencies that need to be possessed by them. This issue inspired the researcher to explore the set of skills required by professional accountants, specifically in Malaysian small to medium accounting practices to perform the tasks efficiently. Professional accountants in these type organisations have important roles in providing business support to their clients such as corporations, small and medium organisations, individuals and other organisations. It will, however, be interesting to study the competencies developed in the use of IT in different cultural settings.

This study is structured as follows: the next section describes the concept of IT competencies and new requirement of IT competencies for professional accountants, derived from the relevant literature. The second section is the proposed framework for IT competencies and the last section outlines a conclusion.

2 Related Work

Competence is defined as the state of quality of being adequately or well qualified or a specific range of skills, knowledge and abilities [8]. Competence, therefore, is the link between skills of employee and the job requirements [9]. In the accounting context, Carnaghan [10] views IT competencies as the qualities which are demonstrated by activities, such as the capacity to create a spread sheet or database for a particular purpose, or the ability to use software. According to IFAC [6], professional accountants are expected to possess necessary IT competencies. In fact, the credibility of the accounting profession in general depends on their success in fulfilling this obligation. Thus, every professional accountant is expected to act as a user, designer, manager, planner or evaluator of information systems; or in a combination of these roles [3]. It has to be acknowledged that these roles require technical skills, organisational skills, conceptual skills, and other social skills [6].

The IFAC through International Education Guideline (IEG 11) guidelines have been used by accounting educators and accounting practitioners as a guide to improve professionalism of accountants. This guideline refers to the technical skills required by accountants to be able to apply their skills in relevant accounting or business context [11]. However, the standard required of IT competencies is not specified and does not give a specific approach on how to develop the IT skills and competence. Therefore, the information systems literature has been reviewed to identify what dimensions or elements have been employed in the previous studies. Unfortunately, most of the studies have provided a list of IT skills such as the ability to use

spreadsheets, word processor, accounting packages and web browser that accountants must be proficient with [12 and 13].

These results are not consistent with the others studies stated that accountants need new skills rather than technical skills to improve their career and professionalism [10]. Although professional accountants' IT competencies are required by the professional standard set by IFAC, very little is known about their IT competencies levels, especially in developing economies such as Malaysia [1]. There are very few studies available that have investigated IT competencies for professional accountants in developing economies. These studies, however only use one dimension i.e. IT skills to measure accounting practitioners' competence in using IT.

2.1 New Requirement of IT Competencies for Accountants

IFAC through International Education Guideline 11 (IEG11) encourages professional accountants to have competencies in IT. However, the standard required of IT competencies is not specified and does not give a specific approach on how to develop the IT skills and competence. Therefore, the information systems literature has been reviewed to identify what dimensions or elements have been employed in previous studies. Most of the outcomes of these studies have been a list of IT skills such as the ability to use spreadsheets, word processor, accounting packages and web browser that accountants must be proficient with [13 and14]. Information systems literature indicates that IT related competencies dependent upon a number of other dimensions or skills [10]. Modern or contemporary professional accountants are required to provide leadership and management support in addition to their routine jobs. It is therefore, essential that professional accountants must have requisite organisational, management, behavioural, and people skills. These skills provide necessary support to IT skills so that professional accountants can perform their jobs effectively. It is, however also important to note that experience of accountants, the culture of organisations, and the formal training of accountants will always have significant influence on the level of competence in accountant possesses in operating, designing and using IT [9, 6, 15].

Table 1. Classifications of IS abilities/knowledge/skills

Skills	Categories/ Elements
Technical Skills	Analysis and design, programming language, specific application and general IS knowledge, information system product, database and data communication, advanced applications, computer applications systems, systems theory and concepts, business functional knowledge, technology management knowledge, operating systems, network, personal computer tools.
Organisational Skills	Time management, priority, information organisation.
People Skills	Organisational skills, organisational unit, interpersonal, communication, interpersonal relationships, management, professionalism, business, management, social, society, personal trait, professional skills, business knowledge.
Conceptual Skills	Problem solving, abstraction, strategic planning.

Source: Derived from Various Authors

Table 1 illustrates multi-dimensional IS knowledge, skills and abilities derived from various authors. Literature suggests four different set of skills that are required by a knowledge worker in the contemporary paradigm [16]. These skills are technical, organisational, people and conceptual skills. Technical skills involved specialised knowledge about methods, processes, and techniques designed to carry out specialised activity. Organisational skills are skills enable employees or workers to plan and carry on activities effectively. People skills deal with human behaviour and interpersonal process and conceptual skills include analytical ability, creativity, efficiency in problem solving and ability to recognise opportunities and potential problems.

Professional accountants' skills for success are highly required to react quickly and effectively in organisations. Thus, to be an effective accountant, the right mix of skills has to be developed to sustain the implementation of skill set of IT competencies. This skill set support accountants in everything they do during the accounting processes such as, auditing, recording daily financial transactions, preparing financial statements and making decisions. For that reason, AICPA through Core Competency framework asserted the values of professional accountants as competitive by identifying skill set of IT competencies elements such as communication and leadership skills, negotiation, strategy, problem solving and critical thinking and personal improvement as well as project management cited by Institute of Management Accountants [16]. In fact, the good skill set of IT competencies is critical to the prosperity and even the survival of organisations [17].

3 Proposed Framework

The research framework illustrated in Fig. 1 shows foundation for this study which demonstrated the multi-layered and multi-tiered framework of IT related competencies for professional accountants. The main research question is "*How IT competency profile changes with change in job description of a professional accountant working in Malaysian small to medium accounting practices?"* followed by three sub-questions: (1) what are the IT related competencies that help professional accountants to perform their job better?, (2) what are the IT skill sets required for professional accountants at each stage of their professional lifecycle?, and (3) what is an appropriate framework for developing IT related competencies for professional accountants in Malaysian small to medium accounting practices.

The inner layer of the framework represents the first sub-question of the study which deals with the generic skill set of IT competencies. It takes a comprehensive view of IT competence for professional accountants, and investigates the issue at hand in technical, organisational, people, and conceptual dimensions. The second layer of the framework seeks to find the answers for second sub-question i.e. to identify the IT related competencies that help professional accountants to perform their job properly. The aim is to employ the above mentioned skill set in the perspective of experience, organisational culture, formal accounting education and the international standard for accounting practices. These elements are major contributing factors that set the

standard of quality of professional accountants. The third layer of the framework recognises the actual competencies of professional accountants in using IT for jurisdiction specific. For the purpose of this study, it will be focused to Malaysian jurisdiction specifically within Malaysian small to medium accounting practices. It addresses improvement of the main accounting professional regulators in Malaysia who set the national accountants standard through national accounting board. In addition, higher education has been the main education provider in generating accounting graduates in Malaysia.

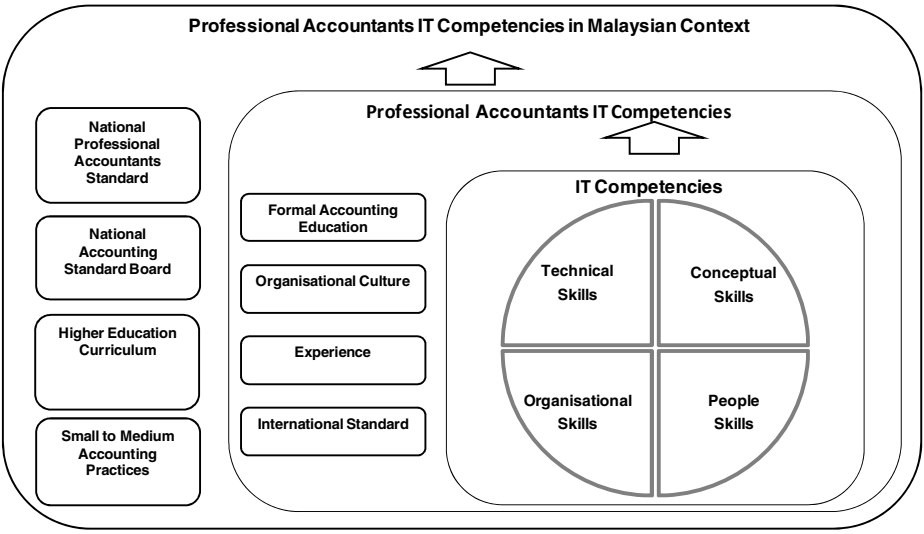


Fig. 1. Conceptual framework of the research

In order to answer the research question, this study will employ a qualitative interpretive research methodology with exploratory research study. Qualitative research methodology approach is represented by distinctive techniques and tools [18]. It involves non-numeric data to provide a deeper understanding of phenomena within its context and creates a strong relationship between the phenomena under study and the researcher [6]. This method provides the best gaining insight to comprehend the issues. Data will be collected through interviews from accountants (including junior, senior accountants and accountants who involve as managers of small to medium accounting practices), academics and representatives from accounting professional regulators in Malaysia. The range of their industry setting helped to understand the phenomena in a broader scope.

4 Conclusion

This study discusses the skills and competencies for professional accountants, particularly in small to medium sized accounting practices in Malaysia, which emphasise the

importance of understanding and being competent in the use of IT. This study provides a novel view, but proposing an all encompassing view of what IT skills and competencies are necessary for contemporary accounts and how do they complement each other to enable professional accountants throughout their professional lifecycle. Through this research, therefore, the authors aim to develop theory related to IT competencies relevant to the entire life cycle of the work process of professional accountants. The outcomes of this research will be invaluable for accounting profession, particularly in Asia. At the same time, these outcomes could be used by academic institutions, i.e. university and polytechnic or business schools for curriculum design so as to incorporate different levels of competences for courses aimed at different levels of professional accountants' professional lifecycle. It is expected that the findings of this study will formalise professional accountants' IT competencies framework and will provide support to on job training and career planning for professional accountants.

References

1. Ismail, N., Abidin, Z.: Perception towards the Importance and Knowledge of Information Technology among Auditors in Malaysia. Journal of Accounting and Taxation 1(4), 061–069 (2009)
2. Winograd, B.N., Gerson, J.S., Berlin, B.L.: Audit Practices of Pricewaterhouse Coopers, Auditing. Journal of Practitioner Theory 19(2), 175–182 (2000)
3. Wessel, P.J.: The Identification and Discussion of Strategies for Implementing an IT Skills Framework in the Education of Professional Accountants. South African Journal of Accounting Researh 22(1), 147–181 (2008)
4. Chang, C., Hwang, N.: Accounting Education, Firm Training and Information Technology: A Research Note. Accounting Education 12(4), 441–450 (2003)
5. Jones, G., Abraham, A.: Educational Implications of the Changing Role of Accountants Perceptions of Practitioners, Academic and Students. In: The Quantitative Analysis of Teaching and Learning in Business. Economics and Commerce, Forum Proceeding, pp. 89–105. The University of Melbourne (2007), http://ro.uow.edu.au/commpapers/296/
6. International Federation of Accountants Education Committee (IFAC): Information Technology for Professional Accountants (2003), https://www.imanet.org/pdf/ITPA.pdf
7. Senik, R., Broad, M.: Information Technology Skills Development for Accounting Graduates: Intervening Conditioned, International Education Studies. Canadian Centre of Science and Education 4(2), 105–110 (2011)
8. Oxford Dictionary, competence definition, http://www.answers.com/topic/competence
9. Pollard, C., Steczkowicz, A.: Assessing Business Managers' IT Competence in SMEs in Regional Australia: Preliminary Evidence from A New IT Competence Instrument. In: Proceedings 7th PACIS, Adelaide, pp. 1005–1020 (July 2003)
10. Carnaghan, C.: Discussion of IT Assurance Competencies. International Journal of Accounting Information Systems 5, 267–273 (2004)
11. Wessels, P.J.: Critical Information and Communication Technology Skills for Professional Accountants. Mediatory Accountancy Research 13(1), 87–103 (2005)

12. Lai, M.L., Nawawi, N.H.A.: Integrating ICT Skills and Tax Software in Tax Education: A Survey of Malaysian Tax Practitioners' Perspectives. Campus-Wide Information Systems 27(5), 303 (2010)
13. Greenstein-Prosch, M., McKee, T.E.: Assurance Practitioners' and Educators' Self Perceived IT Knowledge Level: An Empirical Assessment. International Journal Accounting Information Systems 5, 213–243 (2004)
14. Duedhl, M., Andersen, J., and Sein, M. K.: When Models Cross the Border: Adapting IT Competencies of Business Managers. Paper Presented at the SIGMIS-CPR (2005)
15. Havelka, D., Merhout, J.W.: Towards a Theory of Information Technology Professional Competence. The Journal of Computer Systems 50(2), 106–116 (2009)
16. Aldag, R.J., Kuzuhara, L.W.: Organizational Behavior and Management: An Integrated Skills Approach, Thomson Learning (2002)
17. Dillon, T.W., Kruck, S.E.: The Emergence of Accounting Information Systems Programs' Management Accounting Quarterly (Spring 2004)
18. Sekaran, U.: Research Methods for Business: A Skill Building Approach. Wiley & Sons Australia (2003)

Enterprise Governance of IT and the Evolutions in COBIT: An Academic Perspective

Steven De Haes and Wim Van Grembergen

University of Antwerp – Antwerp Management School, Prinsstraat 13,
2000 Antwerp, Belgium
{steven.dehaes,wim.vangrembergen}@ua.ac.be

Abstract. Recently a new version of COBIT, the good-practice framework for enterprise governance of IT, was released. Although is it not always clearly articulated in its documentation, this new COBIT 5 framework incorporates many concepts and theories out of the IT and general management literature. This paper wants to position COBIT as a framework for enterprise governance of IT, and illustrates how the core principles and building blocks of the framework are derived from insights from literature. Based on this discussion, research questions for future research around enterprise governance of IT and COBIT 5 are proposed and discussed. As such, this paper wants to contribute to the exploration on how a practice-oriented framework such as COBIT could be used in future research activities.

Keywords: Enterprise Governance of IT, IT governance, IT management, COBIT, business/IT alignment, balanced scorecard, IT savvy, organizational system.

1 Introduction

Information technology (IT) has become crucial in the support, sustainability and growth of most enterprises. To overcome the IT productivity paradox as described by Strassman [21] and Brynjolfsson [2], this pervasive use of IT calls for a specific focus on enterprise governance of IT (EGIT) [4, 22]. Enterprise governance of IT is an integral part of enterprise governance and addresses the definition and implementation of processes, structures and relational mechanisms in the organization that enable both business and IT people to execute their responsibilities in support of business/IT alignment and the creation of business value [25].

In the field, many best-practice frameworks are developed and promoted to guide managers in implementing enterprise governance of IT. [25] One of these frameworks is COBIT, of which a new –fifth- version has been released in April 2012 [13]. COBIT (Control Objectives for Information and Related Technologies) is a freely available industry framework that describes a set of best practices for management, control and assurance of information technology, and organizes them around a logical framework of IT related processes.

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 88–103, 2012.

While organizations are adopting COBIT in practice [3, 25], still little academic research is available that leverages COBIT as an instrument in executing research programs. However, as many of the core principles of COBIT heavily build on models, concepts and theories out of the IT and general management literature, there are certainly opportunities for COBIT-based research to be explored.

In this article, we discuss how the COBIT 5 framework embraces new and innovative concepts from literature and we provide directions and challenges for doing future research leveraging COBIT 5. In this way, this paper wants to explore possibilities to better introduce COBIT in academic literature.

This paper will first define the concept of Enterprise Governance of IT in more detail and will then position COBIT as a framework for enterprise governance of IT. Next, it is explained how COBIT 5 embraces insights from IT and general management literature. This will be used as a basis do discuss directions for future research around enterprise governance of IT and COBIT. Finally, a closing section will bring some concluding remarks together.

2 Paper Development Approach

The authors of this paper have been actively engaged in the COBIT development the past decade. Since 2001 and up till 2012, they held membership in the COBIT Steering Committees and COBIT Development Groups, groups of international senior practitioners worldwide responsible for continuously driving the development of COBIT forward. In these groups, the author's mandate was to define and execute new research projects in support of the future evolutions of COBIT. These research projects were built on models and concepts from literature, such as the "Balanced Scorecard" and the concept of "Strategic Alignment", which each time were refined and framed to fit into the COBIT models and presentations [25].

Based on these experiences in the past decade, the authors want to put forward some of the concepts and models used in the development of COBIT. These insights can help in better explaining the COBIT framework, and in finding opportunities for future research based on COBIT.

3 Enterprise Governance of IT

IT governance is a concept that has increasingly become an important issue in the IT area. It is not exactly clear when the concept as we understand it now originated. In 1998, the IT Governance Institute (www.itgi.org) was founded to disperse the IT governance concept in the practitioner area. In academic literature, articles mentioning IT governance in the title or abstract also emerged in the 1990's (eg. [1, 20], mostly focusing on the centralization – decentralization debate. In the context of one of the mini-tracks at the academic Hawaii International Conference on Systems Sciences (HICSS), IT governance was defined more from an holistic organizational perspective as "the organizational capacity exercised by the board, executive

management and IT management to control the formulation and implementation of IT strategy and in this way ensure the fusion of business and IT" [23].

After the emergence of the IT governance concepts, the notion received a lot of attention. However, due the focus on "IT" in the naming of the concept, the IT governance discussion mainly stayed a discussion within the IT area. In the field, many IT governance implementations are driven by IT, while one would expect that the business would and should take a leading role [4] . Many authors agree that that the involvement of business is crucial [22, 28, 16] which initiated a shift in the naming of the concept towards "enterprise governance of IT", focusing on the business involvement. Enterprise governance of IT is an integral part of corporate governance exercised by the Board overseeing the definition and implementation of processes, structures and relational mechanisms in the organization that enable both business and IT to execute their responsibilities in support of business/IT alignment and the creation of business value from IT enabled investments [25]. This adapted definition indicates more clearly that IT governance is the responsibility of the Board and that the execution lies with (executive) management.

Enterprise governance of IT clearly goes beyond the IT responsibilities and expands towards (IT related) business processes needed for business value creation. Also the standardization organization ISO moved into this direction, with the release in 2008 a new worldwide standard defined as "Corporate Governance of IT" [9]. In this standard, ISO puts forward six principles for governance of IT, addressing primarily business' roles and responsibilities, expressing preferred behaviour to guide IT related decision-making.

4 COBIT History

COBIT is developed by ISACA (Information Systems Audit and Control Association), an international professional membership association for IT professionals and IT auditors counting about 100.000 members worldwide. COBIT initially originated mid nineties out of the (financial) audit community. Those audit professionals were confronted more and more with automated environments. To guide their work in these IT-related environments, COBIT was initially developed as a framework for executing IT audit assignments, built around a comprehensive set of Control Objectives for IT processes. Building on this IT auditing basis, the COBIT framework was developed further becoming a broader IT management framework, with in 2000 the addition of "Management Guidelines" in COBIT version 3, including metrics, critical success factors and maturity models for IT processes. In 2005 again a new release was issued, COBIT 4, containing several new management and governance concepts, such as (i) the alignment of business and IT goals and their relationship with supporting IT processes, (ii) roles and responsibilities within IT processes and (iii) the inter-relationship between IT processes. With these extensions COBIT wanted to continue to establish itself as a generally accepted framework for IT governance. [25]

In the shift from IT governance to enterprise governance of IT, as discussed in previous section, ISACA complemented its IT governance best practices framework COBIT, focusing on IT processes and responsibilities, with the Val IT and RISKIT framework [11, 12], addressing the IT related business processes and responsibilities in value creation (VALIT) and risk management (RISKIT). In the field, COBIT, RISKIT and VALIT are considered to be strong reference frameworks guiding managers to implement enterprise governance of IT in their organization, as visualized in Figure 1. [25]

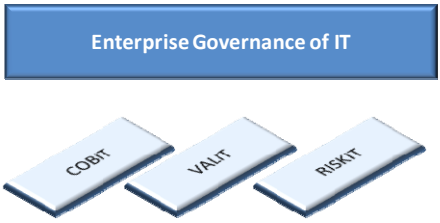


Fig. 1. COBIT, VALIT and RISKIT as frameworks for Enterprise Governance of IT

In April 2012, the latest version COBIT 5 was released, referencing the concept of enterprise governance of IT on its cover. According to the ISACA website, "COBIT 5 provides a comprehensive framework that assists enterprises to achieve their objectives for the governance and management of enterprise IT. ... COBIT 5 enables IT to be governed and managed in a holistic manner for the whole enterprise, taking in the full end-to-end business and IT functional areas of responsibility, considering the IT-related interests of internal and external stakeholders." [14] COBIT 5 will integrate all knowledge previously dispersed over the three important ISACA frameworks COBIT, VALIT and RISKIT, as such becoming a "one-stop-shop" to enter ISACA's body of knowledge.

5 COBIT 5 Principles Linked to Literature

The COBIT 5 manuals state that the framework is built around five core principles, as visualised in Figure 2. Each of those principles are discussed in this section and related to concepts and insights from general and IT literature.

5.1 Meeting Stakeholder Needs: Strategic Business/IT Alignment

According to ISACA, Principle 1 (Meeting Stakeholder Needs) implies that COBIT 5 provides all the required processes and other enablers to support business value creation through the use of IT, as such meeting all stakeholder needs. This principle closely links to the "Strategic Alignment" discussion as initiated by Henderson and Venkatraman in 1993[8].

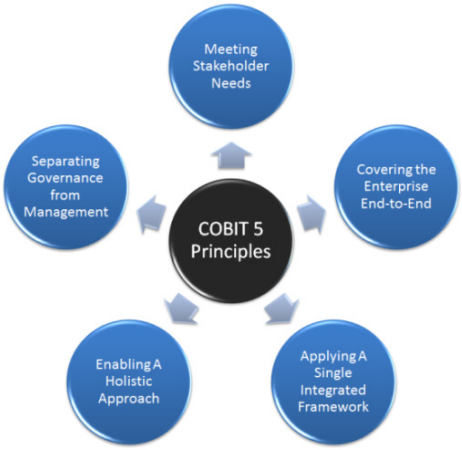


Fig. 2. COBIT 5 Principles

The idea behind strategic alignment is very comprehensive and has been present in all COBIT documentation since the first release. However, the challenge is how organizations can achieve this ultimate objective. To answer this question, a series of research steps has been set up by the development team in order to provide guidance in understanding how enterprise goals drive IT related goals and vice versa. Throughout these research projects, and based on many in-depth interviews in different sectors and expert team/delphi interrogations, a generic list of enterprise goals, IT related goals and its inter-relationship was established, as shown in Figure 3. This cascade now constitutes the core entry point of COBIT 5, implying that organizations should always start with analysing their alignment situation through defining and linking enterprise goals and IT related goals [6, 26].

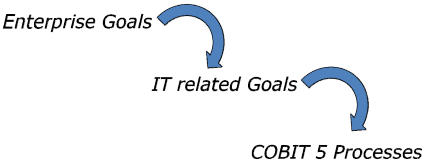


Fig. 3. Enterprise Goals, IT related Goals and COBIT 5 processes

Important to mention is that COBIT 5 uses the word “Enterprise Goals” instead of “Business Goals” as referenced in COBIT 4. With this shift, COBIT 5 wants to explicitly include both profit and non-profit (government) types of enterprises. Also, COBIT 5 talks about “IT related” goals and not about “IT goals” anymore as in

COBIT 4. The reason for that is also explained in the next "COBIT principle", addressing the conviction that both business and IT people have "IT related" responsibilities in realizing value out of IT.

As an illustration of this cascade, Figure 4 shows that the enterprise goal of "External compliance with laws and regulation" requires primary focus (P) on the IT related goals of "security of information and processing infrastructure". In the documentation of COBIT 5, the importance of this IT related goal will in turn lead to (see Figure 5) primary focus on some of the 37 COBIT 5 processes, for example "Manage Risks", "Manage Security and "Manage Changes".

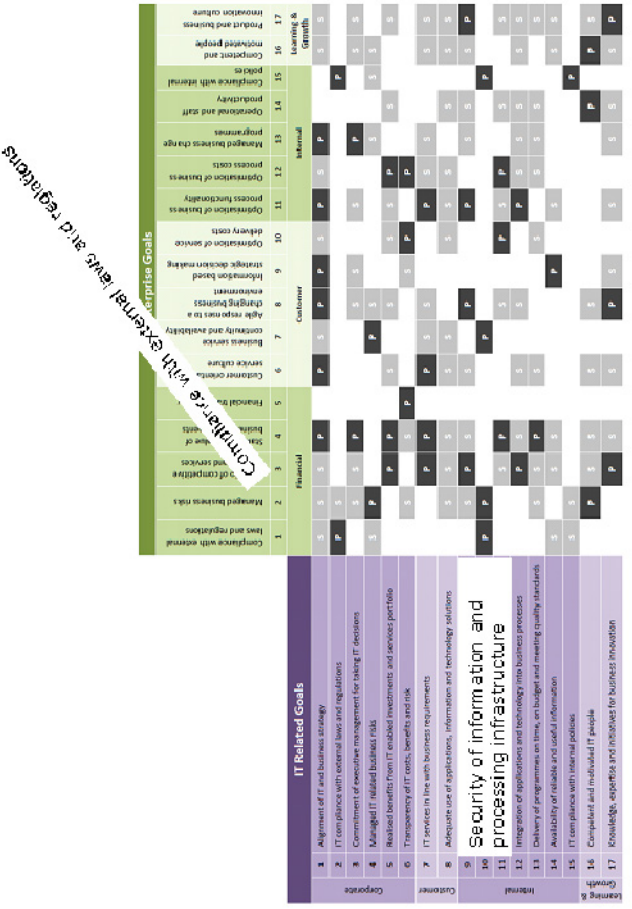


Fig. 4. Cascade of Enterprise Goals and IT related Goals

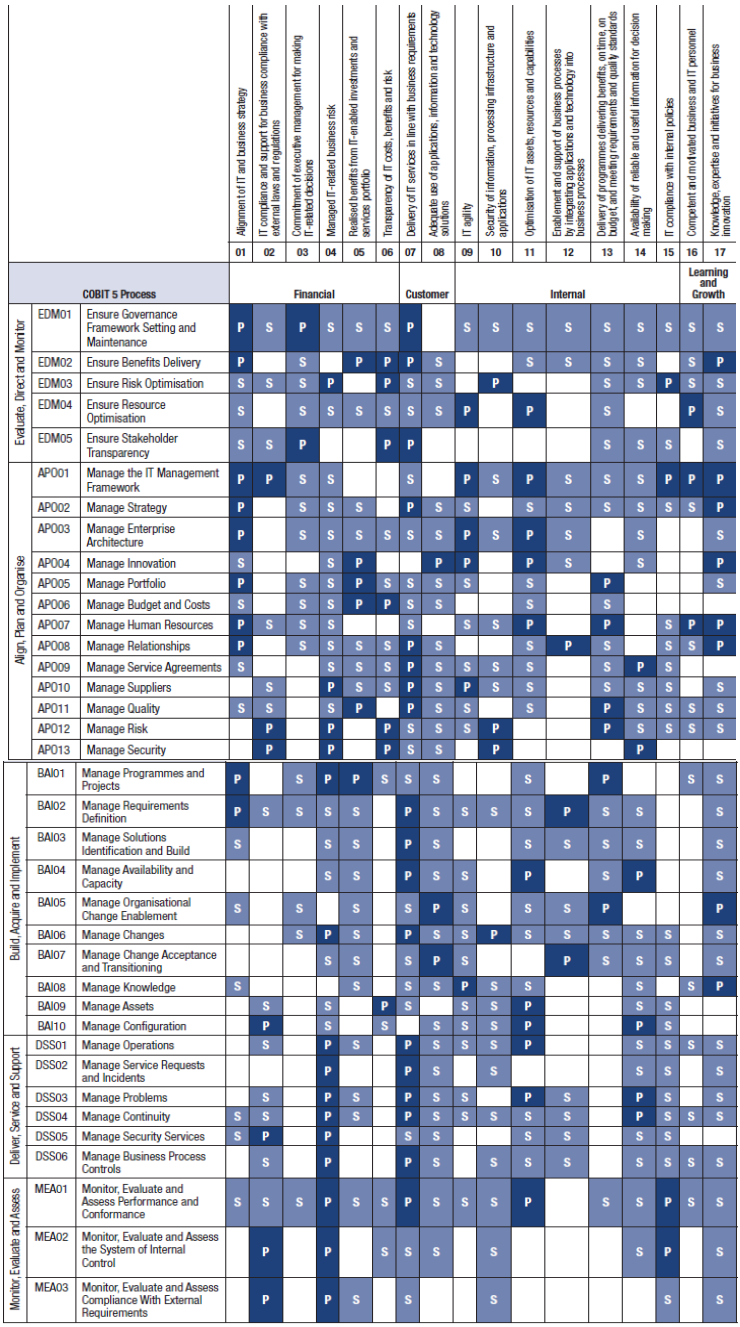


Fig. 5. Cascade of IT related Goals and COBIT 5 Processes

5.2 Meeting Stakeholder Needs: The Balanced Scorecard

To verify whether stakeholder needs are indeed being met, a sound measurement process needs to be established. Traditional performance methods such as return on investment (ROI) capture the financial worth of IT projects and systems, but reflect only a limited (tangible) part of the value that can be delivered by IT [25].

The enable a broader measurement process, the developers of COBIT have built on the concepts of the (IT) balanced scorecard as developed by Kaplan and Norton [15] and Van Grembergen et al [24]. As shown in Figure 4, all enterprise goals and IT related goals are grouped in the balanced scorecard perspectives. COBIT also provides samples of outcome metrics to measure each of those goals and to really build a scorecard for IT related activities. Figure 6 provides some examples of such metrics for the "customer perspective" of the enterprise goals and the IT related goals.

BSC Dimension	Enterprise Goal	Metric
Customer	6. Customer-oriented service culture	• Number of customer service disruptions due to IT service-related incidents (reliability) • Percent of business stakeholders satisfied that customer service delivery meets agreed-on levels • Number of customer complaints • Trend of customer satisfaction survey results
	7. Business service continuity and availability	• Number of customer service interruptions causing significant incidents • Business cost of incidents • Number of business processing hours lost due to unplanned service interruptions • Percent of complaints as a function of committed service availability targets
	8. Agile responses to a changing business environment	• Level of board satisfaction with enterprise responsiveness to new requirements • Number of critical products and services supported by up-to-date business processes • Average time to turn strategic enterprise objectives into an agreed-on and approved initiative
	9. Information-based strategic decision making	• Degree of board and executive management satisfaction with decision making • Number of incidents caused by incorrect business decisions based on inaccurate information • Time to provide supporting information to enable effective business decisions
	10. Optimisation of service delivery costs	• Frequency of service delivery cost optimisation assessments • Trend of cost assessment vs. service level results • Satisfaction levels of board and executive management with service delivery costs

BSC Dimension	IT-related Goal	Metric
Customer	07 Delivery of IT services in line with business requirements	• Number of business disruptions due to IT service incidents • Percent of business stakeholders satisfied that IT service delivery meets agreed-on service levels • Percent of users satisfied with the quality of IT service delivery
	08 Adequate use of applications, information and technology solutions	• Percent of business process owners satisfied with supporting IT products and services • Level of business user understanding of how technology solutions support their processes • Satisfaction level of business users with training and user manuals • Net present value (NPV) showing business satisfaction level of the quality and usefulness of the technology solutions

Fig. 6. Balanced scorecard metrics for enterprise goals and IT related goals

Moreover, COBIT 5 provides outcome measure at the level of the 37 detailed COBIT 5 processes. An example in shown in Figure 7 for the process of Managing Security, providing specific process goals and related metrics. Consolidating all these metrics, at enterprise level, IT related level and COBIT processes level, enables organization to build a comprehensive scorecard for the entire IT related environment, as in instrument to verify whether stakeholder needs are being met.

Process Goals and Metrics	
Process Goal	**Related Metrics**
1. A system is in place that considers and effectively addresses enterprise information security requirements.	• Number of key security roles clearly defined • Number of security related incidents
2. A security plan has been established, accepted and communicated throughout the enterprise.	• Level of stakeholder satisfaction with the security plan throughout the enterprise • Number of security solutions deviating from the plan • Number of security solutions deviating from the enterprise architecture
3. Information security solutions are implemented and operated consistently throughout the enterprise.	• Number of services with confirmed alignment to the security plan • Number of security incidents caused by non-adherence to the security plan • Number of solutions developed with confirmed alignment to the security plan

Fig. 7. Balanced scorecard metrics for the Security Process

5.3 Covering the Enterprise End-to-End: IT Savviness

The second principle (Covering the Enterprise End to End) articulates that COBIT 5 covers all functions and processes within the enterprise. COBIT 5 does not focus only on the "IT function", but treats information and related technologies as assets that need to be dealt with just like any other asset by everyone in the enterprise. This statement related to the work of Weill and Ross [28] on IT Savviness, concluding that business people should take up responsibility in managing their IT related assets. Their work clarifies the need for the business to take ownership of, and be accountable for, governing the use of IT in creating value from IT-enabled business investments.

This implies a crucial shift in the minds of the business and IT, moving away from managing IT as a 'cost' toward managing IT as an 'asset' to create business value. As Weill and Ross describe in their 2009 'IT Savvy' book [28]: "If senior managers do not accept accountability for IT, the company will inevitable throw its IT money to multiple tactical initiatives with no clear impact on the organisational capabilities. IT becomes a liability instead of a strategic asset".

Related to this discussion, COBIT 5 talks about "Covering the Enterprise End to End". COBIT 5 does cover both IT processes and IT related business processes. As a demonstration of this, COBIT 5 provides RACI charts (Responsible, Accountable, Consulted, Informed) in which both business roles and IT roles are included. To illustrate this, an example RACI chart for the process "Manage Service Agreements" is shown in Figure 8. This RACI chart indicates that for the SLA process both business and IT functions have accountabilities and responsibilities.

5.4 Applying a Single, Integrated Framework: COBIT/RISKIT/VALIT

Principle 3 (Applying a Single, Integrated Framework) explains that COBIT 5 aligns with other relevant standards and frameworks at a high level, and thus can serve as the overarching framework for governance and management of enterprise IT. COBIT 5 becomes on overall integration, or one-stop-shop, of all previous ISACA related materials as published COBIT 4, VALIT and RISKIT [10, 11, 12].

APO09 RACI Chart																											
Key Management Practice	Board	Chief Executive Officer	Chief Financial Officer	Chief Operating Officer	Business Executives	Business Process Owners	Strategy Executive Committee	Steering (Programmes/Projects) Committee	Project Management Office	Value Management Office	Chief Risk Officer	Chief Information Security Officer	Architecture Board	Enterprise Risk Committee	Head Human Resources	Compliance	Audit	Chief Information Officer	Head Architect	Head Development	Head IT Operations	Head IT Administration	Service Manager	Information Security Manager	Business Continuity Manager	Privacy Officer
APO09.01 Identify IT services.		C		R	R	R	C		I							I	I	R	I	C	C	C	A	I	I	
APO09.02 Catalogue IT-enabled services.					I	I			I							I	I	R	I	C	C	C	A	I	I	
APO09.03 Define and prepare service agreements.					R	C			C		C					C	C	R		C	R	R	A	C	C	
APO09.04 Monitor and report service levels.		I		I	I	R					C							I		I	I	I	A			
APO09.05 Review service agreements and contracts.					A	C			C		C					C	C	R		C	R	R	R	C	C	I

Fig. 8. End-to-end responsibility in Managing Service Agreements

In this overarching approach, COBIT identifies 37 processes spread over a governance and a management domain, as visualized in Figure 9. The five governance processes are the board's responsibilities in IT, covering the setting of the governance framework, responsibilities in terms of value (e.g. investment criteria), risks (e.g. risk appetite) and resources (e.g. resource optimisation) and providing transparency regarding IT to the stakeholders. In the management area, four subdomains are defined: Align, Plan, Organise (APO), Build, Acquire and Implement (BAI), Deliver, Service and Support (DSS) and Monitor, Evaluate and Assess (MEA). The domain APO concerns the identification of how IT can best contribute to the achievement of the business objectives. A management framework is required, and specific processes related to the IT strategy and tactics, enterprise architecture, innovation and portfolio management. Other important processes in this domain address the management of budgets and costs, human resources, relationships, service agreements, suppliers, quality, risk and security. The domain BAI concretises the IT strategy through identifying in detail the requirements for IT and managing program and projects. This domain further talks about managing capacity, organizational change, IT changes, acceptance and transitioning, knowledge, assets and configurations. The domain Delivery and Support refers to the actual delivery of required services. It contains processes around managing operations, service requests and incidents, problems, continuity, security services and business process controls. The fourth management domain, MEA, includes those processes that are responsible for the quality assessment in compliance with the control requirements for all previous mentioned processes. It addresses performance management, monitoring of internal control and regulatory compliance. [13]

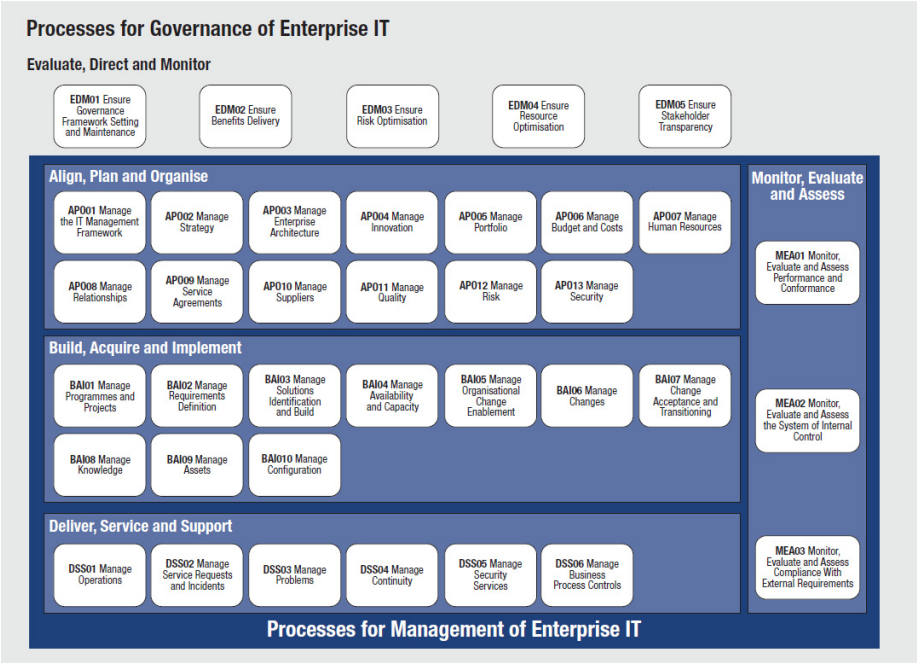


Fig. 9. COBIT 5 Processes

5.5 Applying a Single Integrated Framework: IT Savviness

Comparing this set of processes with previous versions of COBIT clearly demonstrates the extension towards business roles and responsibilities in governing and management IT. Again, this extension fully aligns with the requirement of business people taking up accountable in managing IT, as put forward in the IT savviness discussion.

For example, newly inserted processes that address specific business roles are APO3: Manage Enterprise Architecture, APO4: Manage Innovation and BAI05: Manage Organizational Change. In this context, specific attention goes to "DS06: Manage Business Process Controls". This process was removed out of the previous version of COBIT, as the previous version focused more on the responsibilities of the IT department. Managing business process controls (application controls) was considered to be out of scope, being a prime accountability of the business. However, COBIT 5 aims to include both business' and IT' responsibilities in managing and governing IT. For that reason, the process on business process controls was re-included in COBIT 5.

As a side note, it should also be mentioned that there are much less processes in the "Deliver, Service and Support" domain (6) as compared to the number of processes in the "Deliver and Support" domain of previous COBIT version (13). Many of these processes are moved to a higher domain. A typical example is the shift of Manage

Service Agreements to APO argumented by the recent evolution of the externalisation of IT operations through outsourcing and cloud computing.

5.6 Enabling a Holistic Approach: Organizational Systems

The fourth principle (Enabling a Holistic Approach) explains that efficient and effective implementation of governance and management of enterprise IT requires a holistic approach, taking into account several interacting components, such as Processes, Structures and People.

This implementation challenge is related to what is described in strategic management literature as the need for an organizational system, i.e. "the way a firm gets its people to work together to carry out the business". [7] Such organizational system requires the definition and application, in a holistic whole, of structures (e.g. organisational units and functions) and processes (to ensure tasks are coordinated and integrated), and attention to people and relational aspects (e.g. culture, values, joint beliefs, etc).

Peterson [17] and De Haes van Van Grembergen [4, 5], have applied this organizational system theory to the discussion of enterprise governance of IT. These authors conclude that organizations can and are deploying enterprise governance of IT by using a holistic mixture of various structures, processes and relational mechanisms. Enterprise governance of IT structures include organizational units and roles responsible for making IT decisions and for enabling contacts between business and IT management decision-making functions (e.g. IT steering committee). This can be seen as a form of blueprint for how the governance framework will be structurally organized. Enterprise governance of IT processes refers to the formalization and institutionalization of strategic IT decision making and IT monitoring procedures, to ensure that daily behaviours are consistent with policies and provide input back to decisions (e.g. IT balanced scorecard). The relational mechanisms are ultimately about the active participation of, and collaborative relationship among, corporate executives, IT management, and business management, and include mechanisms such as announcements, advocates, and education efforts.

COBIT 5 builds on these insights and talks about "Enablers" in its framework. Enablers are defined as factors that, individually and collectively, influence whether something will work in this case, governance and management over enterprise IT. The COBIT 5 framework describes seven categories of enablers (see Figure 10), of which the "processes", "organisational structures" and "culture, behaviour and ethics" are closely related to the organizational systems concept.

5.7 Separating Governance from Management: ISO 38:500

Principle 5 finally, (Separating Governance from Management) is about the distinction COBIT 5 makes between governance and management, which, as discussed before, heavily build on the position put forward by ISO38:500 [9]. In COBIT 5, ISACA states for the first time that these IT governance and IT

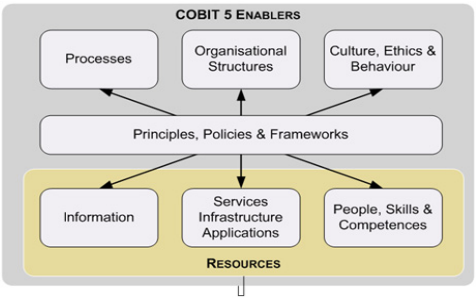


Fig. 10. Organizational systems of Enablers

management processes encompass different types of activities. The governance processes are organized following the EDM model ("Evaluate – Direct – Monitor"), as proposed by the ISO 38:500 standard on Corporate Governance of IT. IT governance processes ensure that enterprise objectives are achieved by evaluating stakeholder needs, setting direction through prioritisation and decision making; and monitoring performance, compliance and progress against plans. In enterprises, IT governance should be the responsibility of the board of directors under the leadership of the chairperson. Based on these governance activities, business and IT management plans, builds, runs and monitors activities (a COBIT translation of Deming's PDCA circle Plan, Do, Check, Act) in alignment with the direction set by the governance body to achieve the enterprise objectives. [13]. This all is in line with the (adapted) definition as formulated in a preceding paragraph: IT governance is the Board's accountability/responsibility and the execution is executive's accountability/ responsibility [25].

6 COBIT 5 Principles Linked to Potential Research Questions

As COBIT 5 is a practitioners framework, it is often challenged as an instrument that can be used for academic research. Very little research is available that builds on the international frameworks COBIT. However, many of the concepts of COBIT 5 are derived from IT and general management literature, opening opportunities to explore potential future research programmes where COBIT can play a role. In this section, some potential research questions are discussed.

6.1 Meeting Stakeholder Needs → Research Question: How Do Organizations Define and Align Enterprise Goals and IT Related Goals?

The concept of business/IT alignment is not new, but still high on the agenda of many organizations. Building on the strategic alignment model of Henderson and Venkatram [8], COBIT provides an approach on how to define enterprise goals and IT related goals. Through case study research, it could be analysed whether

organizations are clearly articulating enterprise goals and IT related goals, and to which degree this is done in a symbiotic way or disparate. Specific questions can be:

- Are businesses clearly articulating their priorities to IT?
- Is IT pro-actively engaged in the business strategic discussion?
- Is the business involved in defining the IT related goals?

6.2 Meeting Stakeholder Needs → Research Question: How Do Organizations Measure the Performance of IT?

Measuring the value of IT is a complex challenge. As COBIT leveraged the balanced scorecard insights, it provides a reference to build conceptual measurement frameworks for IT as a whole or for specific parts of IT. Research projects could work on building such conceptual frameworks based on COBIT, and then validate whether such measurements instruments are used or can be optimized based on empirical findings. Specific questions can be:

- Are organizations using COBIT to build balanced scorecards?
- Are the metrics suggested in COBIT 5 usable for practice?
- How are enterprises organizing the performance management process?

6.3 Covering the Enterprise End-to-End → Research Question: How Involved Is the Business in Enterprise Governance of IT?

COBIT 5 talks about end-to-end responsibilities in governing and managing the IT assets. The RACI charts in COBIT 5 provide useable templates to, in case research, analyse whether business people are really taking up their IT related responsibilities. Specific questions can be:

Are business people aware of the responsibilities as assigned in the COBIT 5 RACI charts?

- Do business people take up the responsibilities as assigned in the COBIT 5 RACI charts?
- What are enablers and inhibitors for business people to take up the responsibilities as assigned in the COBIT 5 RACI charts?

6.4 Enabling an Holistic Approach → Research Question: How Are Organization Implementing Enterprise Governance of IT?

As stated early, many contemporary organisations recognise the importance of enterprise governance of IT, but they are still struggling with getting such governance practices implemented and embedded into their organisations. Through case and/or survey research, it would be interesting to verify whether and how organizations are adopting enterprise governance of IT. Building on the organizational systems theory, COBIT 5 can be used to build an interview protocol, ensuring that the appropriate questions are asked during data gathering. Specific questions can be:

- Which COBIT 5 processes and related practices/structures are most adopted in organizations?
- Which COBIT 5 processes and related practices/structures are perceived being most effective?
- Which COBIT 5 processes and related practices/structures are perceived being easy/difficult to implement?

6.5 Separating Governance from Management → Research Question: How Is the Board of Directors Involved in Enterprise Governance of IT?

COBIT 5 clearly distinguishes between governance and management. Limited research is available on how boards are taking up responsibility in IT. Looking at annual reports or through case/survey research, it would be interesting to verify whether board are taking up the 5 areas of responsibility as discussed in COBIT.

- Which of the five governance processes are really taken up by Boards?
- What are boards reporting on their IT governance roles in the annual report?
- What is the relationship between boards' involvement and IT governance performance?

7 Summary and Conclusion

In 2012 a new version of COBIT, the international best practice framework for enterprise governance of IT, was released. COBIT 5 primarily is a framework made by and for practitioners, but in the past decade, it has also incorporated many insights coming from IT and general management literature. However, still little academic research is available that leverages COBIT as in instrument in executing research programs. Through clearly indicating how the core elements of COBIT 5 are built on IT and general management insights, this paper want to contribute to the exploration of the use of COBIT in future research activities. A first list of potential research questions is provided, as source of inspiration for researchers in this field, mainly leading to qualitative research approaches. A lot of opportunities are out there.

References

1. Brown, C.V., Magill, S.L.: Alignment of the IS functions with the enterprise: towards of model of antecedents. MIS Quarterly 18(4), 371–403 (1994)
2. Brynjolfsson, E.: The Productivity Paradox of Information Technology. Communications of the ACM 36(12), 66–77 (1993)
3. Debreceny, R.: IT Governance and Process Maturity: A Field Study. In: Proceedings of the 42th Hawaii International Conference on System Sciences, HICSS (2009)
4. De Haes, S., Van Grembergen, W.: An Exploratory Study into the Design of an IT Governance Minimum Baseline through Delphi Research. Communications of AIS (22), 443–458 (2008)

5. De Haes, S., Van Grembergen, W.: An exploratory study into IT governance implementations and its impact on business/IT alignment. Information Systems Management 26(2), 123–137 (2009)
6. De Haes, S., Van Grembergen, W.: Prioritising and linking business goals and IT goals in the financial sector. International Journal of IT/Business Alignment and Governance 1(2), 47–67 (2010)
7. De Wit, B., Meyer, R.: Strategy Synthesis: Revolving Strategy Paradoxes to Create Competitive Advantage. Cengage Learning EMEA (2005)
8. Henderson, J.C., Venkatraman, N.: Strategic alignment: leveraging Information Technology for transforming organizations. IBM Systems Journal 32(1) (1993)
9. ISO/IEC: 38500:2008: Corporate governance of information technology (2008), http://www.iso.org
10. ISACA: COBIT 4.1 (2007), http://www.isaca.org
11. ISACA: Val IT 2.0 (2008), http://www.isaca.org
12. ISACA: RISKIT (2009), http://www.isaca.org
13. ISACA: COBIT 5 (2012), http://www.isaca.org
14. ISACA, http://www.isaca.org/cobit (visited on April 2012)
15. Kaplan, R., Norton, D.: The balanced scorecard – measures that drive performance. Harvard Business Review 70(1), 71–79 (1992)
16. Peppard, J.: Unlocking the Performance of the Chief Information Officer. California Management Review 52(4), 73–99 (2010)
17. Peterson, R.: Crafting Information Technology Governance. Information Systems Management 21(4), 1–16 (2004)
18. Pferrer, J., Sutton, R.: The Knowing-Doing Gap. HBR Press (2000)
19. Ross, J.W., Weill, P.: Six IT Decisions your IT People shouldn't make. Harvard Business Review 80(11), 84–92 (2002)
20. Sambamurthy, V., Zmud, R.W.: Arrangements for Information Technology Governance: a theory of multiple contingencies. MIS Quarterly 23(2), 261–290 (1999)
21. Strassman, P.: The business value of Computers. The Information Economics Press, New Canaan (1990)
22. Thorp, J.: The Information Paradox. McGraw-Hill Reyerson (2003)
23. Van Grembergen, W.: Introduction to the Minitrack: IT Governance and its Mechanisms. In: Proceedings of the 35th Hawaii International Conference on System Sciences, HICSS (2002)
24. Van Grembergen, W., Saull, R., De Haes, S.: Linking the IT Balanced Scorecard to the Business Objectives at a major Canadian Financial Group. Journal for Information Technology Cases and Applications 5(1) (2003)
25. Van Grembergen, W., De Haes, S.: Enterprise Governance of IT: Achieving Strategic Alignment and Value. Springer, New York (2009)
26. Van Grembergen, W., De Haes, S., Van Brempt, H.: Understanding how business goals drive IT goals (2008), http://www.isaca.org
27. Van Grembergen, W., De Haes, S., Van Brempt, H.: How does the business drive IT? Identifying, prioritising and linking business and IT goals. Information Systems Control Journal 6, 1–5 (2007)
28. Weill, P., Ross, J.: IT Savvy: What Top Executives Must Know to Go from Pain to Gain. Harvard Business Press, Boston (2009)

Paradigm Change: Aligning Business and IT with a Business Software Integration Method

Susanne Altendorfer

Industrial Logistics, Montanuniversität Leoben
Franz-Josef-Strasse 18, 8700 Leoben, Austria
susanne.altendorfer@unileoben.ac.at

Abstract. Times change and so do business requirements: Within the years the requirements to do business have changed and got even more complex. Thus the companies' structures have to respond to these new situations with changing the way they are doing business. Traditional structures are no longer appropriate. Therefore a business reengineering is normally not enough. A software reengineering is also necessary to support the business processes in the best way. Business Software is here the key. This paper presents the paradigm change from monolithic standalone software to business software that ideally supports the business processes and thus helps streamlining the processes. It fosters the implementation of the term business software, especially in the technical area and describes the underlying software architecture. Furthermore the advantages of business software are outlined within the automotive industry.

Keywords: business processes, process streamlining, business software, software modeling, business software integration.

1 Introduction and Motivation

Information systems are a crucial factor for companies in all lines of business. Within the years the requirements to do business have changed and got more complex. [18] Thus the companies' structures have to respond to these new situations with changing the way they are doing business. Traditional structures are no longer appropriate. Over the years organizations with traditional structures happened to create business departments as silos: They became big and shore less [9].This leads to blockades and information-silos. Therefore new ways of managing the organizations have to be found. Business Reengineering is one way to optimize the business processes according to the company's requirements responding to customer needs [9]. Especially the information system structure has to adapt to the new situation. A business process reengineering encompassed by a radical architecture change is necessary to gain long-lasting improvements.

The concept of IT as enabler of process change dates already back to [10] and [6] and is still on the forefront of process reengineering. A change in information systems is possible as the technological progress over the years has opened new possibilities to support the organizational reengineering. However, information system aspects have

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 104–115, 2012.

often been left out of consideration in reengineering projects. Information systems often have had and still have the status of being a matter of course and therefore their integration is often not thoroughly considered [11]. Thus companies are far too often behind in their information technologies. Software systems are often out-dated and poorly structured. Thus it is impossible to react with such systems to organizational changes. The need for agile software architecture to support business processes becomes evident when the need for more flexibility and reduced costs in the daily business urges companies to restructure the processes. Especially a system adaption is necessary to support the restructuring of the processes otherwise there will always be a break in the organization's structure. It, however, is often not enough to stick to a system adaption.

Although it is obvious that any reorganization project has to start with the company's strategy and should impact the business structure and strategy, it should not stop at this point. Fig. 1. illustrates that the components business strategy, business reengineering and software reengineering are ideally combined like gear wheels. Changes in any of the components influence the other parts, too.

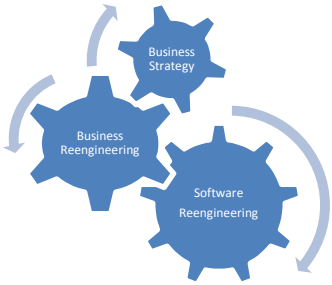


Fig. 1. Reengineering Interactions

Although the business strategy is mainly the impulsive force the other two components support the strategy to be fulfilled when they interact as a closed system. This is often left out of management considerations. The information systems part (illustrated as 'software reengineering') is not considered properly when changing the business strategy and structures, like processes and workflows, although the software influences the processes and the work of the employees to a high impact.

As technologies enable new software architectures, it makes sense to introduce not only standalone software systems but business software to support business processes. In this context Sundblad [19] states that business software is often introduced for exactly one reason: It should support the business and its activities to increase the productivity and efficiency of the business. The advantage of business software lies in the fact that business software, contrary to standalone software, can be integrated in all relevant business processes – like sales, production, after sales - to get a higher scale effect. Exactly this shift to business software means a huge paradigm change for companies all over. It is often not that easy to step into such a radical change project as the results are not easily predictable and case studies on a long time horizon are not available. However, introducing business software would support companies in

restructuring their business and making their processes more efficient. The most known and widespread business software is ERP (Enterprise Resource Planning) systems [23]. However, business software is not only restricted to ERP systems though a concentration on those systems is recognizable, due to the fact, that ERP systems are very common no matter from which provider, and the integration cycle is documented and rather similar no matter what's the company's business. Business software, however, can also be implemented for technical software systems. Thus a better process integration and support would be granted. A big hindrance in this area nowadays is that almost no examples of business software integration in the technical area exist.

This paper will therefore deal with this topic and investigates on the paradigm change to business software in the technical area within the automotive industry. Especially in the automotive industry many technical software systems exist, which could be replaced by integrated business software, which could then have a high impact on the workflow of the business processes. It should be noted, that the paper contains a brief outline and short description of a prototypical example of the specific business software to support the business processes in a very special use case, as the work is still in progress. Nevertheless it shows the general potential business software has on a process value chain within any organization.

2 Monolithic Software and Business Software Impacts

Many software systems are used over decades, with just slightly adapting the system. Business needs are not really considered in depth but only to a point where adaptations to existing systems are risky and dangerous. This fact is not a phenomenon within the automotive industry. This is quite a common phenomenon. Raisch [16] describes this phenomenon as "*Ambidexterity*". This means that an organization differs between two cycles of software reengineering: (1) the Exploration and (2) the Exploitation phase. The former signifies a phase where new systems are tested and introduced. This phase, however, requires a high level of readiness to assume risks, as many unknown components are predominant in this phase. The latter concentrates on exploiting existing systems as long as possible. Here just small and almost insignificant adaptations are made at the software systems. This means no risks and no changes to existing workflows. Fig. 2 illustrates how the two phases interchange each other. Normally the exploration phase is significantly longer.

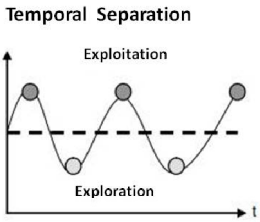


Fig. 2. Temporal Separation Phases in Software Development [16]

Companies wait extremely long, almost too long, before changing a system. Over the years the requirements and the user experience have changed and the technological feasibilities enable a completely different approach to problem solving. This evolution makes it easier for the user to cope with the difficult tasks and complexity could even be reduced. Therefore it is necessary to break up with existing system and to start from scratch in order to change the current situation. Monolithic software architecture is often of great hindrance instead of supporting the business processes. It does not support the business requirements and as standalone software it even makes the processes more complex due to many interfaces needed. Streamlining business processes is therefore hardly possible without a major overhaul of the software architecture that forms the basis for supporting the business processes.

3 Research Domain

The area of test facility automation systems in the automotive industry is an ideal research area for this paradigm change as test facility automation systems are very complex and consist of many software-intensive products. The goal of the research group has been to develop a framework for model-driven generation of automation system configuration parameters, which radically simplifies the configuration and operation of test facility systems.

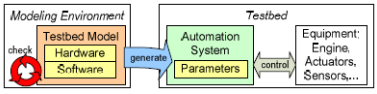


Fig. 3. Model-driven generation of automation system parameters [1]

This framework should then be the basis for business software to be implemented along the whole value chain. As a consequence, the associated business processes should be significantly streamlined.

3.1 Monolithic Architecture with Unstructured Parameters

Test facility systems need to be tailored to customer demands in a straight-forward way, which is neither well supported by the current monolithic software architecture nor does the software support the process chain.[13] The existing software has evolved over the past decades and comprises million lines of code, mainly written in C++ and C, which makes adjustments quite difficult. It requires error-prone editing of parameters (in the order of tens of thousands) in spreadsheet-like tables, as well as the adaptation of configuration files and scripts scattered in the file system. Thus the handling is time-consuming and expensive. Required changes, that would make the software systems flexible to business requirements, are neither possible with the existing software architecture nor is any potential effort for writing code for transforming the data justifiable. A software engineering has to be launched to make

the system flexible to current requirements. Therefore it is highly recommended that the new tool supports the whole processes. A lot of information, which is nowadays entered into the tool in the project execution phase already have to be available in an earlier stage, e.g. in the sales phase. Therefore it is useful to rethink the product integration and use the tool as business software along the whole process chain.

3.2 Current Business Process Landscape

The current software is restricted to the project execution process without having any connection to the up- and downstream processes. Fig. 4 illustrates this situation. The handover points from one process to the other are without any tool support. Thus relevant information might be lost throughout the process chain and information needed in the progression of the project is not available as the importance of these data might not be clear in an earlier phase.

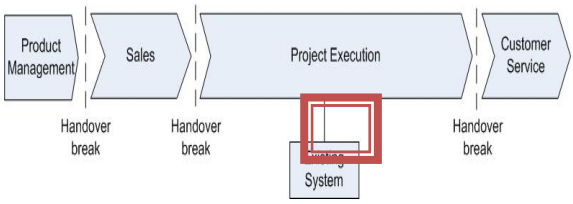


Fig. 4. Position of the current software system

Currently, each process uses different tools, like Excel sheets or proprietary tools. These gaps cause a significant information loss and extra manual conversion and transfer efforts.

4 Business Software Modeling Approach

To show the feasibility of any business software concept both sides the business processes and the software engineering have to be considered. It is crucial that a software system matches the requirements of an organization and that the software enables the business processes to adapt to business requirements. For the described research area the overall complexity of test facility systems demands for business software that applies stepwise refinement. The aim is that a unified method and tool set is established, with which everybody throughout the process chain can interact. The resulting model of the overall test facility should then correspond to the real-world test facility in all its relevant details.

The principal idea is to reduce the complexity through an object-oriented model which enables a graphical 1:1 representation of real-world items of an engine test facility. Thus, this kind of modeling is referred to as deep virtualization of the corresponding real-world test facility. In the beginning a coarse-grained model can be defined which is incrementally refined along the value chain. Models are the main

artifacts describing a system under test, and a model at a certain level of abstraction can be transformed into another model at a possibly different level of abstraction: the hypothesis is that a software system that allows a stepwise modeling of a test facility automation system would support the overall business process chain and thus forms the basis for process streamlining and cost cutting.

By examining the examples of how to model an engine and its attached sensors, problems of conventional modeling languages such as UML [15] were identified. Instead the decision was made for what Atkinson and Kühne [3] have called clabject-based modeling, an approach that unifies the notion of classes and objects. Each model element has a compartment for the name, and a combined compartment for the type facet and the instance facet. The dashed arrows between the levels represent the "instance of" relationship. With a uniform representation of type facets and instance facets, our example can be modeled in a natural way. By definition, the clabjects at the top-level only have a type facet, whereas the clabjects at the bottom level only have an instance facet. [2] The advantage hereby is that the domain engineers do not have to concentrate on specific differences between instances, classes and object. They can mainly concentrate on the engineering of the domain elements.

With the clabject-based approach a library of all relevant elements is developed. The element library is then for the software usage relevant. It is filled in the beginning of the value chain (in the product management) and then used throughout the processes. The modeling of the product library, however, is essential for the whole product management and can also enable a better product management concerning a reduction of the complexity of product variants. (A separate research is initiated in the product variants area.)

Thus this approach supports the stepwise refinement, as people from different departments can work with the tool and without having to know the underlying software architecture.

5 Process Reorganisation through Business Software

A typical business process chain, shown in Fig. 5, covers aspects from product management, sales, project execution to the customer service. Each process has its special requirements and associated roles on the tool.

Fig. 5. Schematic Process Chain

5.1 Business Software Integration Method to Structure the Reorganization Project

Integrating business software into the tool landscape is encompassed by a lot of changes to the existing landscape. Introducing a new software modeling environment

implicates potential for conflict and opposition which should not be underestimated. When the new software modeling approach should furthermore be operated as business software, it is an even more challenging task. It has to be done very carefully, also considering psycho-social effects of performing a change project within an organisation, as introducing new software especially as business software is a huge change project.

As mentioned in the introduction ERP implementation projects exist manifold [17]. However, for introducing a business software in the automation area there are almost no use-cases available and therefore, within the context of the project, a method has been developed to structure this change process. Fig. 6 shows the main phases of the BSI (Business Software Integration) method on a chronological timeline.

Fig. 6. Business Software Integration Method (BSI-Method)

The BSI method has been developed according to Lewin's [12] three typical phases for change. The three phases are: (1)Unfreezing, (2) Moving and (3) Refreezing.

The introduction of business software is mainly divided into three separate phases, which, nevertheless, interact with one another. Referring to Lewin [12] the different phases correlate as follows: (1) The first phase—decision making—corresponds to the unfreezing phase. Here the management decision to change the status-quo has to be made. (2) The second phase —analysis & requirements —represents to some extend also the unfreezing and in more detail the moving phase. (3) The development and implementation phase corresponds to the refreezing phase. It is split into two phases as the development phase is more complex and needs a separate focus.

Each of these single phases consists of several individual steps. Although the stepsare adapted to the organisation's situation, there are some common steps which should be included in the corresponding phase. The most important steps are highlighted in Fig. 7.

Fig. 7. Phases of the BSI-method with corresponding steps

This method can be seen as manual for any technical business software integration project, which all will mainly have these phases. The steps, however, can differ

according to the maturity level of the organisation's processes and the software landscape. Each of these phases is then supported by adequate methods. Although each of these phases are not new, the combination of these phases while introducing a companywide business software has rarely been discussed in literature as no technical business software has up to now been implemented.

5.2 Benefits of the Modeling for the Process Chain

This clabject-based modeling approach can be integrated as business software along the process chain. Fig. 8 demonstrates the core benefit of such a modeling tool. Through the integration as business software, indicated as a constant banner (refered to as DeepVTool in Fig.8), it supports all essential processes and it will have predefined interfaces to other tools, which are relevant for the processes. While the new business software is the technical backbone, the other tools, indicated as tool 1 till tool 4, are mainly used for economical matters, like a sales tool or a calculation tool. All technical solution components from the business software can be automatically exported to other tools when needed. Thus, the workflow is well supported and needed information can be easily accessed at the right time.

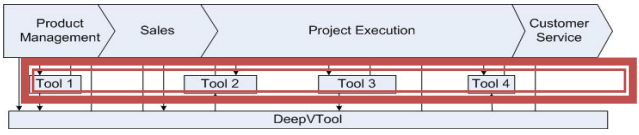


Fig. 8. Envisioned tool supported process chain

Another effect of the integration of the business tool is that several monolithic standalone tools can also be completely replaced. Thus the software landscape will be cleaned up. Thus streamlining makes the work of the people easier, as fewer tools have to be used. On the other hand the reduction of tools is also significantly displayed in the cost structure of the IT department.

Realizing this clabject-based modeling approach also has manifold benefits for the whole workflow of an organisation, which are based on some well known concepts:

- Reusability
- Correctness
- Quality assurance
- Visualization

The application of these concepts has its individual benefits on the different processes within an organisation. Therefore all relevant information should be added gradually to the test facility model in each process phase according to the particular requirements. Thus the level of detail represented by the model increases over the project progression. Fig. 9 shows the step by step refinement schematically: (1) The Product Management is responsible to have all components available for usage in the tool. (2) In the Sales phase general project information and requirements are applied in the model. Thus a first visual test facility system model is generated, which can be

used as discussion base with the customer. (3) In the Project Execution phase important data and details are added, according to their relevance. For example, detailed information on the utility requirements (power supply, water supply, …) are added. (4) In the operations phase the most detailed test facility model is used to configure and operate the whole system.

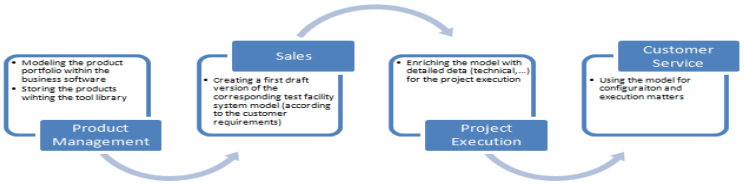


Fig. 9. Step by step refinement along the process chain

A further advantage is that errors or dependencies in the design and development of test facilities can be identified already at an early stage in the sales phase, avoiding costs that are due to late detection and fixing of such errors in the subsequent project execution phase.

The changes in the workflow also require a change in the role model of the existing processes. It is an advantage of the tool that it is not intended to be a tool for experts only. Employees working in all business areas –from sales to operator engineers - should use the tool in their context. The granularity and level of detail of information however will be adapted to the role concept. Thus the business and software reengineering has also a high effect on the people within a company. It is not only that they have to change what and how they do but the basic ways how they think is altering. Therefore any reengineering project should be accompanied by a special communication strategy within the company to prepare all people affected by the change. Describing the communication strategy within our research project would go beyond of the paper's scope.

5.3 Quantitative and Qualitative Effects

In general there are two relevant points where first savings can be achieved: (1) a reduction of cycle times can either be achieved with an elimination of process steps or (2) with the shift of process steps to upstream processes. The first one can be achieved, as the system offers the possibility that process steps are done automatically. The latter one enables a cost reduction as several work steps can be done in house, which is obviously cheaper, than doing the same task at the customer's location.

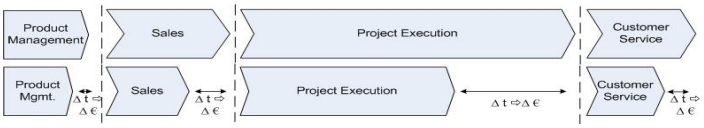


Fig. 10. Process chain streamlining

First analyses show that savings between ten and fourteen per cent can be achieved. This first analysis, however, is only based on scenarios where the business software usage is mainly concentrated on the project execution phase. With a process-wide implementation, as shown in Fig. 10, even more savings can be achieved. Costs savings in a two-digit percentage range might be.

Furthermore the integrated tool with its visualization approach can also increase the customers' satisfaction, as the tool handling becomes easier through the 1:1 representation. The satisfaction will also increase with internal customers. Training times for new employees can be radically shortened, which has on the one hand a motivational effect on the employee and on the other hand a positive financial effect for internal and external customers.

6 Related Work

Addressing the paradigm to introduce a "technical" business tool is generally not dealt with in literature. There are mainly publications dealing with ERP and CRM (Customer Relationship Management) software solutions, compare e.g. [4], [18] and [23]. The introduction of a new software tool and introducing the tool in the organisation as business software, to make use of all the advantages through the integrative approach are not at all discussed.

Technical business software systems, however, are not yet described in literature. Literature on a business software introduction and the effect on the organisation are almost not available. With our project we investigate in both areas: We develop and integrate new software system architecture and investigate on the effects of this business tool on the business processes and thus help to streamline the business processes. An area which is often discussed in literature, which is familiar to our topic, is the whole are of business or enterprise engineering. In the area of business engineering methods mainly concentrate on change projects and rarely describe the effects of this engineering on operations for longer times [8]. It, however, is not so deeply on a specified alignment of IT and business aspects. Furthermore there exists literature on enterprise engineering or parts of enterprise engineering, like [4]. This approach, however, is reduced to only one little aspect of the usage of information systems. Further approaches in the area of enterprise modeling are described and dealt with e.g. by [7]. Thus our research approach deals with an important and quite new topic. The idea of visualization realized in the software architecture already proved right in e.g. the building industry. [7,20]

7 Conclusion and Future Work

In this paper the paradigm change from monolithic standalone software to business software is described in the context of the automotive industry. Therefore a software reengineering is necessary and the therein used architecture is described in this paper. It described the effects of a business tool usage on the processes and on the company as a whole. We argue that monolithic software architectures impede business

processes to adapt to changing requirements whereas business software based on a so called clabject-based modeling approach offers more flexibility and thus forms the basis of a business process reorganisation that enables major improvements.

Generally speaking it can be stated and also proven with first results, that an integrated business tool instead of many monolithic software systems can have an enormous effect on the business process within an organisation. It does not matter in which line of business business software is used. The applied research in the area of industrial automation systems proves applicable and undermines the positive effects of business tool usage. This approach promises to overcome shortcomings in legacy systems and most important it supports business processes.

References

1. Aschauer, T., Dauenhauer, G., Pree, W.: Towards a Generic Architecture for Multi-Level Modeling. In: Proceedings of the 2009 Joint Working IEEE/IFIP Conference on Software Architecture & European Conference on Software Architecture WICSA/ECSA 2009, pp. 121–130 (2009)
2. Aschauer, T., Dauenhauer, G., Pree, W.: Multi-Level Modeling for Industrial Automation Systems. In: Proceedings of the 35th Euromicro Software Engineering and Advanced Applications (SEAA 2009) Conference 2009, Patras, Greece, August 27-29 (2009)
3. Atkinson, C., Kühne, T.: Model-Driven Development: A Metamodeling Foundation. IEEE Software 20(5), 36–41 (2003)
4. Bucher, T.: Ausrichtung der Informationslogistik auf operative Prozesse. Kovac Verlag, Hamburg (2009)
5. Chandler, A.D.: Strategy and structure: Chapters in the history of the industrial enterprise, 22nd print. Edition. MIT Press, Cambridge (2001)
6. Davenport, T.: Process Innovation: Reengineering Work Through Information Technology. Harvard Business School Press (1993)
7. Eastman, C., Teucholz, P., et al.: BIM Handbook. A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers, and Contractors. Wiley & Sons Inc., New Jersey (2008)
8. Frank, U.: Multi-Perspective Enterprise Modeling (MEMO) - Conceptual Framework and Modeling Languages. In: Proceedings of the Hawaii International Conference on System Sciences (HICSS-35), Honolulu 2002, 10 p (2002)
9. Gadatasch, A.: Grundkurs Geschaeftsprozessmanagement. Vieweg + Teubner, GWV Fachverlage GmbH, Wiesbaden, Germany, 5., erweiterte und überarbeitete Auflage edition (2008)
10. Hammer, M., Champy, J.: Reengineering the Corporation: A Manifesto for Business Revolution. Harpers Paperbacks. Rev. Upd. (2003)
11. Kawalek, J.: Rethinking Information Systems in Organizatons. Integrating Organisational Problem Solving. Routledge, New York (2008)
12. Lewin, K.: Researching Organizational Values and Beliefs. Kindle Edition (2001)
13. Martyr, A.: Engine Testing: Theory and Practice, 3rd revised edn. Butterworth Heinemann (2007)
14. Muthum, S., Withman, L.: Business Process Reengineering: A consolidated methodology. In: Proceedings of the 4th Annual International Conference on Industrial Engineering Theory, Applications and Practices, Texas (1999)

15. Object Management Group: Unified Modeling Language Infrastructure. Version 2.1.2. (2007)
16. Raisch, S.: Balanced Structures:Designing Organisations for Profitable Growth. In: Long Range Planning, pp. 483–508 (2008)
17. SAP Website (2010), http://www.sap.de
18. Semar, W.: Weiss + Appetito Holding AG - Software für den ganzen Konzern. In: Wölfle, R., Schubert, P. (Hg.) Dauerhafter Erfolg mit Business Software - 10 Jahre Fallstudien nach der Experience Methodik, pp. 51–64. Carl Hanser, München (2009)
19. Sundblad, S., Sundblad, P.: Business Improvement through better architected software. Microsoft Architect Journal 10 (2007)
20. The Economist. From blueprint to database (2010), http://www.economist.com/
21. Ungar, D., Smith, R.B.: Self: The power of simplicity. In: Proceedings of OOPSLA 1987. ACM SIGPLAN Notices, vol. 22, pp. 227–242 (1987)
22. Winter, R.: Business Engineering. In: Integrierte Informations Systeme, pp. 17–37. Springer, Berlin (2008)
23. Wölfe, R., Schubert, P. (eds.): Integrierte Geschäftsprozesse mit Business Software. Hanser Verlag, Munich (2005)

An Inventory of the Business and IT Alignment Research Field

Julia Kaidalova and Ulf Seigerroth

School of Engineering, Jönköping University
P.O. Box 1026
551 11 Jönköping, Sweden
{julia.kaidalova,ulf.seigerroth}@jth.hj.se

Abstract. The area of business and IT alignment (BITA) has received increased attention during the last decade. As a consequence of this the amount of literature has also been growing increasingly. Since the interest for the field has grown it has also become important to follow and monitor trends and dominant directions of ongoing research within the field. The purpose of this paper is therefore to make an inventory of areas and directions that have attracted our attention in the literature. As a result of such inventory we present a set of categories that shows the focus and diversification of the BITA field. In the findings we can observe that the area of the highest research interest is development of a new instrumental support (methods, frameworks, approaches, etc.) for BITA. One dimension that seems to be partly neglected in the literature is procedural guidance about HOW to conduct and achieve business and IT alignment. The existing instrumental support that are presented in the literature has a strong focus on WHAT to do and WHAT to deal with (regarding both theoretical and practical issues concerning BITA), but they are missing the HOW dimension to a large extent.

Keywords: Business and IT alignment, Instrumental support, Literature review.

1 Introduction

A key issue in today's enterprise functioning is information technology (IT) that supports business needs, processes and strategies (Silvius, 2009). Nowadays, when a broad variety of CASE tools and IT systems are available for business functioning support, it becomes more and more complicated to control and follow their dynamic development and transformation and even more difficult to keep track of those IT tools that could potentially support current enterprise functions. The importance of using appropriate and effective IT means to facilitate business functions have been acknowledged and discussed by practitioners and scientists over the past two decades (Vargas, 2011). It has become a crucial issue for enterprise success – IT systems that fit to the business needs. The problem in this context is all the more complex due to the dynamic and evolving nature of both sides – business and IT (Luftman, 2003). To address this problem practitioners and researchers have used such terms as

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 116–126, 2012.

"harmony," "linkage," "fusion," "fit," "match", "integration", but in the long run the term "alignment" has been put to use. Luftman (2003) explains Business and IT Alignment (BITA) as a problem of how to get technical and business people to interpret things in the same way and to understand each other's side.

In the research community BITA is often addressed as a top concern of IT and business practitioners (Chan and Reich, 2007; Luftman and McLean, 2004). It is possibly caused by scientific and practical recognition of organizational benefits that BITA can bring to the table (Vargas, 2011). Researchers have discussed different aspects of aligning business and IT: initially the problem is often studied as relating business plan and IT plan; later on BITA have been considered as linking business strategy with IT strategy. Eventually the literature started to propose frameworks that could combine several aspects of BITA. One theory that is often referred to as the foundation model of the field is the Strategic Alignment Model (SAM). The essential idea in SAM is that organization should strive for alignment between four areas: business strategy, IT strategy, organizational infrastructure and IT infrastructure. In this relation SAM emphasize strategic fit and functional integration. Strategic fit is linking strategy and infrastructure for both business and IT, functional integration is fitting together business and IT strategy and business and IT infrastructure (De Haes et al, 2010).

Interestingly enough, the number of unsolved problems within BITA does not fall over time, which determines the increasing attention to this area. Several researchers indicate growing interest to the BITA within both the academic domain and the practice domain (Luftman, 2003). The number of aspects that are noticed and studied within BITA keeps increasing over time (Chan and Reich, 2007). Scholars propose models and frameworks that interpret BITA through various dimensions. Another popular direction of the research is supportive guidelines (for example, methods, frameworks, approaches, etc.) that can help to achieve, sustain or improve BITA (Seigerroth, 2011). Among other issues under discussion in the literature are factors that can influence BITA and various case studies.

With enduring attention to the question the amount of related literature is growing increasingly, therefore, it becomes more important to keep track of publications regarding BITA, since it is necessary to detect and follow the trends and important issues in the field. Thus, the purpose of this paper is to describe the state of the art within the domain of BITA on overall level. More specifically, the aim is to investigate the literature related to business and IT alignment in order to make an inventory about what is in focus in this research domain. The research question of this work is the following:

What are the main interest areas in business and IT alignment research?

The rest of the paper is structured in the following way: Section 2 describes the research approach that have been applied to perform our literature study (inventory), in Section 3 results are presented and analyzed, furthermore, directions for future work and validity threats of the study are discussed. The paper is then ending with conclusions in Section 4.

2 Research Approach

The approach that is used in the research has involved four stages: first it was required to define the aspects to outline the selection of literature, then extracting the papers according to defined selection parameters, further categorization of extracted papers and generation of results, which is concluding with answering the research question. The outline that depicts research approach is presented in Figure 1.

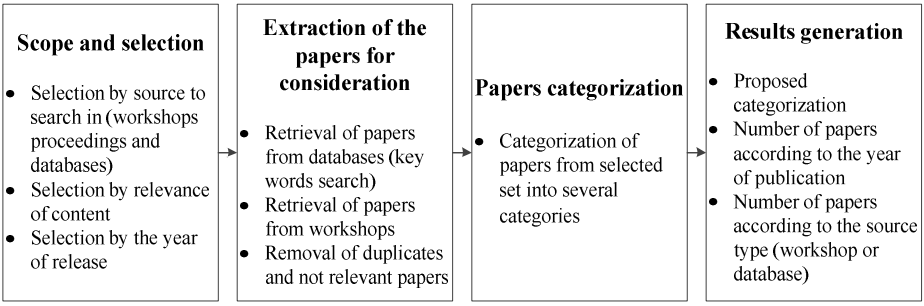


Fig. 1. Research approach

2.1 Scope and Selection

An important question to answer before performing this research was about the literature coverage. Papers for further consideration have been selected according to the following aspects:

1) Sources to extract papers
2) Relevant content
3) Time of publication

The first issue is related to sources that have been used to retrieve papers for consideration. There are several sources that have been used to seek through for relevant papers. It was decided to browse two types of sources: scientific databases (SpringerLink, ACM Digital Library and Emerald) and conference/workshop proceedings (BITA and BUSITAL). The reason to work through indicated scientific databases is that all three of them involve solid collection of scientific papers from variety of research areas and allow to use search interface that simplify the identification process. The reason to consider works that have been presented on BITA and BUSITAL workshops was that these workshops are recognized thematic events in the BITA field. Subsequently, works that have been presented there are most likely relevant for our research.

The second issue regarding relevant content of the papers has been settled in the following way. Taking into consideration increasing number of works related to BITA that are available via scientific databases, it was quite natural to limit the search using the keywords. Pondering over the most suitable keyword compositions, it was noticed that the literature shows slightly different ways to name alignment between

business and IT: business-IT alignment (for example, Luftman and Brier, 1999), business/IT alignment (for example, De Haes and Van Grembergen, 2009), IT/business alignment (for example, Saat et al., 2011), business and IT alignment (for example, Wegmann et al, 2005). Thus, it was decided to search for "Business IT alignment" within aforementioned databases. It was also decided to consider only those papers that had required phrase only in the title. Works that have been presented on BITA and BUSITAL workshops have been taken for granted in terms of relevance, thus they did not need any sorting by key words.

The third aspect to limit the extent of the literature coverage was the time of publication. According to Luftman and Brier (1999) the importance of alignment has been well known since late 70s, whereas Schlosser et al. (2012) assert that the beginning of intensive development of BITA research falls on the early 90s. It is also well known that researchers have used different terms to talk about alignment of business and IT - integration, fit, strategic alignment, harmony and other terms (Mendoza, 2009). It is fair enough, since just as any other research field, the field of BITA was developing and elaborating its own terminology. It seems reasonable within present research to consider papers that have been published after the BITA terminology has been established. Maes et al. (2000) criticize BITA research of the day because of its ambiguous nature: "In general, alignment is defined in an indefinite and vague way, if at all!" (Maes et al., 2000, p.7). Thus, we strongly believe that this point of time (year 2000) can be considered as a reference point for elaboration and unification of BITA research domain. We inclined to consider that approximately at this point BITA research area has crystallized dominating research directions and started to use more full-fledged and elaborated terminology. Thus, during present literature review we will consider papers that have been published within time frame of years 2000-2012.

2.2 Extraction of the Papers for Consideration

Extraction of papers for further categorization has been done taking into account aspects that are described above in the section 2.1. We have searched for "Business IT alignment" phrase within titles of papers from three databases (publication years from 2000 to 2012). Papers that have been presented on BITA and BUSITAL workshops have been included into the preliminary set of papers for further categorization without any filtering, since it was assumed that they are a priori relevant to the BITA research field. Preliminary number of papers that were supposed to be categorized is presented in Table 1.

Table 1. Preliminary number of papers for further consideration

	SpringerLink	ACM Digital Library	Emerald	BITA	BUSITAL	Total
Number of papers	43	53	21	10	49	176

After removing duplicates the total number of papers has decreased to 162. One more abridgement of preliminary set of papers has been done by removing false positives – works that do not deal with BITA, but that for one or another reason has been included into preliminary set. The number of papers has then decreased to 138.

2.3 Papers Categorization

After the set of relevant papers has been outlined, it was needed to categorize them. This was done simultaneously with the progress of reading papers. The process of creating categorization was inductive – papers have been studied one by one, with adding logical tags to each of them. Afterwards it was possible to gather papers of corresponding sort into one category, with distinguished differentiation between categories. Along the categorization process the following interest areas within BITA research domain evolved into what is presented in the list below:

- Paper with aim to develop BITA instrumental support (for example, method, framework, strategy, tool or other)
- Paper with aim to evaluate existing BITA instrumental support (for example, method, framework, strategy, tool or other)
- Paper with aim to apply existing BITA instrumental support (for example, method, framework, strategy, tool or other)
- Paper with aim to identify factors that can be used to influence or give indication of BITA (quite often these two types of factors have been discussed within one work)
- Paper with aim to study current state of BITA research field
- Paper that studies other issues of BITA

The elaborated content of this categorization is presented below in the section 3.1.

2.4 Generation of Results

After categorization has been done it was possible to collect the results. In order to answer the research question it was decided to analyze the results qualitatively from two points of view: number of papers per category according to the year of publication and number of paper per category according to the source type (database or workshop), see result of this in section 3.

3 Results and Discussion

3.1 Analysis of Results

Results of the work consist of two parts: the set of categories that has been generated to categorize relevant papers from the BITA research field and quantitative data that have been obtained from this categorization. Categories that have been formed are presented in Table 2 below, with corresponding description for each category.

Table 2. Categories and corresponding definitions

Category – focus in paper	**Definition**
1. Development of a new instrumental support item for BITA*	The purpose of the work is to develop instrumental support for BITA, for example, method, approach, strategy, framework, model, tool or other item. Often development of a new supportive item is accompanied by description of its usage, but the work emphasis is done on creation of it. The category also involves papers that present modified supportive items (for example, a new framework that have been produced by modifying existing one).
2. Evaluation of existing instrumental support item for BITA	The work purpose is to evaluate existing BITA instrumental support item. Evaluation of an item can be accompanied by description of its application, but the emphasis is done on the analysis or evaluation of it.
3. Application of existing instrumental support item for BITA	The purpose of the work is to apply existing BITA instrumental support item. Quite often it is a case study work that presents application of BITA supportive tool in practice. The category also involves papers that describe application of instrumental support from other areas for BITA. The category does not introduce procedural guidelines of how to apply one or another BITA method or tool, rather describes particular case of application.
4. Key performance indicators for BITA	The purpose of the work is to describe factors that can be used as key performance indicators for BITA. Often among them are mentioned: • Factors affecting BITA • Enablers and inhibitors of BITA • Effects of BITA • BITA antecedents and outcomes
5. BITA state of the art study	The purpose of the work is to study current state of BITA field. The category also involves papers that introduce framework or model that can be used to structure BITA literature in one or another way.
6. Other BITA issues	Work describes an issue that does not deal with BITA straightly, but that has indirect relation to BITA. For example, work can discuss specific problems and/or solutions that exist within a certain BITA research aspect. Alternatively, work can discuss problems of different research area, but describe the relation that these problems have to BITA issues. Work also can explain the potential influence of some factor on BITA, but does not clearly prove it (in contrast to works from category "Key performance indicators for BITA", where relation of these factors to BITA is grounded and justified).

* Instrumental support can include one or several of the following items: method, approach, strategy, framework, model, tool (Seigerroth, 2011)

As it was mentioned above, the set of relevant papers has been sorted according to created categories (1-6). Table 3 shows number of papers per each category by publication years. The highest number of paper for each year is marked in bold.

Table 3. Number of papers in each category (1-6) according to the year of publication

	1	2	3	4	5	6	Number of papers per year		
							DB	WS	Total
2000	0	0	**1**	0	0	0	1	0	**1**
2001	**1**	0	0	0	0	0	1	0	**1**
2002	0	0	0	0	0	0	0	0	**0**
2003	**1**	0	0	0	0	0	1	0	**1**
2004	0	0	0	**2**	0	0	2	0	**2**
2005	**3**	0	0	0	0	0	3	0	**3**
2006	**6**	0	0	3	0	2	3	8	**11**
2007	2	1	3	3	0	**6**	8	7	**15**
2008	**7**	1	3	3	1	**7**	13	9	**22**
2009	**9**	1	0	7	2	7	19	7	**26**
2010	**8**	2	4	4	1	5	11	13	**24**
2011	**12**	4	3	3	1	8	16	15	**31**
2012	0	0	0	0	1	0	1	0	**1**
	49	9	14	25	6	35	79	59	**138**

It is noticeable that the total number of papers per year has stable tendency to grow yearly – from year 2000 with 1 paper to year 2011 with 31 papers. For transparency reason number of papers per year is represented with the help of three numbers: number of papers that were retrieved from databases (DB column), number of papers that were retrieved from workshops proceedings (WS column) and total number of papers. This is done in order to see both overall decrease or growth tendency and this tendency depending on the source type. Interestingly enough, number of papers that were retrieved from databases tends to increase yearly with slight deviations, the same as total number of paper per year irrespective the source type. The distribution of papers into categories is uneven, which illustrates quite natural research phenomenon of higher interest to one research direction and comparatively lower attention to another one. The dominant area of research interest is development of instrumental support for BITA.

The largest by number of papers category consists of papers that develop new item of instrumental support to facilitate BITA. Among items for instrumental support have been introduced various methods, approaches, strategies, frameworks, models, CASE tools and other artifacts that can provide guidelines for BITA achievement, improvement, maintenance or assessment. In some cases new instrumental support items have been introduced as a solution for well-known BITA problems by modifying or improving existing ones, whereas another part of papers has introduced a solution for newly discovered BITA problems. It is also noticeable that, in the same

way as the total number of BITA publications, the number of instrumental support items steadily grows year by year - from year 2001 with 1 paper to year 2011 with 12 papers regarding this topic, reaching total number of 49 papers within considered period of time. A question that could be relevant to ask is "Do we need yet another method?".

In order to understand what kind of instrumental support is being introduced within the field it was decided to divide papers from this category according to the character of proposed instrumental support. One type of work has the purpose to develop instrumental support that involves procedural guidelines for BITA, whereas other type of work is rather creating notions or models that can facilitate BITA. These two types of instrumental support have originated by drawing an analogy with idea of Stirna and Persson (2007) that assert equal importance of two aspects involved into enterprise modeling. They argue that to obtain high quality results both modeling process and modeling language are important. According to our analysis, the number of papers that introduce some sort of procedural guidelines is much fewer than the number of other papers (see Table 4 below).

Table 4. The character of presented instrumental support

Procedural guidelines	Non-procedural guidelines
17	32

Quite minor attention in the BITA research domain is paid to evaluation of existing item for BITA instrumental support. Papers that belong to this category mostly deal with discussing strengths and weaknesses of existing BITA methods or frameworks. Comparatively higher research interest is dedicated to application of existing BITA instrumental support items. This category involves work that describes applications of BITA methods or tools in practice, but it does not propose procedural guidelines for how to apply different instrumental support. Number of papers per category is 9 and 14 respectively. Works that applies existing BITA instrumental support items mostly presents case studies of using one or another BITA framework or method in practice. Interestingly enough, papers that develop a new instrumental support items quite often describe practical application of them too, but in contrast to this type of papers, such application is not their focal research issue.

The second largest category includes work that deals with other BITA issues. This kind of work often describes minor or particularistic issues of BITA - issues that have indirect relation to BITA. Quite often authors describe problems that have originated from the BITA area, but have eventually segregated into separate research fields and started to require particular scrutiny. Alternatively, authors illuminate an issue from another research domain, but discuss potential relation of it to BITA. Within studied set of papers this type of papers started to emerge in 2006. Among concepts that are considered within papers from this group it is possible to mention the following: business requirements, requirements engineering, IS requirements, business needs, business model, goal model, value model, business process model, business process management, enterprise architecture, enterprise modeling, service modeling and others.

The third largest category consists of work that describes key performance indicators for BITA. The purpose of such work is to describe factors that can be used as influencing or indicating BITA. Often these two types of factors have been discussed together. Among issues that are being debated are factors affecting BITA, enablers and inhibitors of BITA, effects of BITA, BITA antecedents and outcomes and others. The number of papers in this category has a tendency to grow too, but quite irregularly, so that it reached its peak in 2009 with 7 published works.

The smallest category in number of papers is BITA state of the art. This category has 6 papers that have been presented within studied period of time.

Another part of extracted quantitative data includes number of papers per each category by source type, where DB stands for databases, WS stands for workshops (see Table 5). The highest number of papers from each source type is marked in bold.

Table 5. Number of papers in each category according to the source

	Development of a new instrumental support item for BITA	Evaluation of existing instrumental support item for BITA	Application of existing instrumental support item for BITA	Key performance indicators for BITA	BITA state of the art study	Other BITA issues	Total
DB	**35**	8	12	24	6	0	85
WS	14	1	2	1	0	**35**	53

The most significant finding that we can draw from this view on data is that papers that have been presented on workshops are mostly dealing with minor or particularistic issues of BITA. It is quite reasonable, since such thematic event as workshop has purpose of discussing the interest area in details. It leads to quite significant diversification of the research directions and elaboration of specified problems closely and particularly.

3.2 Future Work

Potential direction for further work is elaboration of presented categorization by adding new layers to it. These layers can involve the following aspects:

1) What types of problems is BITA instrumental support supposed to solve?
2) For what processes can BITA instrumental support serve as guidelines? (BITA achievement, improvement, maintenance, assessment, etc.)
3) What nature does BITA research domain have? (theoretical vs. practical)

3.3 Source of Errors

There are several threats to validity in our research. First, categories have been generated in parallel with reading selected papers. It is possible that another category of BITA research domain does exist. We could possibly not include it in our

categorization, since considered set of papers might simply not have papers that belong to this category.

Categorization of papers as it has been done in this research is a materially subjective process, since it is a result of interpretation of phenomenon by human. However, we tried to avoid it by providing detailed definition of each category, so that they are clearly distinguished from each other.

We have also limited out search within databases by choosing only those papers that have required key words in the title. It is double-edged sword: from one point of view it increased the probability to get mostly relevant papers, but from another point of view it prevents getting broader picture of BITA research area. Thus, we cannot guaranty that considered set of papers provides comprehensive view over the BITA field.

4 Conclusion

The purpose of this paper is to provide overall picture of the BITA research domain in terms of the main interest areas of BITA literature. To achieve this we have developed a set of categories that were used to classify the literature in the field. Categories that have been generated are the following: development of a new instrumental support, evaluation of existing instrumental support, application of existing instrumental support, key performance indicators, state of the art studies, and other BITA issues. These categories do not indicate very specific aspects of the BITA research domain, thus, the created categorization can to a large extend be considered as a quite general way to distinguish research directions within the field. This is due to the purpose of the paper to get an overall and high-level description of the research domain without going into details of each research direction. Elaboration of presented classification will be one aspect to pursue in future work.

Evidently the total number of papers per year has a stable tendency to grow. Research direction that attracts the highest attention consists of work that develops new instrumental support artifacts for BITA. Such work focuses primarily on WHAT to do and WHAT to deal with in order to facilitate BITA, in terms of both theoretical and practical issues, whereas dimension of providing procedural guidelines about HOW to conduct and achieve BITA seems partly neglected. The category of papers that takes the third place by number consists of work discussing factors that influence BITA or indicate that BITA has been achieved on one or another way. We strongly believe that these two categories are related to each other in sense that both categories aimed on presenting derived know-how. Evidently, the research community tends to intensively propose theoretical and practical solutions for various problems that emerge within the area. However, a question that could be relevant to ask is “Do we need yet another method?”.

Another area of relatively high interest from research community involves various minor and specific issues of BITA. Interestingly, this category consists basically of papers that have been presented on thematic BITA workshops. It is logical and can be explained by the essence of such thematic event as workshop, since it has to gather

and present all the work from the field and its branches, which causes substantial segregation of the research discussion within respective workshops.

A research direction that has attracted relatively low attention from the research community is study of BITA state of the art. One possible reason for it is the orientation of the BITA research, which tends to be rather practical.

References

1. Chan, Y.E., Reich, B.H.: IT alignment: what have we learned? Journal of Information Technology 22, 297–315 (2007)
2. De Haes, S., Van Grembergen, W.: An Exploratory Study into IT Governance Implementations and its Impact on Business/IT Alignment. Information Systems Management 26(2), 123–137 (2009)
3. De Haes, S., Haest, R., Van Grembergen, W.: IT Governance and Business-IT Alignment in SMEs. In: ISACA, vol. 6 (2010)
4. Luftman, J., Brier, T.: Achieving and Sustaining Business-IT Alignment. California Management Review (1999)
5. Luftman, J., McLean, E.R.: Key issues for IT executives. MIS Quarterly Executive 2004 3(2), 89–104 (2004)
6. Luftman, N.J.: Measure Your Business-IT Alignment. Optimize: Business execution for CIOs Magazine (26) (2003)
7. Maes, R., Rijsenbrij, D., Truijens, O., Goedvolk, H.: Redefining business – IT alignment through a unified framework. PrimaVera Working Paper Series, Universiteit van Amsterdam, Department of Information Management (2000)
8. Gutiérrez Mendoza, A.S.: Alignment of IT projects with business Strategy. School of Information Systems, Computing and Mathematics Brunel University (2009)
9. Saat, J., Winter, R., Franke, U., Lagerström, R., Ekstedt, M.: Analysis of IT/Business Alignment Situations as a Precondition for the Design and Engineering of Situated IT/Business Alignment Solutions. In: 44th Hawaii International Conference System Sciences, HICSS (2011)
10. Schlosser, F., Wagner, H.-T., Coltman, T.: Reconsidering the Dimensions of Business-IT Alignment. In: 45th Hawaii International Conference on System Sciences (2012)
11. Seigerroth, U.: Enterprise Modelling and Enterprise Architecture: the constituents of transformation and alignment of Business and IT. International Journal of IT/Business Alignment and Governance (IJITBAG) 2 (2011) ISSN 1947-9611
12. Silvius, A.J.G.: Business and IT Alignment: What We Know and What We Don't Know. In: International Conference on Information Management and Engineering (2009)
13. Stirna, J., Persson, A.: Ten Years Plus with EKD: Reflections from Using an Enterprise Modeling Method in Practice. In: Pernici, B., Gulla, J.A. (eds.) Proceedings of the 11th International Workshop on Exploring Modeling Methods in Systems Analysis and Design (EMMSAD 2007), pp. 99–108. CEUR-WS.org (2007)
14. Vargas, J.O.: A Framework of Practices Influencing IS/Business Alignment and IT Governance. School of Information Systems, Computing and Mathematics in Brunel University (2011)
15. Wegmann, A., Regev, R., Loison, B.: Business and IT Alignment with SEAM. REBNITA, RE 2005, France (2005)

A Critical to Quality Factors Choice: An Integrated AHP-QFD Model for Information Quality

Abrar Haider and Sang Hyun Lee

School of Computer and Information Science, University of South Australia, Australia
abrar.haider@unisa.edu.au,
leesy116@mymail.unisa.edu.au

Abstract. Quality of information is critical to corporate decision support. However, ensuring quality of information is not a straight forward task, due to the intertwined nature of the information quality dimensions. Particularly, measuring the impact of each dimension on other dimensions has proved to be extremely difficult. This paper presents a case study of measuring information quality in a manufacturing organization. It applies six-sigma approach to the product perceptive of information. In doing so, the paper utilizes analytical hierarchy process to find the correlation between information quality dimensions. The paper then applies the quality function deployment model to determine critical to quality factors for managing information quality.

Keywords: Information quality, information quality dimensions, analytic hierarchy process, quality function deployment.

1 Introduction

With the increase in business automation, quality of information has become critical to business competitiveness and even survival. Information quality (IQ) measurement and management, therefore, are positioned quite high on the agenda of IT managers. At the same time, IQ related research activity has also gained significant momentum in the last decade. As a result, there are a plethora of IQ measurement, control, and management frameworks and models. However, despite the significant number, most of these IQ endeavors only work for specific application contexts and are quite subjective in nature [1]. Stvilia et al. [2] complement this assertion and point out that most of IQ frameworks are ad-hoc, less intuitive and incomplete and cannot produce robust and systematic measurement models. This is because obtaining accurate measurement and cost-effective assessments of IQ has been difficult because of the complexities of information systems [3]. Many research initiatives have attempted to resolve this by taking a product perspective of information, where information lifecycle is treated the same way as a product lifecycle. The product perspective of IQ was first proposed by Wang [4], through Total Data Quality Management (TDQM) that gives convincing answers to many questions in regard to the methods of handing information in organizations. In TDQM, information is treated as a product and information systems are regarded as an assembly line to produce information and

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 127–138, 2012.

information stakeholders as customers that use information for better decision making or organizational management. Despite the fact that TDQM methodology provides full range of IQ management, it does not give clear IQ assessment method due to the high level of abstraction to objectively assess IQ. For example, researches based on TDQM do not account for the correlating or impact of individual IQ dimensions of other dimensions. This paper takes a holistic view of IQ measurement, where subjective and objective aspects of IQ dimensions are correlated and collated. It is based on the IQ management framework proposed by Lee and Haider [5] proposed that measures, analyzes, sustains, and improves IQ by adopting a six-sigma method. This paper reports the findings of a case study in a Korean manufacturing organization, where the framework was applied. This paper explores the finding from the IQ assessment perceptive and extracts critical to quality factors by applying AHP (analytic hierarchy process) and QFD (quality function deployment).

2 Analytic Hierarchy Process

AHP is a multi-criteria correlation approach that illustrates a hierarchical representation of a system by assigning weights to groups of components [6]. It is a practical approach when the decision making process is complex, large, and dynamic because it splits the overall criterion into lesser importance to acquire a global view of a problem for a global decision [7]. Three main processes constitute AHP, i.e. hierarchy construction, priority analysis, and consistency verification [8]. Firstly, in order to build a hierarchy, decision makers are required to deconstruct complex multiple criteria decision problems into its every possible component parts. Secondly, for a pair-wise comparison, decision makers need to compare each components part in the same hierarchy level according to their own knowledge and experience. Finally, the consistency ratio is calculated to confirm the degree of consistency among the pair-wise comparison. Once these three processes are completed, each components part in the same hierarchy level is operated by comparing components to obtain their relative importance (or weights) throughout the hierarchy. This pair-wise comparison is a key step of the AHP. With the benefit of AHP and its intuitive deployment and flexibility, it has been applied in diverse industries such as manufacturing, finance, engineering, public sectors, management and more [8, 9, 10, 11]. In this paper, we conduct pair-wise comparison of IQ dimensions' relative importance, and then the weights of IQ dimensions (relative importance) are utilized to QFD to extract critical to quality factors.

3 House of Quality

House of Quality (HOQ) is the core of QFD [12]. As shown in figure 1, this is a matrix that composes to sub-matrices that are related to one another. Here, customer requirements matrix (block number 1) is replaced to information customer responses collected from the case study. Technical requirements matrix (block number 2) is represented in the hierarchy of IQ dimensions. Inter-relationships matrix

(block number 3) is calculated according to the results of AHP. Correlation of IQ dimensions matrix (block number 4) is filled by the results of IQ relationship of AHP and its directions of improvement (block number 5) are defined accordingly. Finally, in targets matrix (block number 6), objective or subjective measures are defined to each IQ dimension respectively.

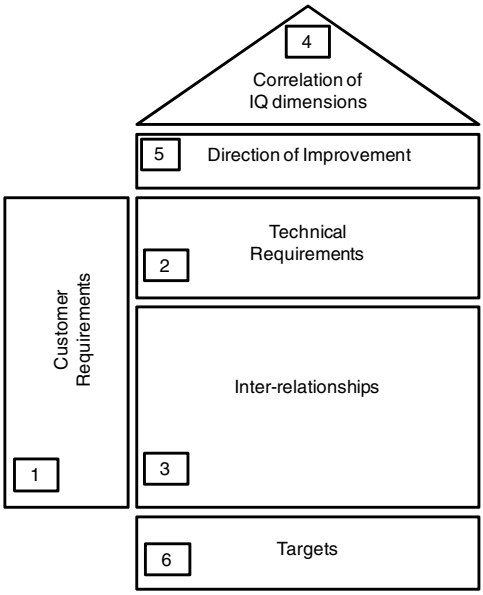


Fig. 1. Quality Function Deployment (House of Quality)

4 The Case Study

As per the University of South Australia ethics committee instructions, the case organization cannot be identified with its real name; therefore it will be referred to as company A. It has four distinct business domains in electronics manufacturing areas working in over 120 operations including 80 subsidiaries world-wide. Company A has centralized IT operations. In order to reveal relative importance of IQ dimensions and their mutual relationships, a survey was conducted at the purchase division of the company headquarters in Seoul in between June and July, 2011. The survey was distributed among 80 employees, and 68 valid responses were received. The details of the respondents' profiles are illustrated in the figure 2.

4.1 Identifying IQ Problems

Identifying IQ problems can be the most important activities as it is obviously starting point of quality improvement and linked with outcomes. Company A has been suffered from information which is dealt with different data because of non-standardized information management.

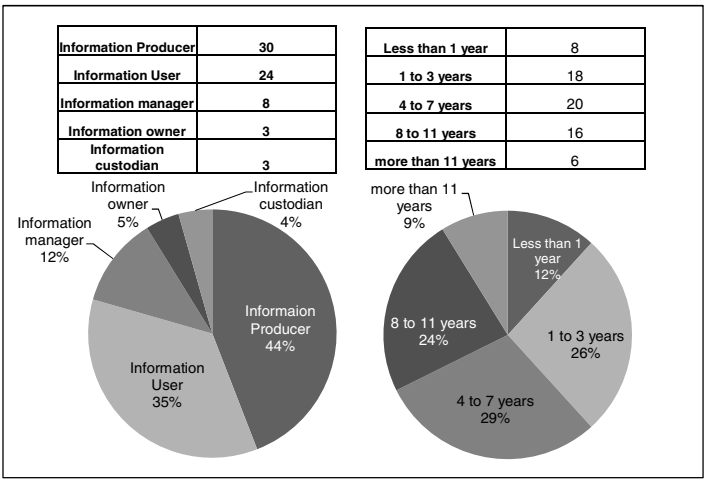


Fig. 2. Respondents' profiles

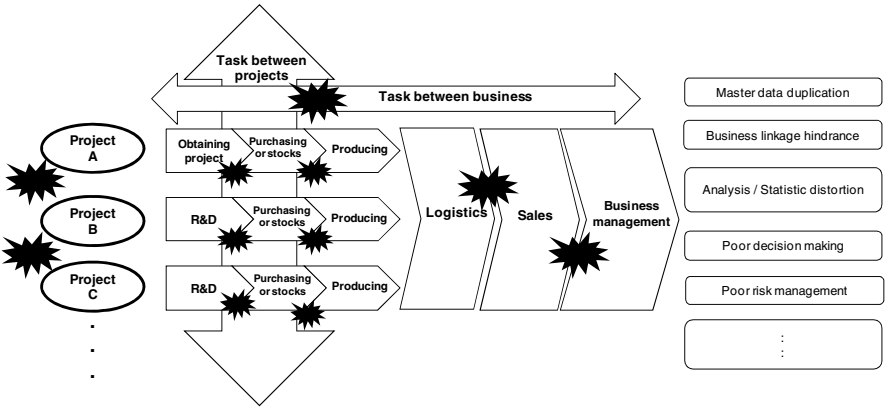


Fig. 3. Information related problems in company A

This means that although the subsidiaries and other offices around the globe have their own IT functions, yet the subsidiaries and other operating offices are required to update information regarding their operation to the company headquarters. It causes different management of information according to the difference between projects and business. Figure 3 shows information related problems in company A. As the information throughout project life-cycle is differently handled, purchasing associated divisions have to manage information in their own policies. Tasks between projects and tasks between businesses also handle information in different management process. Even though the company A recognizes that information should be shared throughout all divisions, each division and subsidiary operates own information management system and it affects line of business effectiveness and decision making. In addition, it costs more money because handling information system in different

way at one company causes duplicated investment for products components purchasing as well as wasting time. Through interview with focused groups, IQ related problems in company A were addressed as follows:

- Master data duplication: non-standardized information, inconsistency of same information
- Business linkage hindrance: Low efficiency of LOB (Line of Business), different operation of IS, lack of master data
- Analysis/statistic distortion: operating poor information, not solid information
- Poor decision making: utilizing biased information or not comprehensive information
- Poor risk management: incomplete management due to missing information

4.2 Analytic Hierarchy Process

The survey instrument consists of three sections i.e. general information regarding IQ initiatives, relative importance of IQ dimensions, and the relationship of IQ dimension with each other. In the relative importance of IQ dimensions section, questions are asked about IQ dimensions and their relative importance to other dimensions so that pair-wise comparison of IQ dimensions is carried out. Since there are many IQ dimensions, it is impossible to assess each IQ dimensions at the same time.

Table 1. Summary of AHP applied to Respondents' Responses

Categories			Quality Perspective		IQ Dimensions	Weight	Local order
Information Quality Dimensions for improvement	Conformance to Specifications	0.474	Information Product Quality	0.638	Free of Error	0.290	1
					Conciseness	0.284	2
					Completeness	0.215	3
					Consistency	0.211	4
			Information Service Quality	0.362	Timeliness	0.502	1
					Security	0.498	2
	Meets or Exceeds Customer's Expectations	0.526	Information Product Quality	0.314	Appropriate Amount	0.065	5
					Relevancy	0.240	2
					Ease of Understanding	0.167	3
					Interpretability	0.163	4
					Objectivity	0.365	1
			Information Service Quality	0.686	Believability	0.510	1
					Accessibility	0.102	3
					Ease of Operation	0.073	4
					Reputation	0.315	2

Therefore, we classified IQ dimensions into 'conformance to specification' and 'meets or exceeds customer's expectation' broadly. The hierarchy of IQ dimensions is based on PSP/IQ Model [13]. The six IQ dimensions in the 'conformance to specification' have been selected. These dimensions are 'free of error', 'conciseness', 'completeness', 'consistency', 'timeliness' and 'security'. However, in order to gain the relative importance of all IQ dimension in the IQ hierarchy and implement the results to QFD, AHP survey questions contains all IQ dimension and their relative

importance were computed. Table 1 shows the results of the AHP applied to the respondents and hierarchy of IQ dimensions. The survey results indicate the 'meets or exceeds customer's expectation' is slightly higher than the 'conformance to specifications'. It is, therefore, not unreasonable to postulate that being a manufacturing organization, company A, is aware of importance of both customers and product manufacturing based quality perspectives. This is because information users, custodians, and managers can easily see the impact of the quality, or lack of quality of information on their work. It, therefore, relates easily to 'conformance to specification' than 'information service quality' perspective, which is more abstract and relative in terms of task, user, or context. On the other hand, it is obvious that the weight of 'information service quality' is higher than the 'information product quality' in the 'meets or exceeds customer's expectations' category, because the 'information service quality' is more connected with subjective perspective.

Analysis of IQ Dimensions' Relationship

Since the survey was conducted in a manufacturing organization, only four IQ dimensions, i.e., 'free of error', 'conciseness', 'completeness', and 'consistency' were analyzed in this study. The most related IQ dimensions against each four dimension were determined to discover how much they affect each other. The 'free from error' dimension embraces the information attribute that are related to reflecting the real-world such as correctness, reliability, and certifying values.

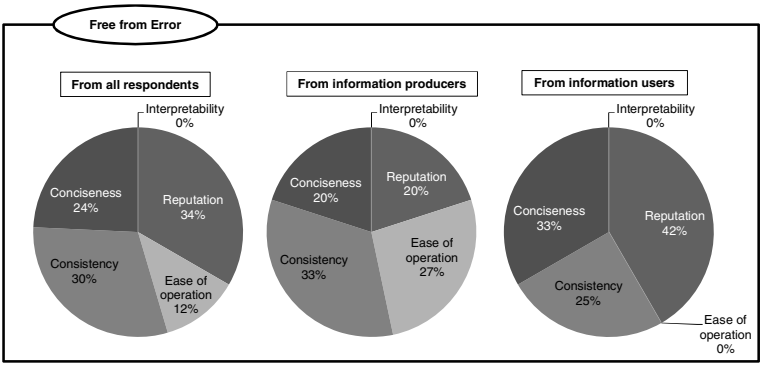


Fig. 4. Relationship of 'Free from Error' IQ Dimension

Therefore, the set of IQ dimensions, i.e. 'reputation', 'ease of operation', 'consistency', 'conciseness', and 'interpretability' were chosen as the list for respondents to rate the relative importance of 'free from error' against. Figure 4 shows the results of 'free from error' IQ dimension. From all respondents' point of view, 'reputation' is outlined as the most related dimension with 34% to 'free from error'. Dimension of 'reputation' represents the raw information sources that are used to create the information product. The quality of 'free from error', therefore, interacts strongly with the 'reputation' compared to the rest of dimensions. Interestingly, while

the 'ease of operation' is rated by 27% of information producers, no one in the information users choose this dimension. It appears likely to us that the information users prefer to use information from reliable sources for correctness rather than 'ease of operation'; given the fact that the 'reputation' dimension is rated by 42% of the respondents as having an important relationship with 'free from error' in the information users group.

The 'completeness' dimension represents the information attribute that is of sufficient breath, depth, and scope for the task at hand. Therefore, the set of IQ dimensions which are regarded as the most related dimensions, include, 'objectivity', 'ease of operation', 'relevancy', 'timeliness', and 'believability' with the 'completeness' were asked to the respondents.

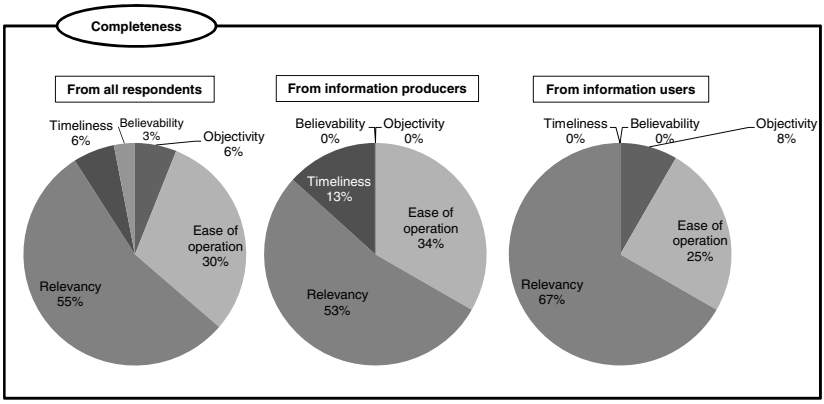


Fig. 5. Relationship of 'Completeness' IQ Dimension

Figure 5 shows the relationship of 'completeness' with other dimensions. From all respondents' point of view, 'relevancy' is outlined as the most related dimension with 55% response to 'completeness'. Given the attribute of 'relevancy', it should not surprise us that the dimension marks the highest relationship percentage. The attributes of 'relevancy' relate to the information that is applicable and helpful for the task at hand. As the dimension is mostly related to supporting a given process, helping decision making, adapting to other application and linkage, there is a striking connection between 'completeness' and 'relevancy'. This connection is supported here, since both information producers and users groups gave the highest percentage to the 'relevancy' dimension even though the information users rated it 13% higher than information producers.

The 'consistency' dimension embodies the information attribute that is always presented in the same format and is compatible with previous information. Therefore, the set of IQ dimensions related to this dimension include, 'free from error', 'timeliness', 'completeness', 'interpretability', and 'ease of operation' were asked to the respondents. Figure 6 shows the relationship of 'consistency' with these dimensions as deemed by the survey respondents. From all respondents' point of view, 'interpretability' has the highest score with 37%, whereas 'timeliness' is second

with 27%, followed by 'completeness' rated as 18%. A sharp contrast is drawn between the two groups toward the relationship with 'consistency'; whereas information producers indicate that 'timeliness' is most related with 'consistency' dimension, 'interpretability' was chosen by information users as the most related dimension for 'consistency'. The significant disparity seems to be explained by the difference task roles of the two groups. In general, information producers are supposed to generate information based on their task. As the 'timeliness' concerns how long the information remains valid it can be related to the compatibility with previous information when they generate new information.

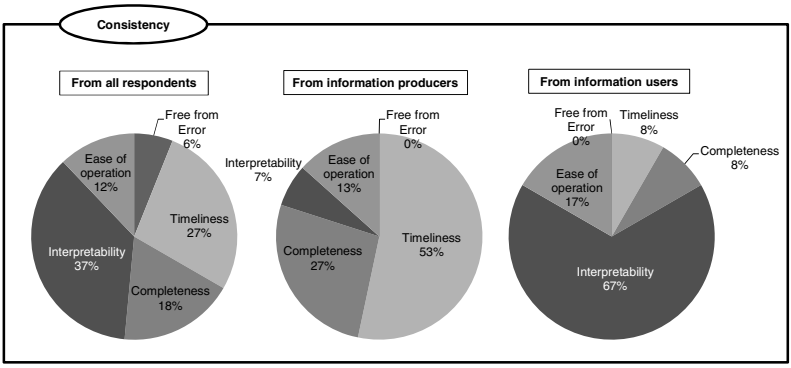


Fig. 6. Relationship of 'Consistency' IQ Dimension

It is highly reasonable to note that 'interpretability' was ranked as the highest with 67% for the relationship with the 'consistency' by the information users. 'Interpretability' describes the attributes of the information that is appropriate language and units and clear definition, it has strong connection with the factors of the 'consistency' such as same format or compatibility from the information users' point of view. The 'conciseness' dimension represents the information attributes that are compactly represented without being overwhelming. Therefore, the set of IQ dimensions related to this dimension including 'free from error', 'completeness', 'consistency', 'ease of understanding', and 'timeliness' were asked to the respondents. Figure 7 shows the relationship of 'conciseness' with other dimensions. From all respondents' point of view, the 'ease of understanding' dimension has the highest relationship with the 'conciseness'. Since 'ease of understanding' pursues the information that is to be clear without ambiguity and easily comprehended, it is highly reasonable that the 'conciseness' has the strongest relationship with the 'ease of understanding'. The reason why information users group ranked 'ease of understanding' higher than information producers group can be explained by the fact that information users, in general, amend, update, and view information and thus this dimension is helpful for their job.

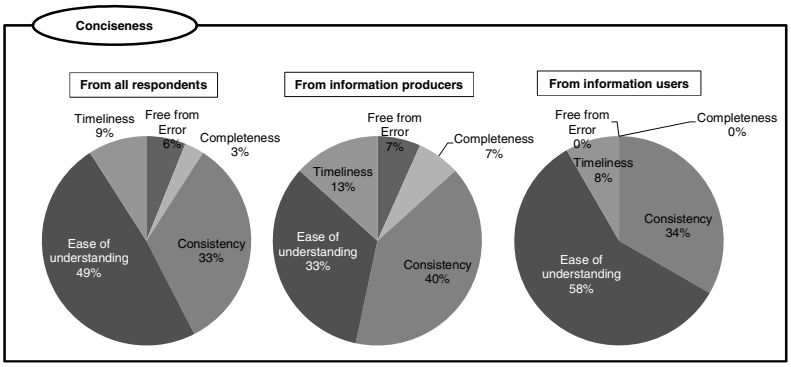
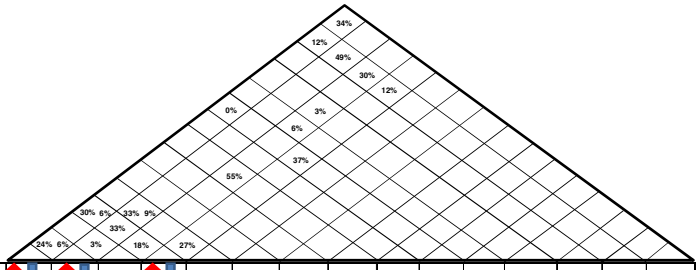


Fig. 7. Relationship of 'Conciseness' IQ Dimension

4.3 Quality Function Deployment

Based on HOQ methodology, QFD of company A is filled out to extract CTQ (critical to quality). Firstly, customer requirements (block number 1) are listed on the QFD matrix. And then the weights of importance for customer requirements are assigned with the scale 5 to 1 based on the frequency from interview and survey data. The most frequent customer requirement is assigned scale 5 and the lowest frequency one is assigned scale 1. In the technical requirements (block number 2), all IQ dimensions are place in according the IQ hierarchy defined the phase 2 of IQ framework. Accordingly, each weights of importance of IQ dimension (as calculated from AHP) are allocated. In the inter-relationships matrix (block number 3), the importance value between customer responses and IQ dimensions is calculated, e.g. "high dependency to often changed data policies" is related to "free of error", "conciseness", "completeness", and so on. In this case the weight of importance of customers requirement is multiplied by each related the weight of importance of IQ dimension. When the inter-relationship matrix is conducted, filling the matrix with stakeholders who are actually involved in the information flow is critically important in order to clearly know which customers requires are associated with certain IQ dimensions. Based on the results of relationships of IQ dimension, the relationship rate is filled in the correlation of IQ dimensions (block number 4) matrix respectively. For example, as the "completeness" dimension has 27% of relationship with "Timeliness" dimension, the relationship value is filled in the matrix between "completeness" and "Timeliness" in the correlation of IQ dimensions. Here the direction of improvement (block number 5) indicates of mutual inter-relationship. Finally, targets matrix (block number 4) contains assessments methods for each IQ dimension about objective (O) or subjective measurement (S) of IQ as well as each IQ specification. In addition, CTQs are determined based on the results of inter-relationship results. Here, 'believability', 'timeliness' and 'conciseness' are determined as CTQs of company A. Figure 8 shows the QFD results of company A.



Direction of improvement																
HOWS / WHATS	Weights of Importance	Conformance to Specification						Meets or Exceeds Customer's Expectations								
		Information Product Quality				Information Service Quality		Information Product Quality					Inforamtion Service Quality			
		Free of Error	Conciseness	Completeness	Consistence	Timeliness	Security	Appropriate Amount	Relevancy	Ease of Understand	Interpretability	Objectivity	Believability	Accessibility	Easy of Operation	Reputation
Weights of importance		0.290	0.284	0.215	0.211	0.502	0.498	0.065	0.240	0.167	0.136	0.365	0.510	0.102	0.073	0.315
High dependency to often changed data policies	2	0.580	0.568	0.430	0.422	1.004			0.480	0.334		0.730	1.020		0.146	0.630
Lack of license to assess IQ	1													0.102		
Difficulty of data trace	2				0.422				0.480						0.146	
Data distortion when integrating	3		0.852	0.645	0.633										0.219	
Long process time to update data	3		0.852			1.506									0.219	
My information is open to everyone	1						0.498									
Manual data handling always makes poor information.	4	1.160											2.040		0.292	
Poor definition of data cycle.	3		0.852			1.506			0.720					0.306		
Lack of data ownership	5	1.450							1.200		0.680		2.550		0.365	1.575
Lack of professional skill to IQ	3	0.870	0.852	0.645	0.633	1.506	1.494	0.195	0.720	0.501	0.408	1.095	1.530	0.306	0.219	0.945
Late delivery of information	3					1.506								0.306	0.219	
Various format of data	4	1.160	1.136	0.860	0.844			0.260	0.960	0.668				0.408	0.292	
Undefined tolerance of each IQ dimensions	3	0.870	0.852	0.645	0.633	1.506	1.494	0.195	0.720	0.501	0.408	1.095	1.530	0.306	0.219	0.945
Lack of understanding which data is more important for the task	2	0.580	0.568	0.430	0.422	1.004	0.996	0.130	0.480	0.334	0.272	0.730	1.020	0.204	0.146	0.630
Various way to enter data	3	0.870								0.501					0.219	
Too much data that are not often used in IS	3							0.195	0.720						0.219	0.945
Poor understanding of data structure	3		0.852							0.501	0.408	1.095		0.306	0.219	
Difficulty of Real-time analysis	2														0.146	
Un-synchronized data with business framework	4				0.844				0.960	0.668					0.292	
Too much transaction code	3		0.852	0.645	0.633										0.219	
poor definition of data entity	3		0.852	0.645	0.633	1.506	1.494		0.720	0.501		1.095	1.530	0.306	0.219	0.945
non-standardized data representation	5	1.450	1.420	1.075	1.055	2.510	2.490	0.325	1.200	0.835	0.680	1.825	2.550	0.510	0.365	1.575
Too much restriction to assess data	2						0.996							0.204	0.146	
Too much time to retrieve data from repository	3	0.870	0.852	0.645	0.633	1.506	1.494	0.195	0.720	0.501	0.408	1.095	1.530	0.306	0.219	0.945
Lack of data storage	1							0.065							0.073	
Assessment methodology		O	O,S	S	S	O,S	O	O,S	S	S	S	S	S	O,S	S	S
Six sigma specification																
Absolute importance		9.860	11.360	6.665	7.807	15.060	10.956	1.560	10.080	5.845	3.264	8.760	15.300	3.570	4.818	9.135
CTQ selection			3			2							1			

Fig. 8. QFD Results of Company A

4.4 Analysis of Critical to Quality Factors

The cause and effect diagram helps to identify the potential causes for a problem. This was applied to CTQs from the QFD results. Figure 9 illustrates the results of fishbone diagram for the timeliness dimension.

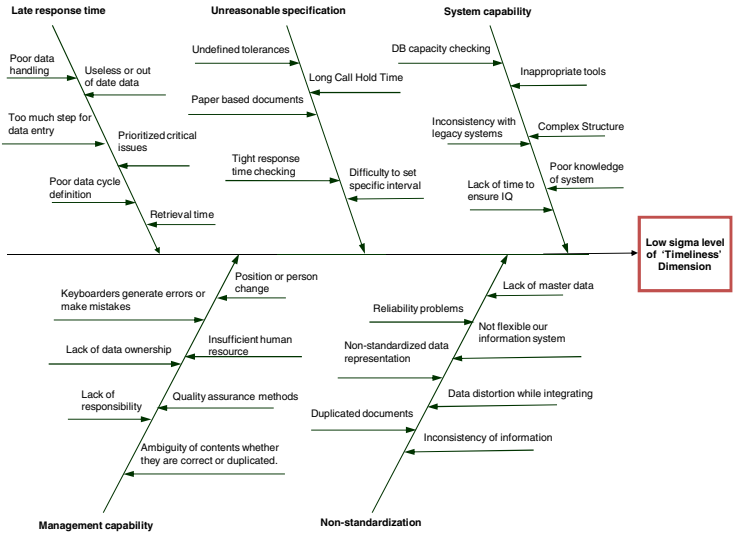


Fig. 9. Fishbone Diagram for 'Timeliness' Dimension

This fishbone diagram has been developed with the inputs from information managers at company A. These managers are actively involved in the flow of information and are thus well placed to comment on the issue at hand. From this step, the following issues have been listed as potential reason of CTQs.

- Late response time, i.e. prioritized critical issues, too much steps for data entry.
- Unreasonable specification, i.e. undefined tolerances.
- System capability, i.e. poor knowledge of system.
- Management capability, i.e. insufficient human resources, position or person change.
- Non-standardization, i.e. inconsistent information.

5 Conclusion and Future Work

This paper has attempted to sketch out the method of CTQs extraction using AHP-QFD model and has illustrated the process to ascertain the reasons to these CTQ. The paper demonstrates that by taking into account the product perspective of information, IQ dimensions can be converted to qualitative forms by establishing the mutual

relationship between IQ dimensions as relative importance using AHP method. With the relative importance, the paper also shows the extraction of CTQ factors utilizing QFD. The results of this study, thus, provide useful grounds on how subjective nature of IQ can be assessed numerically. There are, however, a number of problems that remain to be research further. Firstly, conversion of subjective qualitative factors into quantitative objective elements; secondly, there is no agreement of IQ dimensions' selection for extracting CTQs; and thirdly, the study about IQ dimensions' selection according to the different types of industries should be conducted. Additionally, as an IQ dimension does not work solely, negative or positive correlations of IQ dimensions should also be addressed so that potential reasons of CTQs are further clarified.

References

1. Knight, S., Burn, J.: Developing a framework for assessing information quality on the World Wide Web. Informing Science: International Journal of an Emerging Trans. Discipline 8, 159–172 (2005)
2. Stvilia, B., Gasser, L., Michael, B., Linda, C.: A framework for information quality assessment. Journal of the American Society for Information Science and Technology 58(12), 1720–1729 (2007)
3. Madnick, S.E., Wang, R.Y., Lee, Y.W., Zhu, H.: Overview and Framework for Data and Information Quality Research. Journal of Data and Information Quality (JDIQ) 1(1), 1–22 (2009)
4. Wang, R.Y.: A product perspective on total data quality management. Communications of the ACM 41(2), 58–65 (1998)
5. Lee, S.H., Haider, A.: Asset Lifecycle Information Quality Management - A Six-Sigma Approach. In: 6th World Congress on Engineering Asset Management, Cincinnati, OH, USA (2011)
6. Saaty, T.L.: How to make a decision: the analytic hierarchy process. European Journal of Operational Research 48(1), 9–26 (1990)
7. Saaty, T.L.: Time dependent decision-making; dynamic priorities in the AHP/ANP: Generalizing from points to functions and from real to complex variables. Mathematical and Computer Modelling 46(7-8), 860–891 (2007)
8. Ho, W.: Integrated analytic hierarchy process and its applications-A literature review. European Journal of Operational Research 186(1), 211–228 (2008)
9. Chen, M.K., Wang, S.C.: The critical factors of success for information service industry in developing international market: Using analytic hierarchy process (AHP) approach. Expert Systems with Applications 37(1), 694–704 (2010)
10. Yang, J., Shi, P.: Applying analytic hierarchy process in a firm's overall performance evaluation: a case study in China. International Journal of Business 7(1), 31–46 (2002)
11. Byun, D.H.: The AHP approach for selecting an automobile purchase model. Information & Management 38(5), 289–297 (2001)
12. Przekop, P.: Six Sigma for business excellence. McGraw-Hill, London (2006)
13. Kahn, B., Strong, D.M.: Product and Service Performance Model for Information Quality: An Updat. In: Proceeding of the 1998 Conference on Information Quality, Massachusetts Institute of Technology, Cambridge (1998)

Multi-dimensional Visualization in Enterprise Modeling

Frank Christiner, Birger Lantow, Kurt Sandkuhl, and Matthias Wißotzki

University of Rostock,
Chair of Business Information Systems,
Albert-Einstein-Str. 22, 18059 Rostock, Germany
win.office@uni-rostock.de
http://www.informatik.uni-rostock.de/index.php?id=2169

Abstract. Enterprise modeling (EM) is a discipline supporting business and IT alignment by providing means for capturing, visualizing and improving different perspectives of an enterprise, including processes,organization structures, products, systems, and business objectives. However, there is a lot of relevant information besides the one presented in enterprise models. Including such information into enterprise models or an integrated presentation of model and data view is supposed to ease decision making for stakeholders in organizations by providing contextual information for the decision at hand. Additional information however usually means additional complexity. This paper explores possibilities of an integrated presentation guided by the following questions: (1) What kind of complementary information should be visualized in an enterprise model? (2) How can the information be visualized? (3) How can the content of a specific visualization be adapted by the business stakeholder using it?

Possibilities and benefits of enhancing the existing enterprise models with visualization of additional information are illustrated using a small case study.

Keywords: Enterprise Architecture (EA), Enterprise Architecture Management(EAM), Enterprise Modeling, Business Alignment, Multidimensional Visualization, EKD, EKM.

1 Introduction

In highly competitive markets with globalized sourcing structures, enterprises' competitiveness depends to a large extent on the ability to quickly adapt to changing market and regulatory environments. This includes business and organizational functions as well as the information systems and IT-infrastructures supporting the business, i.e. systematic and continuous alignment of business and IT is crucial. Enterprise modeling (EM) is a discipline supporting business and IT alignment by providing means for capturing, visualizing and improving different perspectives of an enterprise, including processes,organization structures, products, systems, and business objectives.

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 139–152, 2012.

Many enterprise modeling methods and notations capture the complexity of businesses by using different sub-models for the above mentioned perspectives and by including the mutual relationships between these sub-models. Examples for such methods are EKD[1] and MEAF[2]. Such enterprise models allow to identify dependencies between perspectives and to determine what effects changes in one of the sub-models would have on the other sub-models. This feature supports business and IT alignment in general, but it should ideally be complemented by integration of key performance indicators from back-office systems or real-time information from operations. For a business process this might be the throughput time, number of exceptions or consumed capacity. For the relationship between a resource and a business process it might be the utilization or availability.

Including such information into enterprise models or an integrated presentation of model and data view is supposed to ease decision making for stakeholders in organizations by providing contextual information for the decision at hand. This reduces time required for information retrieval and it increases control on possible side effects or on target achievement as well. Some of the existing tools for enterprise modeling address this need for additional information. However, from our point of view the full potential of visualization techniques has not yet been tapped. Furthermore, little effort is spent on the selection of provided additional information in enterprise modeling approaches.

Adding information to model visualizations as described means visualizing the real world complexity of enterprises. In this context we seek for possible answers to the following questions:

1. What kind of complementary information should be visualized in an enterprise model?
   Adding complementary information not only enriches models but also adds complexity. Thus, options for limiting the overall complexity should be considered before visualization. Reduction of complexity is often an important step before its visualization. The paper investigates existing approaches and best practices. The results of complexity reduction define possible inputs for visualization tools.
2. How can the information be visualized?
   Adding information to a visualization means increasing its dimensionality. Hence, we have a look into approaches to visualize multidimensional data in different areas of science. Furthermore, the paper includes a short view into existing tools for enterprise modeling and their visualization capabilities.
3. How can the content of a specific visualization be adapted by the business stakeholder using it?
   Approaches to multidimensional visualization also address the problem of setting the focus on relevant information. This can be done by interactive navigation through and selection of the visualization content. We will have a short exploratory look into existing operations in this field.

[1] Enterprise Knowledge Development, see: section 4.
[2] Metis Enterprise Architecture Framework, see: `www.troux.com`

Based on the elaboration of question one in section 2 and of the remaining two questions in section 3, we are going to motivate the integration of indicators from operative systems in enterprise models. As an example for illustrating our approach, we selected the goal model of the EKD method. The EKD approach was chosen because of its free availability and existing experiences in enterprise modeling with this approach. This is taken as an example for the addition of new visualization possibilities to existing enterprise modeling approaches in section 4. A case study in section 5 gives an example of a possible implementation of the suggested visualization enhancements and their use.

2 Complexity and Complexity Reduction

This section briefly introduces the term complexity and general possibilities to complexity reduction.

2.1 What Is Complexity?

In general usage, complexity tends to be used to characterize something with many parts in intricate arrangement. In science, there are a number of approaches to characterize complexity. A general, complexity can be defined as follows:

That property of a language expression which makes it difficult to formulate its overall behavior even when given almost complete information about its atomic components and their inter-relations.[1]

There are different views on the described difficulty of behavior formulation. In software technology it might be a matter of processing time. From a managerial perspective it is a matter of costs or of human cognition capabilities. The latter one is most important in our context, since we have a focus on the information needs of business stakeholders and tools for enterprise modeling.

2.2 Complexity Reduction

Models are complexity reduced representations of real world phenomena. However, there are complementary techniques and practices to modeling that aim to give a proper way of complexity reduction within models. Improper complexity reduction in models like oversimplification may cause wrong conclusions based on the model. We give a short explorative view into complexity reduction approaches in the domain of management and enterprise modeling.

Principles of Proper modeling (PoPM by Becker (see [2]) form a preemptive approach. The PoPM define a validated methodical framework which supports the creation of information models in terms of clarity, quality, and consistency assurance. The model creation is supported by PoPM with the effect that the model is given comparability and that multi-perspectivity is facilitated [2, p. 3-7]. PoPM is a good example for the approach to aim at complexity reduction by defining the modeling process and its outcomes. Each enterprise modeling method handles complexity at least this way.

Simulation allows a demonstrative or an experimental access to complex issues. Simulation models provide positive contribution to complexity reduction in the analysis of enterprise models. Business process simulation (BPS) as an example is the formation and (experimental) running of models to analyze business processes. Here, processes are considered as chronological system state histories [3, p. 1-2].

Modularization is a general approach in software architecture. Applying principles of modularization to enterprise models means to identify partial models with a high level of interdependency within the partial models but not between them. This does not refer to classical submodels of enterprise modeling which represent different views/aspects of an enterprise as a whole but rather to functional domains or single processes of an enterprise. Roger Sessions promotes an practical approach to this topic. His SIP(Simple Iterative Partition)-method helps to identify possible partial models. Complexity is reduced because the state space of the partial models can be investigated almost independently (see [4]). The idea of picking parts out of the whole model is also represented in some of the visualization approaches shown in section 3.2.

Mathematical approaches exist that allow to identify and emphasize on relevant information. Examples are scale decomposition, sensitivity analysis, and statistics in general (see [5]). Taking sensitivity analysis, it could be used to identify relevant interdependencies.

Indicators can be seen as classification numbers which provide leading information on performance. Their change allows conclusions on behaviour of an unobservable or difficult to predict system [6, p. 402].

A large variety of indicators has been developed in managerial science. Therefore a context based selection of used indicators is necessary. Besides a one dimensional view on indicators like Return-on-Investment, Residual Income, or Economic Value added there are multidimensional views described by indicator systems like Tableau de Board, Performance Pyramid and Balanced Scorecard [6, p. 406].

For example, the main idea of the Balanced Scorecard approach is to define a few perspectives which reflect main goals of an enterprise. Each of the perspectives is represented by one or two indicators. The aim of the Balanced Scorecard method is a "balance" between the perspectives. In opposition to traditional approaches focusing on the financial perspective only, a more complete picture of an enterprise should be provided by including and presenting various perspectives with their most relevant characteristics. The perspectives originally suggested by Kaplan and Norton (see for example [7]) can be found along with our case study in section 5.

3 Multidimensional Visualization

In order to explore the possibilities of multidimensional visualization we analysed visualization techniques from different areas, like Geography [8], Biology [9], Civil Engineering[10], Computer Science [11], and Electrical Engineering[12]. Additionally, several EM tools namely planningIT®, Troux Architect®, and Enterprise Architect® have been investigated.

As described in section 1, two aspects have to be considered. First is, how different values and dimensions can be expressed. Second is, which manipulations to a visualization are possible in order to focus on relevant information.

3.1 Degrees of Freedom to Express Values in Visualizations

Looking into the different visualization approaches revealed a set of attributes to model artifacts commonly used. Table 1 gives an overview. Taking indicators as described in section 2.2, each indicator can show by its position which artifact it applies to, by its shape the semantics it has, by its color a qualitative risk information, and by text its actual quantitative value. Other visual representations of indicators are possible. Figure 1 illustrates how a large quantity of indicators can be visualized within a given hierarchy. Figure 3 illustrates how state of the art enterprise modeling tools make use of the visualization possibilities by using

Table 1. Overview of graphical attributes

Attribute	Comment
Position qualitative, quantitative	The position within a coordinate system can quantify a certain value. An example would be the arrangement of assets according to their risks in portfolio management. The relative position of an element to other elements can describe relationships/interdependencies between the elements. An example would be the use of swim lanes to show resource assignments.
Size qualitative, quantitative	Size can express quantitative values if a scale is provided. Different sizes may also define the membership in different classes which would be qualitative information.
Color qualitative, quantitative	Like size, color can express quantitative values based on a colormap. Qualitative information like class membership can also be expressed by color.
Shape qualitative	Different shapes can represent different qualities. As described for color and size, class membership can be expressed by shape.
Text qualitative, quantitative	Text is the most versatile graphical element to give information. It can carry qualitative and quantitative Information. However, it needs the most cognitive resources to perceive the given information.
Symbols qualitative	Adding symbols with a defined meaning can give qualitative information. For example, an exclamation mark may be used to identify conflicts.

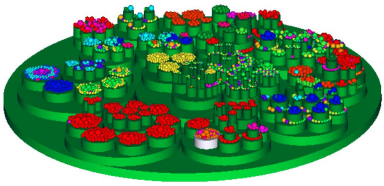


Fig. 1. 3D Nested Cylinders and Spheres (from:[13])

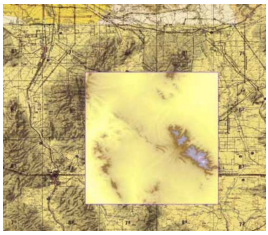


Fig. 2. Semantic lens (from:[14])

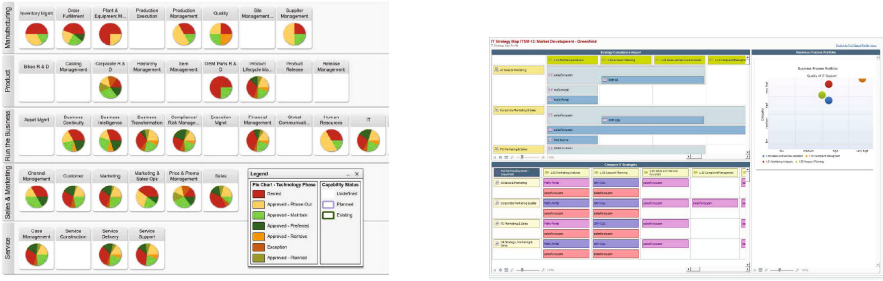


Fig. 3. Left: Capability Risk Dashboard Troux® Right: Scenario Planning dashboard planningIT®

leading tool vendors in the market as example. Both examples include indicator value visualizations.

3.2 Manipulation of Visualizations

The so called OLAP (Online analytical processing) operations are a well known way to describe possible manipulations on a multidimensional data structure. Here, a cube is the base for the multidimensional data model.

The basic manipulations are (see for example [11]):

- *Roll up/Drill down* to change the scale within a dimension.
- *Slice/Dice* to select pieces out of the cube.
- *Rotate* to get another perspective of the cube.
- *Drill across* to change to another cube.

Enterprise Architect® defines enterprise models within such a cube (see figure 4). In theory all of the described manipulations should be possible. However, the proof will be future work.

Besides the OLAP example there exist other manipulations to visualization that allow focusing or emphasizing on relevant information:

- *Filtering/Layering:* General approach of including/excluding elements by certain criteria from visualization. This can be done either globally or locally by a lense. See figure 2 for a semantic lens as used in tools for the visualization of geo data.
- *Assignment:* Possibility to move elements close to another in order to allow a comparison.
- *Colormap manipulation:* Changing colormap shifts emphasis on different classes.
- *Feature identifaction:* Automated highlighting either of artifacts sharing similar qualities or of artifacts showing qualities different from the majority.
- *Context Displays:* Include context information in partial models in order to show dependencies to other partial models.
- *Sequencing:* Cycling through different model visualizations along the values of one dimension, e.g. time. This allows to emphasize on developments. It is close to the simulation approach of section 2.2.

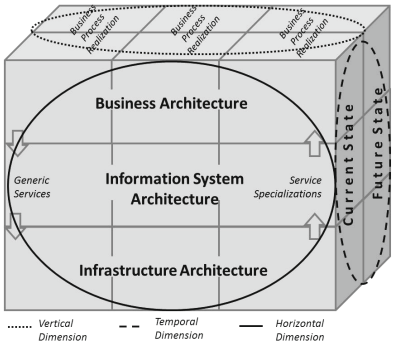


Fig. 4. The 3 Dimensions of Enterprise Architect (see: [15])

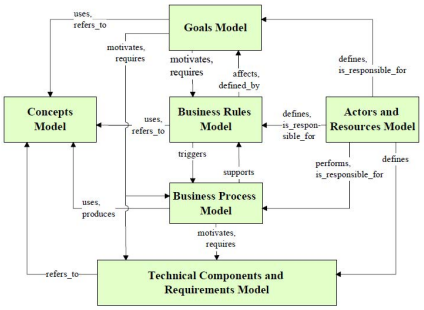


Fig. 5. Model types of the EKD framework (from:[16])

3.3 The Gap

The investigation of selected enterprise modeling methods and tools revealed that there is a gap between the possibilities of visualization and current implementations. Navigation and positioning in three or more dimensions is not supported by the existing methods. Freedom to add information and select information by context is very limited. The power of sequences is not used. And at last, the principle of context displays to give better orientation on navigation in models should be considered to a greater extent.

Table 2. Overview of the EKD sub-models (from:[17, p. 2])

	Goals Model (GM)	**Business Rules Model (BRM)**	**Concepts Model (CM)**
Focus	Vision and strategy	Policies and rules	Business ontology
Issues	What does the organization want to achieve or to avoid and why?	What are the business rules, how do they support organization's goals?	What are the things and *phenomena* addressed in other sub-models?
	Business Process Model (BPM)	**Actors and Resources Model (ARM)**	**Technical Component & Requirements Model (TCRM)**
Focus	Business operations	Organizational structure	Information system needs
Issues	What are the business processes? How do they handle information and material?	Who are responsible for goals and process? How are the actors interrelated?	What are the business requirements to the IS? How are they related to other models?

4 Integrating Indicators in EKD Goal Models

The EKD method defines a number of interrelated sub-models. Each of them represents some aspect of the enterprise (see table 2). The relationships between Goal Model (GM) and other sub-models are shown in figure 5).

There are two main reasons why including indicators to the GM provides benefits to the EKD method:

1. One task of the EKD method is a review and evaluation of goals. Having additional information about the goals available provides a significant support here.
2. While the EKD method covers the process of enterprise modeling it should be embedded in a process of Enterprise Architecture Management (EAM). EAM includes controlling as a task. Here, indicators are a common tool. Thus, considering indicators in EKD is a prerequisite for a smooth integration in EAM processes.

We decided to base our choice of indicators on the Balanced Scorecard approach by Kaplan [18, see]. This has several reasons:

1. The perspectives defined within a Balanced Scorecard reflect general business goals. Thus, adding indicators for these perspectives also adds information about dependencies to these general goals. Compared to a direct modeling of these goals in GM this implicit way saves visual elements and decreases complexity.
2. Especially strategic Balanced Scorecards as originally introduced by Kaplan proofed to be able to condense information about complex systems to a few key indicators.

3. The Balanced Scorecard concept is well known.
4. There is an inherent flexibility in the approach. The selection of perspectives is based on the stakeholders' priorities.
5. If the approach is already implemented in an enterprise, indicator measurement is available and the semantics of used indicators are clear.
6. The Scorecard perspectives are suitable to express existing dependencies to other EKD sub models. The Business Process Perspective can characterize dependencies to the elements within the EKD Business Process Model (BPM).

4.1 Selection of a Multidimensional Visualization Method

Balanced Scorecards are based on a hierarchy of indicators where the top level key indicators with the highest aggregation are the first information provided.

For level 1 visualization (highest aggregation) we suggest bar graphs. As we will see later in the case study, indicator values on this level need to be compared between different goals. This is supported by bar graphs which are assigned to each of the goals and use similar scales.

Drilling down the indicator hierarchy to indicators of lower aggregation requires the visualization of the underlying hierarchy. A good visualization for hierarchical information structures are Nested Circles, Cylinders (and Spheres) Diagrams. They can display large hierarchies in a comprehensible form, that means they use the available visualization space effectively. In contrast to tree visualizations which are widely used to visualize hierarchies, the focus lies on quantities. This supports the use of indicators which are quantifiers in general.

The selected visualizations combine the different graphical attributes (see table 1) as follows. Bar Diagrams (see figure 8) express their semantics by position and color. Quantities are expressed by scale based size along one dimension. Thus, absolute and relative quantities can be visualized here. Nested Circles Cylinders and Spheres express hierarchy levels by nesting child-elements within parent elements (relative position) and quantities by different sizes (see figures 1 and 11). Relative quantities can be visualized here. Generally, there are more visualization techniques available that fit the requirements. Also, several techniques can be combined as we see later.

5 Case Study Toys4Joy(T4J)

In order to illustrate our approach we use a case study from teaching enterprise modeling on bachelor level in a university course of study program. The virtual case study company T4J is active in the northern part of Germany and specialized in selling games for teenagers and young adults. The company consists of four stores and a central warehouse.

An enterprise modeling project was initiated, because a number of changes were needed and some general problems arose. The scorecard driven monitoring showed critical indicator values for finance and customer perspective. In finance

perspective costs went up. In customer perspective, the market share dropped due to decreasing sales. A reason might be the lack of an online-shop. To compete with competitors new e-business solutions should be given priority.

5.1 Goals Model Analysis

Since the aim is not to provide a comprehensive introduction into EKD GM, we concentrate only on "Supports"–relationships (see [16] for reference). In EKD they are used for goal decomposition. Hence, a hierarchy of goals is the result (see figure 6). The main goal of T4J can be found on level 1. Level 2 shows the goals which support the main goal and so on. Moving downwards in the hierarchy the goals become more tangible and may even represent tasks to be performed. The formulated goals reflect the needs resulting from the initially described situation of T4J.

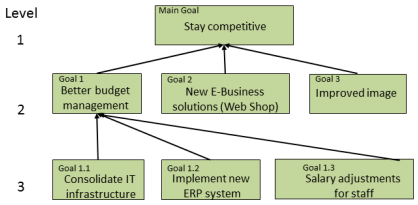


Fig. 6. Goals of T4J

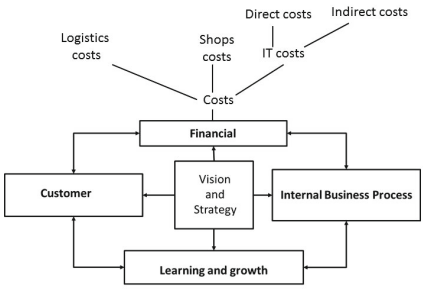


Fig. 7. Balanced Scorecard for T4J

Figure 7 shows the perspectives of the previously described Balanced Scorecard. The costs are distributed on cost centers and categories in order to have an example for further discussion. However, all perspectives are represented by indicators that can be decomposed. Evaluation of formulated goals should consider their impact on the scorecard perspectives. Thus, expected changes to the scorecards indicator values that are connected with reaching a certain goal are valuable information. In general, it is possible to use different indicators for different goal levels. However, comparability within levels should be assured by using the same indicators within a level. It is also possible that certain indicator values or value changes are not available or applicable to some goals. Negative impacts are omitted in this example but are generally possible.

The visualization of the enriched GM is shown in figure 8. Bar graphs show the expected changes to indicator values with respect to the scorecard perspectives where applicable. The main goal "Stay competitive" is considered to be congruent with the scorecard. Thus, it can be omitted for a reduction of visual complexity. Goal 1.3 for example has only financial impact. Goal 1.1 is red in

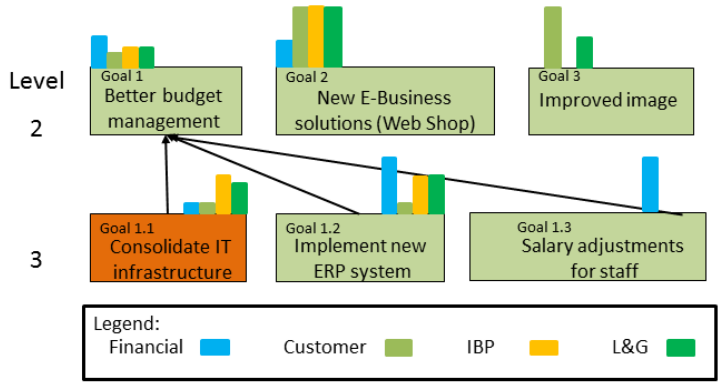


Fig. 8. Goals Model augmented with indicators

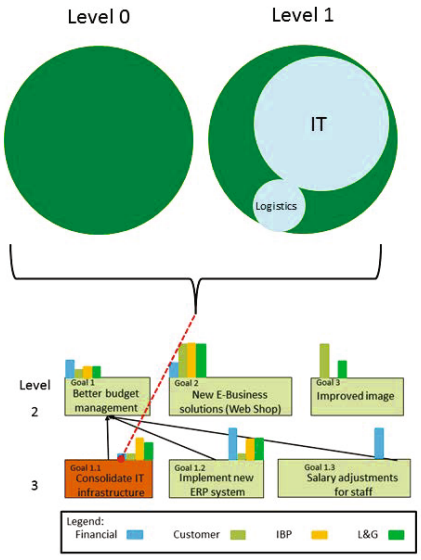


Fig. 9. Refined cost indicators

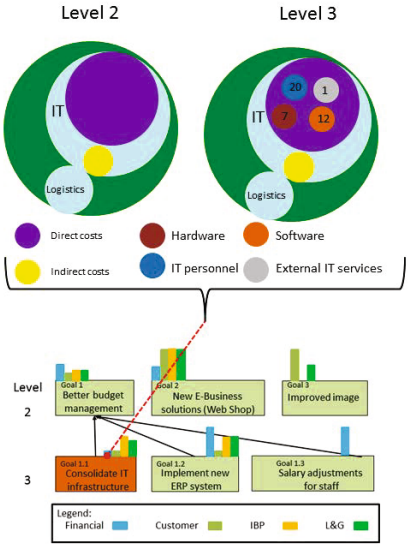


Fig. 10. Further refined cost indicators

order to set the focus on it. This is the result of the application of one of the manipulations like filtering, feature identification, or colormap manipulation as described in section 3.2. A look on the indicator bars reveals that the effect on finance is comparably low in contrast to the other sub-goals of Goal 1. The reasons may be revealed by drilling down (see section 3.2) the financial perspective.

The visualization is based on nested circles. Figures 9 and 10 illustrate the process of constructing such a visualization by adding lower hierarchy levels step-by-step. The depth of cost indicator decomposition may be chosen freely.

In this case, the actual level of visualization would be shown. There are already some findings for the work on the GM on level 1. The biggest effect of goal 1.1 "Consolidate IT infrastructure" is on IT costs. However, the shops do not benefit in terms of costs (figure 9). Going further down the levels, it is also visible that there is little effect on indirect IT costs (figure 10, level 2) and on External IT services costs (figure 10, level 3).

Several measures can be taken based on the given information. A further investigation on the reasons for the lack of effect may be initiated. Time is saved because the questions to answer in the investigations can be defined very precisely already at this stage. Another possibility would be the assignment of a lower priority to goal 1.1, knowing that there is little chance to increase the positive cost effects.

Figure 11 shows the highest possible level of detail in the given case study. The given information matches level 3 in the indicator decomposition. However, the third dimension (cylinders) is used to enhance visualization quality. The different indicator values on level 3 can now be compared visually. Generally, it is possible to visualize a three level hierarchy of indicators while the original context of the GM is still kept within the visualization.

The difficulties of investigating just one of the perspectives for a goal without the suggested visualization are illustrated in figure 12. The original GM on top gives no clue of the general benefits provided by the goals shown. The bottom Excel®-Sheet contains values by goal for the various cost indicators. It is hardly possible to include or to consider the GM context when navigating in these data.

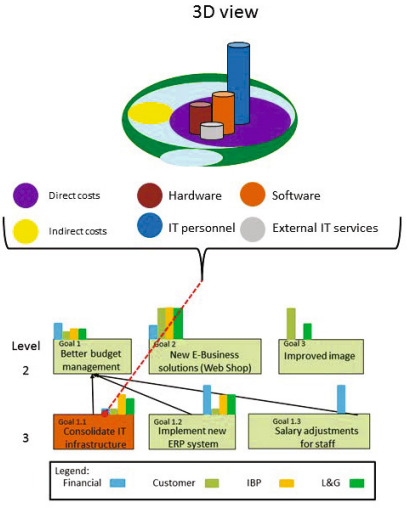


Fig. 11. Final enriched Goals Model

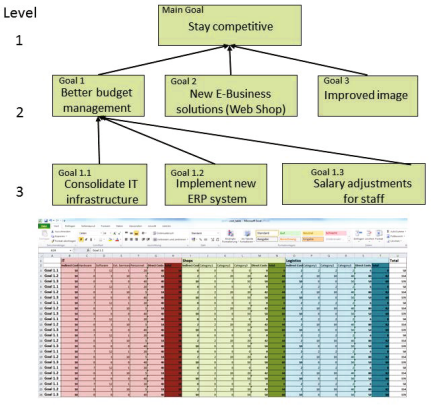


Fig. 12. original Goals Model and Excel-Sheet containing cost indicator information

6 Summary

We started our investigation by posing three questions: (1) What kind of complementary information should be visualized in an enterprise model? (2) How can the information be visualized? (3) How can the content of a specific visualization be adapted by the business stakeholder using it?

With respect to the first question we see a value in integrating operative data supporting decision making into enterprise models. We briefly investigated different approaches for complexity reduction to facilitate such an enrichment of enterprise models without creating an information overload on the user's side. Considering questions two and three, multidimensional visualization techniques turned out to be promising approaches. In this context, we identified a gap between existing enterprise modeling tools and methods on the one hand side and what might be desirable functionality in such tools on the other hand side (see section 3.3). The benefits and possibilities of complementing enterprise models with additional information were illustrated by a case study based example. Due to the explorative character of this study a need to extend and detail the findings by further research and a more systematic approach is evident. Future research on this topic is expected to lead to interesting options how to enrich or improve existing enterprise modeling tools.

References

1. Edmonds, B.: What is complexity? - the philosophy of complexity per se with application to some examples in evolution. Technical report (1999)
2. Becker, J.: Die grundsätze ordnungsmäßiger modellierung und ihre einbettung in ein vorgehensmodell zur erstellung betrieblicher informationsmodelle (1998) (last access: February 02, 2012)
3. van Uthmann, C.: Grundsätze ordnungsmäßiger modellierung (gom) für die geschäftsprozeß simulation (1999) (last access: February 02, 2012)
4. Sessions, R.: A better path to enterprise architecture (2006) (last access: February 07, 2012)
5. Surovtsova, I., Sahle, S., Pahle, J., Kummer, U.: Approaches to complexity reduction in a systems biology research environment (sycamore). In: Winter Simulation Conference, pp. 1683–1689 (2006)
6. Friedl, B.: Controlling. Lucius und Lucius, Stuttgart (2003)
7. Kaplan, R.S., Norton, D.P.: The balanced scorecard: Translating strategy into action. Advanced reader's copy. edn. Harvard Business School Press, Boston and Mass (1996)
8. MacEachren, A.M., Ganter, J.H.: A pattern identification approach to cartographic visualization. Cartographica (1990)
9. Boyle, J., Horch, H., Scharf, M.: Visualising large data sets in molecular biology. In: IDS, p. 8 (1996)
10. Farin, G., Hoschek, J., Kim, M.-S.: Handbook of computer aided geometrical design (2002)
11. Clausen, N.: OLAP: Multidimensionale Datenbanken; Produkte, Markt, Funktionsweise und Implementierung. Addison-Wesley, Bonn [u.a.] (1998)

12. Soonee, S.K., Kumar, D., Saxena, S.C., Kumar, S.: Visualization and human factors in electric power system operation. In: ICPSODR 2006 (2006)
13. Wang, W., Wang, H., Dai, G., Wang, H.: Visualization of large hierarchical data by circle packing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2006, pp. 517–520. ACM, New York (2006)
14. Erika Darling, C.N.: Reducing visual clutter with semantic lenses. In: ESRI User Conference (2005)
15. Merriman, T.(OAD Consulting, I.A.): A multi-dimensional approach to enterprise architecture (2011) (last access: November 14, 2011)
16. Bubenko Jr., J., Anne Persson, J.S.: D3: Appendix B EKD User Guide (2001)
17. Janis Stirna, A.P.: Ten years plus with ekd: Reflections from using an enterprise modeling method in practice
18. Robert Kaplan, D.N.: The balanced scorecard (1992) (last access: January 19, 2012)

Ontology Alignment for IT Integration in Business Domains

Alexander Smirnov, Tatiana Levashova, and Nikolay Shilov

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 39, 14 Line, 199178 St. Petersburg, Russia
{smir,tatiana.levashova,nick}@iias.spb.su

Abstract. Development of ICT offers new solutions facilitating appearance of flexible supply networks. Business and IT alignment in flexible supply networks requires integration of different IT systems, what, in turn, assumes semantic integration of their data and workflow models. Ontologies have shown their usability for this type of tasks. Integration of different enterprise aspects into an ontology has been also researched in a number of works. In the proposed approach an ontological model is used to solve the problem of heterogeneity. This model makes it possible to enable interoperability between heterogeneous information systems due to provision of their common semantics. The paper focuses on description of the ontology-based knowledge fusion and ontology alignment patterns aimed at facilitation of the IT integration processes.

Keywords: Pattern, knowledge fusion, ontology alignment.

1 Introduction

New information technologies open new boundaries for researchers. One of the results of implementation of new knowledge-based ICT (Information & Communication Technologies) applications is evolving of flexible manufacturing systems. Above regard a growing number of research directions have been developed in this area. One of such directions is flexible supply networks assuming complex non-linear relationships between numerous independent participants. Development of ICT offers new solutions facilitating appearance of flexible supply networks.

The integrated European Commission sponsored FP6 project “ILIPT - Intelligent Logistics for Innovative Product Technologies” [1] within the “EU 5-Day Car Initiative” was aiming at transforming the European automotive industry from the "stock push" thinking, to a stockless "build-to-order" (BTO) production strategy. Some of ILIPT results showed that one of the key issues in mass customization is establishing ad hoc agreements between supply network members for assembling and delivering personalised products on the Just-in-Time (JIT) or Just-before-Time (JBT) basis, thus shifting the mass customization idea to the new level of the postponement of the final product assembly. Participation of the authors of the paper in ILIPT gave an inspiration for further research, some results of which are presented below.

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 153–164, 2012.

Business and IT alignment in a flexible supply network requires integration of different IT systems, what, in turn, assumes semantic integration of their data and workflow models. Developing frameworks, with appurtenant models, needs to be based on solid foundations. The alignment problem requires a common ontology capturing business as well as IT [2].

Ontologies have shown their usability for this type of tasks [e.g., 3-5]. Integration of different enterprise aspects into an ontology has been also researched in a number of works. For example, socio-instrumental pragmatism [6] incorporates human, organizational, and IS/IT enabled actions in a coherent ontology. The concern of theorizing actions has also been acknowledged by actor-network theory [7], where technology and people are both seen as social actants.

In the proposed approach an ontological model is also used to solve the problem of heterogeneity. This model makes it possible to enable interoperability between heterogeneous information systems due to provision of their common semantics [8].

In order to analyze the existing ontology matching techniques an extensive state-of-the-art review has been done, which covered about 20 systems/approaches/projects related to ontology matching. Among them the following ones are worth to be mentioned: GLUE System [9, 10], Falcon-AO [11], MLMA [12], Hovy [13], SKAT [14], ONION [15], Promt [16], H-Match [17], CTX-MATCH [18], SMART [19], Cupid [20], COMA [21], Similarity Flooding Algorithm [22], AgreementMaker [23], Pattern Based Approach [24], MinSMatch [25], OntoView [26], Chimaera [20].

Patterns are a proven way to capture experts' knowledge in fields where there are no simple "one size fits all" answers [27], such as knowledge fusion or ontology alignment. Each pattern poses a specific design problem, discusses the considerations surrounding the problem, and presents an elegant solution that balances the various forces or drivers. In most cases, the solution is not the first approach that comes to mind, but one that has evolved through actual use over time. As a result, each pattern incorporates the experience base that senior integration developers and architects have gained by repeatedly building solutions and learning from their mistakes. This implies that we did not "invent" the patterns; patterns are not invented, but rather discovered and observed from actual practice in the field [27].

The paper is structured as follows. This section introduces the problem of ontology alignment in the business and IT alignment domain as well as the proposed approach. Then, in the next section, the patterns of ontology-based knowledge fusion are described. Section 3 shortly describes the knowledge representation formalism. Section 4 describes the ontology alignment patterns. An alignment example is given in sec. 5. The main results are summarized in the conclusion.

2 Ontology-Based Knowledge Fusion Patterns

In the proposed approach the process of knowledge fusion for IT integration is supported by usage of the ontologies. Carried out analysis of the tasks involving knowledge fusion has allowed to identify the list of generic patterns.

Ontology-based knowledge fusion patterns will be illustrated via the following example. There are two ontologies (A and B) with some structures (Fig. 1). There is a tacit mapping between two primary knowledge units, namely a3 from A and b2 from B. It is necessary to integrate these two ontologies preserving their internal knowledge structures and revealing the tacit mapping mentioned above.

Selective Fusion (Fig. 2) is used for common ontology creation. A new ontology is created, which contains required parts of the initial ontologies. The initial ontologies preserve their internal structures and autonomy.

Simple Fusion (Fig. 3) is used for ontology library creation and maintenance. A new "super ontology" is created, which contains initial ontologies. The initial ontologies preserve their internal structures but lose (partially or completely) their autonomy.

Extension (Fig. 4) is used during knowledge map and internal knowledge base maintenance. One of the initial ontologies is extended so that it includes the required part of the other initial ontology, which preserves its internal structure and autonomy.

Absorption (Fig. 5) is used when a new (relatively small) ontology is introduced into a system. The existing initial ontology is extended so that it includes the other initial ontology, which preserves its internal structure but loses (partially or completely) its autonomy.

Flat Fusion (Fig. 6) assumes creating a new ontology, which contains initial ontologies. The initial ontologies dissolve within the new one and do not preserve their internal structures and autonomy.

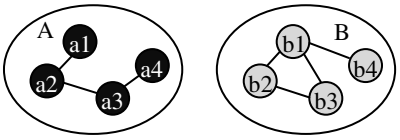


Fig. 1. Source ontologies

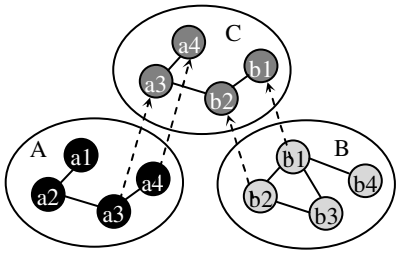


Fig. 2. Selective fusion

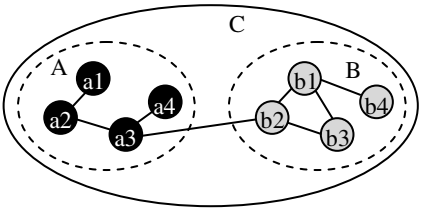


Fig. 3. Simple fusion

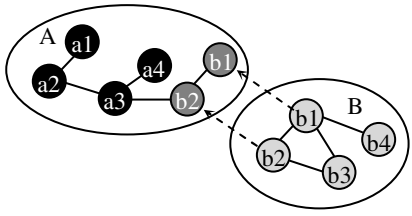


Fig. 4. Extension

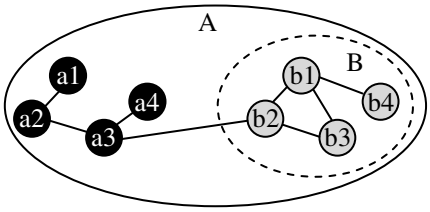


Fig. 5. Absorption

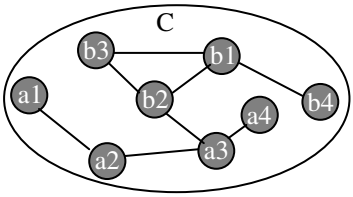


Fig. 6. Flat fusion

Utilizing the ontology fusion patterns considerably accelerates the ontology fusion processes due to typification of fusion schemes.

3 Ontological Knowledge Representation

In the approach the ontological model is described using the formalism of Object-Oriented Constraint Networks (OOCN). Application of constraint networks allows simplifying the formulation and interpretation of real-world problems which in the areas of management, engineering, manufacturing, etc. are usually presented as constraint satisfaction problems [28]. This formalism (its detailed description can be found in [29]) supports declarative representation, efficiency of dynamic constraint solving, as well as problem modelling capability, maintainability, reusability, and extensibility of the object-oriented technology. OOCN provides compatibility of ontology model for knowledge representation and internal solver representations.

The formalism of OOCN describes knowledge by sets of classes, class attributes, attribute domains, and constraints. Concept "class" in OOCN notation is introduced instead of concept "object" in the way object-oriented languages suggest.

The OOCN paradigm defines the common ontology notation used in the system. According to this representation an ontology (A) is defined as: $A = (O, Q, D, C)$ where: O – a set of *object classes* ("*classes*"); each of the entities in a class is considered as an *instance* of the class. Q – a set of class attributes ("*attributes*"). D – a set of attribute domains ("*domains*"). C – a set of *constraints*.

For example, the following classes can be mentioned: **truck** (o_1), **container** (o_2), **vehicle** (o_3), **transportation company** (o_4), and **transportation** (o_5). Examples for attributes are: **capacity** (q_1), **cost** (q_2), and **transportation cost** (q_3). The same attributes can be assigned to different classes and have different domains. An example domain is **real numbers** ($d_1 = \Re$).

For the chosen notation six types of constraints have been defined. Below, these constraints are illustrated with some examples:

$C^I = \{c_1^I, c_2^I, \ldots\}$, $c^I = (o, q)$, $o \in O$, $q \in Q$ – accessory of attributes to classes. E.g. **capacity** (q_1) is an attribute of the class **truck** (o_1): $c_1^I = (o_1, q_1)$; **cost** (q_2) is an attribute of the class **transportation** (o_5): $c_2^I = (o_5, q_2)$; **transportation cost** (q_3) is an attribute of the class **container** (o_2): $c_3^I = (o_2, q_3)$.

$C^{II} = \{c_1^{II}, c_2^{II}, \ldots\}$, $c^{II} = (o, q, d)$, $o \in O$, $q \in Q$, $d \in D$ – accessory of domains to attributes. E.g., the attribute **capacity** (q_1) of the class **truck** (o_1) is described by real numbers: $c_1^{II} = (o_1, q_1, \Re)$.

$C^{III} = \{c_1^{III}, c_2^{III}, \ldots\}$, $c^{III} = (\{o\}, True \vee False)$, $|\{o\}| \geq 2$, $o \in O$ – classes compatibility (compatibility structural constraints). E.g., all instances of the class **container** (o_2) are compatible with all instances of the class **truck** (o_1): $c_1^{III} = (\{o_1, o_2\}, True)$.

$C^{IV} = \{c_1^{IV}, c_2^{IV}, \ldots\}$, $c^{IV} = \langle o', o'', type \rangle$, $o' \in O$, $o'' \in O$, $o' \neq o''$ – hierarchical relationships (hierarchical structural constraints) "is a" defining class taxonomy ($type$=0), and "has part"/"part of" defining class hierarchy ($type$=1). E.g., **truck** (o_1) is a **vehicle** (o_3), i.e., all instances of the class **truck** (o_1) are instances of the class **vehicle** (o_3): $c_1^{IV} = \langle o_3, o_1', 0 \rangle$; an instance of the class **truck** (o_1) can be (but not necessary is) a part of an instance of the class **transportation company** (o_4), i.e., a particular truck can belong to a particular transportation company, or can be independent: $c_2^{IV} = \langle o_4, o_1', 1 \rangle$. The topmost class in the taxonomy of classes of the ontology is **Thing**.

$C^V = \{c^V\}$, $c^V = (\{o\})$, $|\{o\}| \geq 2$, $o \in O$ – associative relationships ("one-level" structural constraints). E.g., an instance of the class container (o_2) can be (but not necessary is) associated with an instance of the class truck (o_1), i.e., a particular container is transported by a particular truck: $c_1^V = (\{o_1, o_2\})$.

$C^{VI} = \{c^{VI}\}$, $c^{VI} = f(\{o\}, \{o, q\}) = True \vee False$, $|\{o\}| \geq 0$, $|\{q\}| \geq 0$, $o \in O$, $q \in Q$ – functional constraints referring to the names of classes and attributes. E.g., **cost** (q_2) of **transportation** (o_5) is a sum of **transportation costs** (q_3) of all **containers** (o_2): $o_5 q_2 = \sum_{o_2} o_2 q_3$.

Below, two auxiliary definitions used in the following section are presented.

Definition 1. Relationship *PseudoParent* of order n between two classes of an ontology: class o_p is a *PseudoParent* of order n for class o of ontology A ($o_P = Pp_n^A(o)$), if $o_p \neq Thing$ and there exists a sequence $o_1, \ldots, o_n, n \geq 2$: $o_1 = o_p, o_n = o, \forall i \neq n \; \exists c = \langle o_i, o_{i+1} \rangle \in C^{IV}$, i.e. class o_p takes a higher position in the taxonomy than o.

Definition 2. Relationship *RemotePseudoParent* of order n between two classes of an ontology: class o_p is a *RemotePseudoParent* of order n for class o of ontology A ($o_P = RPp_n^A(o)$), if $o_p \neq Thing$ and there exists a sequence $o_1, \ldots, o_n$, $n \geq 2$: $o_1 = o_p, o_n = o, \forall o_i, i \neq n \quad \exists c = (o_i, o_{i+1}) \in C^V \vee \exists c = \langle o_i, o_{i+1} \rangle \in C^{IV}$, i.e. class o_p either (1) takes a higher position in the taxonomy than o or (2) is related with class o by an associative constraint directly or indirectly.

4 Ontology Alignment

Ontology alignment is a set of correspondences between two or more (in case of multiple matching) ontologies obtained as a result of the ontology matching process [30]. In this section the complicated ontology alignment situations, which may arise during setting relationships between elements and the rules of their processing, are presented. These rules are valid for both straight and reverse directions.

Notations:

Source – ontology mapped;
Destination – ontology mapped to;
○ – class;
● – attribute;
— – associative relationship;
| \ – hierarchical relationships or "class-attribute" relationship;
⤏ – correspondence relationship.

Class-to-class Alignment. A class *a″* from the Source corresponds (maps) to a class *a′* from the Destination; a subclass *b″* of the class *a″* does not correspond to any class from the Destination. In this case search "in depth" does not stop and if a subclass *c″* of the class *b″* corresponds to a class *c′* from the Destination, the class *c′* becomes a subclass of the class *a′*, and the class *c″* becomes a subclass of the class *a″*. Experts can make a decision about including or not the new class *b″* into the common ontology (Fig. 7).

Attribute-to-attributes Alignment. An attribute *attr″* of the class *a″* from the Source corresponds to several attributes (a set of attributes) *ATTR′* of the class *a′* from the Destination. In this case all the attributes from *ATTR′* and methods for values conversion should be added into the common ontology (Fig. 8).

Class-to-classes Alignment. A class *a″* from the Source corresponds to several classes (a set of classes) *A′* from the Destination. In this situation all the classes from *A′* and conditions of selection among these classes are added into the common ontology. Attributes and subclasses of the class *a″* are mapped into attributes or subclasses of the classes from *A′* (Fig. 9).

Class-to-attribute Alignment. A class *a″* from the Source corresponds to a class *a′* from the Destination; a class *b″* associatively connected to the class *a″* from the Source corresponds to an attribute *attr′* of the class *a′* from the Destination. In this situation all the attributes and subclasses of the class *b″* are mapped to the attribute *attr′* with appropriate conversion methods and conditions are also added (Fig. 10).

Subclass-to-attribute Alignment. A class *a″* from the Source corresponds to a class *a′* from the Destination; subclass *b″* of the class *a″* corresponds to an attribute *attr′* of the class *a′*. In this situation all the subclasses of the class *b″* are mapped to the attribute *attr′* or possibly to other attributes of the class *a′*, with appropriates conversion methods and conditions being also added (Fig. 11).

5 Example of the Ontology Alignment

The common ontology (*A*) taxonomy is presented in Fig. 12. A database being integrated with its ontology ($\hat{A}(S)$) taxonomy is presented in Fig. 13. Examples for the main definitions and patterns of the ontology alignment are presented below.

PseudoParent of level 3 (Fig. 14): $o = \hat{A}.Focus$, $o_p = \hat{A}.Vehicle$; i.e. the class of $\hat{A}.Vehicle$ is a *PseudoParent* of the class of $\hat{A}.Focus$.

RemotePseudoParent of level 2 (Fig. 15): $o = SPI$, $o_p = \hat{A}.Focus$; i.e. the class of $\hat{A}.Focus$ is a *RemotePseudoParent* of the class of *SPI*.

Simple Class-to-class alignment (Fig. 16): $o'' = \hat{A}.Vehicle$, $O' = \{o'\}$, $o' = A.Vehicle$.

Simple attribute-to-attribute alignment (Fig. 20): $q'' = Volume$, $Q' = \{q'\}$, $q' = A.Engine$.

$o'' = SPI$ (Owner of *Volume*); *SPI* has already been mapped into *A.Engine*.

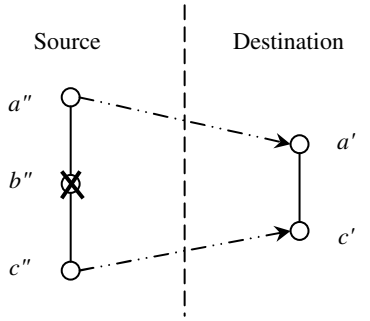


Fig. 7. Class-to-class alignment

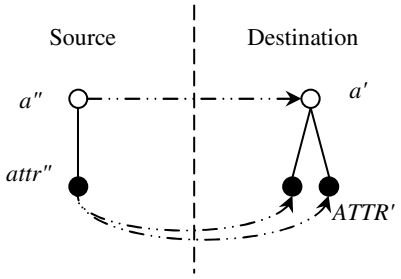


Fig. 8. Attribute-to-attributes alignment

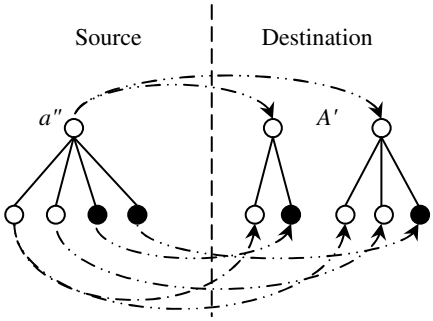


Fig. 9. Class-to-classes alignment

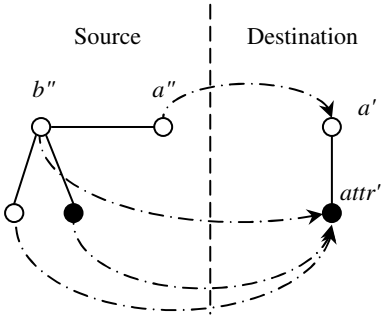


Fig. 10. Class-to-attribute alignment

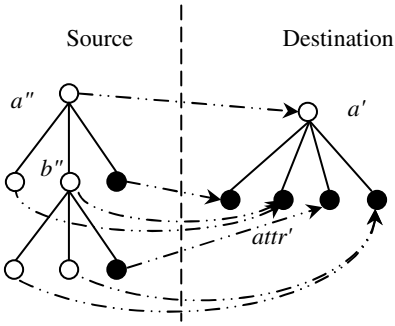


Fig. 11. Subclass-to-attribute alignment

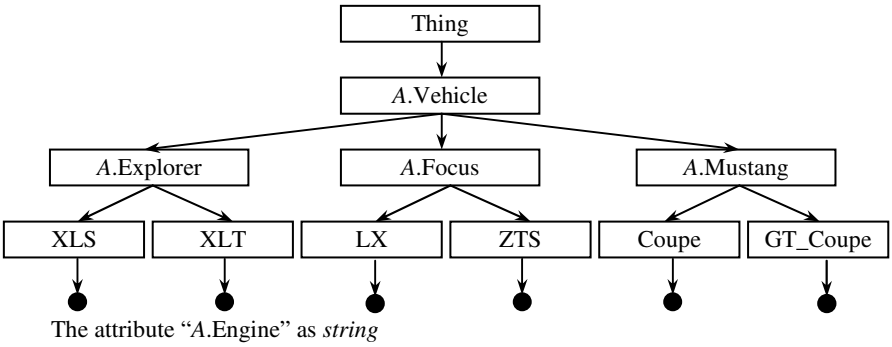


Fig. 12. Taxonomy of the common ontology

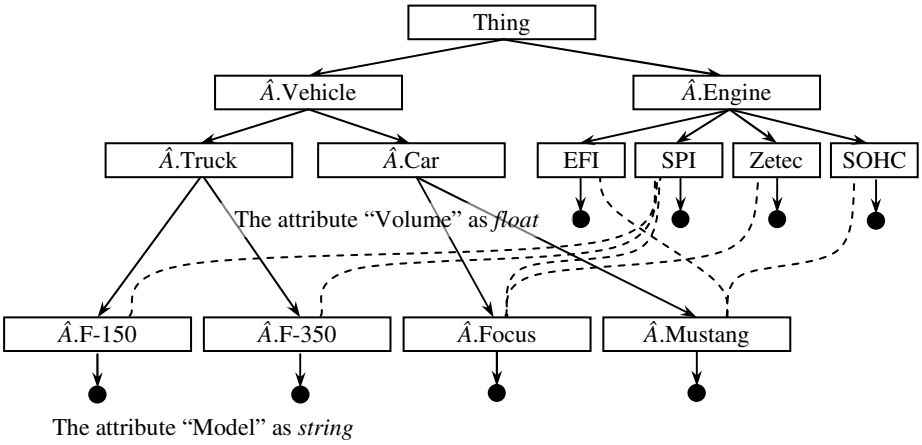


Fig. 13. Taxonomy of the database ontology

Class-to-class alignment (Fig. 17):

$o'' = \hat{A}.Focus$, $O' = \{o'\}$, $o' = A.Focus$.

$o''_1 = \hat{A}.Vehicle$ (*PseudoParent* of $\hat{A}.Focus$), $o'_1 = A.Vehicle$ (*PseudoParent* of $A.Focus$); $\hat{A}.Vehicle$ has already been mapped into $A.Vehicle$.

Subclass-to-attribute alignment (Fig. 18):

$q'' = Model$, $O' = \{o'_1, o'_2\}$, $o'_1 = LX$, $o'_2 = ZTS$.

$o'' = \hat{A}.Focus$ (has already been translated), $o'_1 = A.Focus$ (*RemotePseudoParent* of both *LX* and *ZTS* since it is a direct parent of these classes); $\hat{A}.Focus$ has already been mapped into $A.Focus$.

Class-to-attribute alignment (Fig. 19):

$o'' = SPI$, $Q' = \{q'\}$, $q' = A.Engine$.

$o''_1 = \hat{A}.Focus$ (*RemotePseudoParent* of *SPI*), $o'_1 = LX$ (owner of $A.Engine$), $o'_2 = A.Focus$ (PseudoParent of *LX*); $\hat{A}.Focus$ has already been mapped into $A.Focus$.

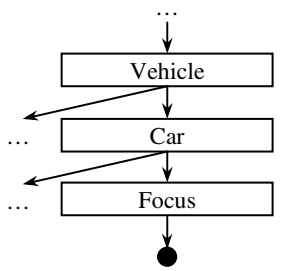


Fig. 14. PseudoParent of level 3

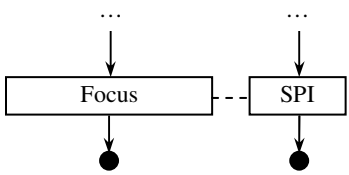


Fig. 15. RemotePseudoParent of level 2

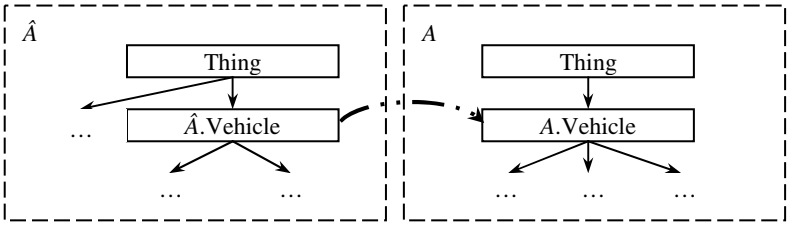


Fig. 16. Simple "Class-to-class" alignment

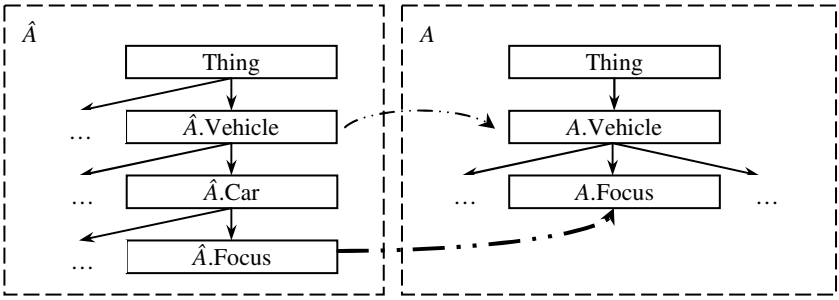


Fig. 17. Class-to-class alignment

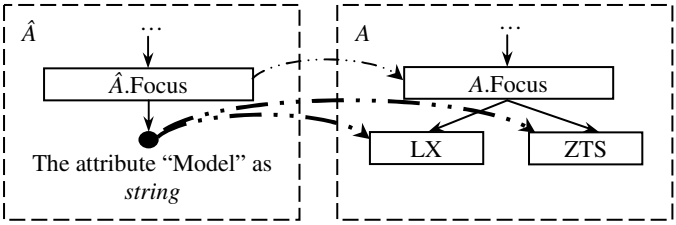


Fig. 18. Subclass-to-attribute alignment

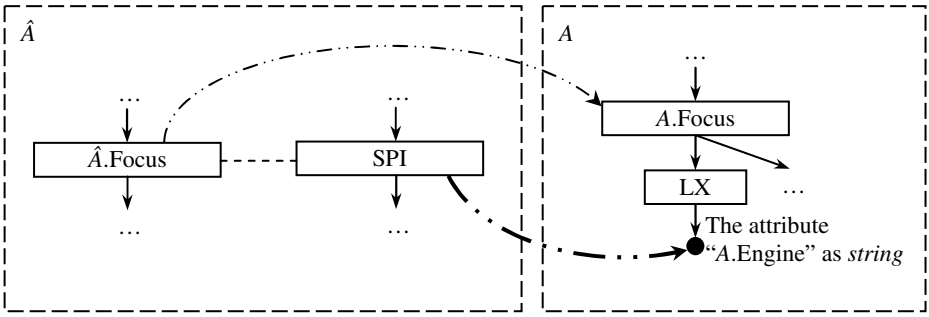


Fig. 19. Class-to-attribute alignment

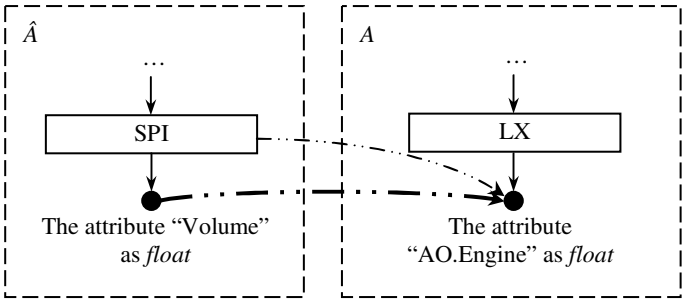


Fig. 20. Simple attribute-to-attribute alignment

6 Conclusion

The paper proposes an approach using an ontological model to solve the problem of heterogeneity. This is one of the key tasks in integration of different IT systems of an enterprise or of different enterprises, which requires semantic integration of their data and workflow models. The paper concentrates on description of ontology fusion and ontology alignment patterns. Utilizing of the patterns considerably accelerates the ontology fusion and matching processes due to typification of fusion and alignment schemes. An illustrative example from an automotive production demonstrates some of the patterns.

Acknowledgements. The research presented is motivated by the Integrated Project FP6-IST-NMP "Intelligent Logistics for Innovative Product Technologies" sponsored by European Commission (2004-2008), and sponsored by grants # 12-07-00298-a, # 12-07-00302-a, # 11-07-00058-a, and # 11-07-00045-a of the Russian Foundation for Basic Research, project # 213 of the research program "Intelligent information technologies, mathematical modelling, system analysis and automation" of the Russian Academy of Sciences, and project 2.2 "Methodology development for building group information and recommendation systems" of the basic research program "Intelligent information technologies, system analysis and automation" of

the Nanotechnology and Information technology Department of the Russian Academy of Sciences. The paper is due to the project COBIT sponsored by the Swedish Foundation for International Cooperation in Research and Higher Education.

References

1. ILIPT Project, http://www.ilipt.org (accessed May 14, 2010)
2. Lind, M., Seigerroth, U.: Multi-Layered Process Modeling for Business and IT Alignment. In: Proceedings of the 43rd Hawaii International Conference on System Sciences (HICSS), pp. 1–10 (2010)
3. Bradfield, D.J., Gao, J.X., Soltan, H.: A Metaknowledge Approach to Facilitate Knowledge Sharing in the Global Product Development Process. Computer-Aided Design & Applications 4(1-4), 519–528 (2007)
4. Chan, E.C.K., Yu, K.M.: A framework of ontology-enabled product knowledge management. International Journal of Product Development, Inderscience Publishers 4(3-4), 241–254 (2007)
5. Patil, L., Dutta, D., Sriram, R.: Ontology-based exchange of product data semantics. IEEE Transactions on Automation Science and Engineering 2(3), 213–225 (2005)
6. Goldkuhl, G., Röstlinger, A.: Towards an integral understanding of organisations and information systems: Convergence of three theories. In: Proceedings of the 5th International Workshop on Organisational Semiotics (2002)
7. Latour, B.: Technology is society made durable. In: Law, J. (ed.) A Sociology of Monsters: Essays on Power, Technology and Domination, pp. 103–131. Routledge & Kegan Paul (1991)
8. Uschold, M., Grüninger, M.: Ontologies: Principles, methods and applications. Knowledge Engineering Review 11(2), 93–155 (1996)
9. Doan, A., Hai, D., Jayant, M., Pedro, D., Alon, H.: Ontology Matching: A Machine Learning Approach. In: Handbook on Ontologies in Information Systems. Springer (2004)
10. Doan, A., Madhavan, J., Domingos, P., Halevy, A.: Learning to map between ontologies on the semantic web. In: Proceedings of the 11th International Conference on World Wide Web, pp. 662–673 (2002)
11. Hu, W., Jian, N., Qu, Y., Wang, Y.: GMO: A Graph Matching for Ontologies. In: Proceedings of the K-CAP Workshop on Integrating Ontologies, pp. 43–50 (2005)
12. Alasoud, A., Haarslev, V., Shiri, N.: An Effective Ontology Matching Technique. In: An, A., Matwin, S., Raś, Z.W., Ślęzak, D. (eds.) ISMIS 2008. LNCS (LNAI), vol. 4994, pp. 585–590. Springer, Heidelberg (2008)
13. Hovy, E.: Combining and standardizing largescale, practical ontologies for machine translation and other uses. In: Proceedings of the First International Conference on Language Resources and Evaluation (LREC), pp. 535–542 (1998)
14. Mitra, P., Wiederhold, G., Jannink, J.: Semi-automatic Integration of Knowledge Sources. In: Proceedings of the 2nd International Conference on Information Fusion, FUSION 1999 (1999)
15. Mitra, P., Wiederhold, G., Kersten, M.L.: A Graph-Oriented Model for Articulation of Ontology Interdependencies. In: Zaniolo, C., Grust, T., Scholl, M.H., Lockemann, P.C. (eds.) EDBT 2000. LNCS, vol. 1777, pp. 86–100. Springer, Heidelberg (2000)
16. Noy, N., Musen, M.: Anchor-PROMPT: Using Non-Local Context for Semantic Matching. In: Proceedings of the Workshop on Ontologies and Information Sharing at the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI 2001 (2001)

17. Castano, S., Ferrara, A., Montanelli, S.: H-Match: an Algorithm for Dynamically Matching Ontologies in Peer-based Systems. In: Proceedings of the 1st VLDB International Workshop on Semantic Web and Databases, SWDB 2003 (2003)
18. Serafini, L., Bouquet, P., Magnini, B., Zanobini, S.: An algorithm for matching contextualized schemas via SAT. Technical report, DIT University of Trento, Italy (2003)
19. Noy, N., Musen, M.: SMART: Automated Support for Ontology Merging and Alignment. In: Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling, and Management (1999)
20. McGuinness, D.L., Fikes, R., Rice, J., Wilder, S.: An Environment for Merging and Testing Large Ontologies. In: Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning, KR 2000 (2000), http://www.ksl.stanford.edu/software/chimaera/ (accessed May 12, 2012)
21. Aumueller, D., Do, H., Massmann, S., Rahm, E.: Schema and Ontology Matching with COMA++. In: Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, pp. 906–908 (2005)
22. Melnik, S., Garcia-Molina, H., Rahm, E.: Similarity flooding: a versatile graph matching algorithm and its application to schema matching. In: Proceedings of the 18th International Conference on Data Engineering, pp. 117–128 (2002)
23. Cruz, I., Antonelli, F., Stroe, C.: Efficient Selection of Mappings and Automatic Quality-Driven Combination of Matching Methods. In: Proceedings of the Fourth International Workshop on Ontology Matching (2009)
24. Ritze, D., Meilicke, C., Šváb-Zamazal, O., Stuckenschmidt, H.: A pattern-based ontology matching approach for detecting complex correspondences. In: Proceedings of the Fourth International Workshop on Ontology Matching (2009)
25. Giunchiglia, F., Maltese, V., Autayeu, A.: Computing Minimal Mappings. In: Proceedings of the Fourth International Workshop on Ontology Matching (2009)
26. Klein, M., Fensel, D., Kiryakov, A., Ognyanov, D.: Ontology Versioning and Change Detection on the Web. In: Gómez-Pérez, A., Benjamins, V.R. (eds.) EKAW 2002. LNCS (LNAI), vol. 2473, pp. 197–212. Springer, Heidelberg (2002)
27. Enterprise Integration Patterns, http://www.eaipatterns.com/ (accessed March 26, 2012)
28. Baumgaertel, H.: Distributed constraint processing for production logistics. IEEE Intelligent Systems 15(1), 40–48 (2000)
29. Smirnov, A., Levashova, T., Shilov, N.: Semantic-oriented support of interoperability between production information systems. International Journal of Product Development, Inderscience Enterprises Ltd. 4(3/4), 225–240 (2007)
30. Euzenat, J., Shvaiko, P.: Ontology matching. Springer (2007)

Cloud-Based Online Learning Platforms

Stefan Bensch and Markus Rager

Center for ERP Systems, Augsburg University, Universitätsstr. 16,
86159 Augsburg, Germany
{stefan.bensch,markus.rager}@wiwi.uni-augsburg.de

Abstract. Increasing organizational investment in technology for online learning platforms emphasizes how important it is for scholars and practice to understand and investigate technology and organizational developments. Online learning platforms are one component of education in university environments. Over the last decade an extensive offer of different online learning platforms and the fitting technology has been established. New technical capabilities are decisive to advance further development. Universities and companies face similar challenges in order to react flexibly to new requirements. The paper at hand strives to answer the question how the cloud computing concept can help technically and organizationally to achieve an efficient and sustainable utilization of resources for online learning platforms. In an organizational respect cloud solutions fit into the concept of an educational network and can stimulate the development of sustainable learning platforms-as-a-service.

Keywords: Online Learning Platform, Cloud Computing, Conceptual Framework.

1 Introduction

The increasingly popular online learning platforms have become the focus of current research. Surveys by the German Federal Institute for Vocational Education and Training (BIBB) established factors that hinder a rapid introduction of online learning platforms [36]. One of those factors is the confusing variety of online learning platform offerings [43]. Apart from these hindering factors especially the technical sphere could be identified as a key factor for the success of online learning platform concepts [10, 36]. Learning platforms are complex technical software architectures due to the variety of integrated applications and media formats. Online learning platforms are also complex software infrastructures with integrated task-specific applications, which support learning scenarios [22]. One of the main tasks for the operation of online learning platform infrastructures is to meet technological challenges of these platforms [35]. Choosing a partner from the confusing market makes the realization of online learning platforms difficult [26] in an inter-organizational respect [8]. It is of interest in both an organizational and a technical respect whether cloud computing improves this situation in value networks as the next generation of online learning platforms. A *value network* provides web tools and collaborative services, such as lesson plans, videos, curriculum resources, student interactives etc. A value network

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 165–176, 2012.

in this sense is a network of organizations that work across multiple value chains [38]. The value network for educational platform services as whole will bring together separate service provider as whole in a network [12]. The object of the value network is to deliver the right educational platform service in the right format, independent from location and time, in an approach driven by customer demand. For this purpose location-independent information technologies are used. Based on information about the network participants the value network can finally be formed. Thus, the path leads from classic online learning platforms to Software-as-a-Service (SaaS) offerings, which can technically and organizationally link multiple data sources, training providers and educational customers to each other. An examination of networks seems to be in vogue and expands the scientifically themed educational networks [8]. The examination of the organizations and institutions promises a high potential for innovation. Currently the interest is promoted by the Federal Ministry of Economics and Technology within the framework of the so-called 'Trusted Cloud' [15].

A systematic examination of the subject of cloud computing shall show future potentials and fields for the design of powerful online learning platforms in order to facilitate teaching in an university environment. This article examines two key issues which will be investigated as a rule through the design science paradigm for design-oriented research as an accepted method of information systems research [29, 37].

- What are the technical and organizational potentials of cloud computing for efficient and sustainable online learning platforms?
- How can current requirements be met?

Thereby shall be shown whether the design of cloud approaches and the provision of services for established and future online learning platforms are likely to meet technical and organizational challenges in an university environment. In chapter 1 the need for the investigation of potentials is motivated and the research design described. In chapter 2 the current state of research on cloud computing is systematically demonstrated with reference to established learning platforms. Furthermore, functions and concepts of e-learning scenarios in the university environment are outlined. Potentials are combined in chapter 3. The article documents in potential requirements for the implementation of online learning platforms in the cloud in an university environment in chapter 4. Chapter 5 gives a summary and an outlook on further research questions.

2 Online Learning Platforms in the University Environment

Online learning platforms in the university context are isolated solutions often used in isolation of the university IT infrastructure [22]. Even in one faculty the use of different online learning platforms can be observed. This is the case although the functions of online learning platforms overlap [22, 43]. This has been documented in recent studies which evaluated and established feature catalogues [43]. Therefore, the multiple uses of online learning platforms, components and contents are interesting from an economic and technological point of view [49]. The multiple uses relate to teaching strategies, multimedia learning materials, system components or entire educational environments. The goal of the multiple uses of multimedia learning

materials can be achieved by employing important standards (e.g. LOM, SCORM, AICC). For the multiple uses of e-learning technologies there are promising approaches such as the development of service-oriented architecture [27], or component application and component system frameworks [41, 52]. Current interest is increased by the growing popularity of cloud computing technology. IT resources and services should be provided for online learning platforms over the Internet [15]. In the past this aspect has already been postulated as the next generation of e-learning infrastructure [17].

2.1 Current State of Research on Cloud Computing in e-Learning

Published scientific articles about online learning platforms in the cloud were collected with a systematic review of literature. An investigation by keywords in the established online databases EbscoHost [21], ACM Portal [1], AIS Digital Library [2], IEEE Xplore [30] and the Internet was conducted for this purpose in April 2012. In the online databases the search strings were alike ((cloud OR cloud computing OR SaaS OR PaaS OR IaaS OR virtualization) AND (e-learning OR blended learning)) in singular and plural (inflectional), in German and English language in abbreviated and full form, scans with and without hyphen in the title and the abstract. The Internet search has been made with the search engine Google.

It is striking that only a few scientific papers have been published starting in 2008. The articles examine the economic *organizational perspective* on the one hand and the conceptual composition of cloud actors and their roles in a new cloud computing value network on the other hand. Thereby the focus is on the conceptual development of a generic value network [33]. Furthermore, positive impact and challenges of cloud computing architectures to e-learning solutions are examined [7, 23, 24]. Proposals for project management are presented [39]. Further costs potentials through cloud computing are examined for the education sector [13]. A framework for user-controlled data analysis is discussed [53]. Efficiency measures have been proposed for the long-term overall efficiency of cloud computing usage in the field of e-learning systems [39]. Other articles share first experiences with the employment of clear cloud architecture in practice [34].

From a *technical perspective* mechanisms of e-learning system architecture with cloud computing infrastructure were examined with respect to stability, balance, efficient usage of resources and sustainability of e-learning ecosystem [17]. Furthermore, architectures, key components and technologies for virtualization are discussed, such as improvements of performance, availability and scalability of e-learning technologies in order to emphasize but also to tackle risks of data management [18].

The results suggest that the introduction of cloud computing in online learning platforms is feasible and that significant improvements can be expected.

2.2 Information Technically Supported Teaching Concepts

Nowadays *e-learning* is a basic component of modern teaching methods [20]. The use of new information and communication technologies in all sectors of education, thus, takes on an important and much-debated role. *Traditional class lectures* like

face-2-face learning stand in contrast to information technically supported teaching concepts, where the teacher teaches educational contents in direct contact with his or her students. To combine the strengths of both methods the use of so-called *blended learning* concepts (often referred to as hybrid concepts) has been increased in recent years [28]. In this concepts traditional classroom phases and self-directed online learning modules take turns. The online learning phases are controlled mainly by synchronous *communication techniques* such as online tutorials and online chats, or even by asynchronous communication techniques via e-mail and online forums [36].

Three different approaches are described to support such concepts using information technology [9, 42]. The *enrichment approach* sums up the present events, which are enriched with multimedia components. A higher proportion of e-learning characterizes the *integrative concept*. The obligatory combination of e-learning and related present events coordinates corresponding learning methods. The *virtualization concept* aims at the realization of a high online-share. In fig. 1 these concepts are classified. E-learning is here described as a form of learning, which is supported organizationally and technically by information and communication technologies and network orientation. E-learning can be based on heterogeneous technologies and support various educational scenarios. Typical technological representatives are in excerpts computer based training (CBT), authoring tools, virtual classrooms and learning management systems. Learning platforms integrate technological applications to e-solutions.

3 Cloud-Based Learning Platforms

The investigation and comparison of characteristics of the technological and organizational integration of e-learning concepts can determine the state of research, taking into account cloud solutions. Thus the starting point to analyze e-learning technologies is formed in the cloud. Starting with the referenced concepts and criteria potentials are described by the use of cloud computing for online learning platforms.

3.1 Cloud Computing

Cloud computing has been consolidated in the research field of information systems. The topic is named in numerous workshops, conferences and journals. There is currently an attempt to establish a precise definition for the cloud computing phenomenon in science and practice. However, there is no commonly accepted definition for cloud computing. Recent articles accumulate for these purpose systematically scientific publications, expert opinions and pragmatic descriptions of practice, and attempt a comprehensive definition e.g. [11, 33]. The procedure turns out basic concepts and general objectives. The definitions have often in common that the term cloud computing addresses infrastructure-, platform- and application layer [51]. Yourseff and Silva [51] develop an ontology with two other layers: kernel software and firmware / hardware to sort access to overlying layers of virtualization technologies and hardware. The term can be summarized as follows: Cloud computing uses virtualization and the

Internet to provide flexible "infrastructure-", "platform-" and "applications-" as-a-service.

Infrastructure-as-a-Service (IaaS) is characteristic of the flexible and adaptive use of IT resources. One example is the use of virtual servers or storage (remote storage) from the cloud, which can be provided dynamically through virtualization. An additional level of abstraction is Platform-as-a-Service (PaaS). Instead of a virtual infrastructure software platforms will be provided if required. Thereby, integrated runtime and development environments are provided as a service. A known representative is the Google App Engine [25]. At the application level services are offered (SaaS). Users can access the hosted cloud services. SaaS will be defined as a method for deployment of software applications available in the Internet as a service. A typical scenario is the access via Internet browser. In the narrow sense of the term "Service", functionality can also be accessed via a "web service" interface. Thus the integration of external services into own applications are possible. The layer provides an alternative to the use of local applications [5]. Examples for this are typical office applications such as office software, which are installed as desktop applications on a remote computer. SyberWorks and other companies have been working "in the cloud“, and have hosted customer training and materials for some time on its own servers [16].

Cloud computing is a concept, a pool of easily usable and accessible resources, and is not considered as new technology. Cloud computing combines available technologies and offers this as service in a consolidated form with new approaches. Thereby, in scientific publications the potential as compared to existing approaches such as service-oriented architectures and grid computing topics with the demand-supply of services and the usage-based payment are discussed novella [33]. The resources can be configured dynamically to meet requirements of the architecture and allows a variable burden distribution. The price for such services is generally determined by pay-per-use models and Service Level Agreements (SLAs) for the use of resources [54].

The young cloud computing market can be explained economically as a value network [11]. Customers purchase services [4] such as platforms [11] and infrastructure [40] individually or in aggregate [31] from the service provider or aggregator. These and other scenarios have in common that the corporate network is of strategic importance for the participating actors.

3.2 Technological and Organisational Criteria of Online Learning Platforms

For further investigation not an isolated online learning system is considered but an e-learning concept taken from a practical context shall be described decomposed from cases of abstract use of current practice (see fig. 1). At the component level the assessment of cloud services is made on the scope of an university environment. For this purpose, a case study of the article "e-learning implementation strategies in higher education" was adapted [44]. The case study was deliberately chosen among the known limitations of the research methodology, since the representation inheres in the systematic teaching concepts [45, 47, 50]. Components reflect once more the technology integration. Elements are assigned to the components as "artifact result" of the

case study. Various elements can thus be combined into components and determine the degree of virtualization of the selected teaching concept.

Criteria in university operating structures for the successful use of e-learning *infrastructure and technology* are reliability, access speed, data security and system utilization. *Organizational* criteria are the basis of ICT skills of the participating organizations, the media-didactic competence, planning online courses or the production of content [32]. It is striking that the idea of networking in a conceptual respect is taken into account neither technically nor organizationally.

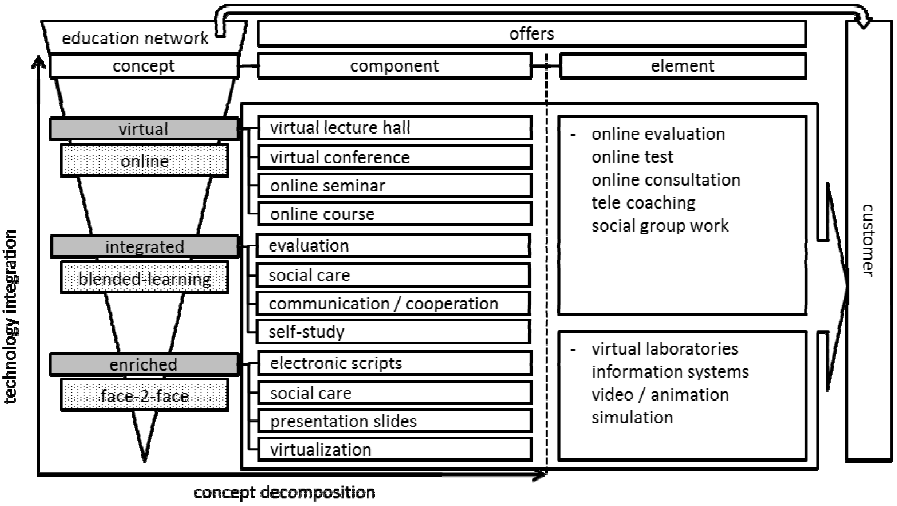


Fig. 1. Concept Decomposition in Learning Platforms

With the increased degree of virtualization through the use of various technologies and curriculums problems can arise between the actors. The integration of new e-learning technologies leads to the integration of new system providers. This integration can result in new challenges for the existing IT infrastructure. Thus, the infrastructure for online learning platforms consists of a number of different components, which are arranged by different organizational units (e.g. e-learning competence centers, media centers, computer centers). These structural units are to be combined with regard to information technology. Characteristic problems are difficulties in providing centralized, personalized and needs-based applications and the need-based data exchange between applications.

Furthermore, integrating educational content providers can cause organizational problems. For example, if such providers are companies that integrate the application system training into teaching, which is based on the dual model, software will be provided. With the introduction of business application software for training purposes, responsibilities on user level and technical operation level as well as senior offices of university and company become dependent. Thus, one has to reckon on difficulties when operating in an university environment, because due to the complexity it cannot

be foreseen in detail whether new online learning components can be used without problems. In many organizations the implementation of business application software causes an inner breaking test: The planned project term and the budget that was fixed in the run-up are often exceeded.

3.3 Potentials in Cloud-Based Online Platforms

Potential for technical and organizational effects extend over the different layers in the cloud for e-learning scenarios. This allows organizations to convert the data center into university wide shared and yet secure clouds, which offer infrastructure and storage as a service [14]. With Cloud solutions multitenancy is developed, which makes potentials of the central software installation, its maintenance and smaller storage requirement usable. Further impacts are expected on falling licensing costs. The establishment of a university-wide *infrastructure* also leads to an integration of services and thus synergies. Quality and efficiency characteristics are emerging. Furthermore, components and services can be combined to online learning platforms. Some integrational advantages, which can be realized on the SaaS layer, are the connection of different kinds of system such as online learning components with applications of libraries, or the linkage to examination management services. The compilation of online learning platforms as-a-service promotes the possibility of providing original needs-based e-learning offers. Such solutions would be extremely difficult to realize without the cloud approach which supports service coupling functions.

While one actor performs all the functions of creating complex training programs in an integrated model, the functions and, thus, the contribution to added value distribute over several parties in an educational network [48]. Based on component analysis in online learning platforms and the conducted systematic literature research it can be concluded that the research area shifts due to the use of cloud computing to value networks. This is the case because the information technological support enables the use of online learning platforms in the first place. Creation of value-added portions are increasingly produced through the creation of a value network in many industries [19]. Often-quoted examples are educational networks [35], or originate from the automotive industry [46]. The young educational cloud computing market can economically also be described as a value network (see 3.1). The corporate network is of strategic importance for the participating companies. Value networks are an integrated combination of companies with the aim to solve a customer-specific problem as a whole, economically and technically. Such concepts are complex and cannot be carried out without information technology support. [6]. The investigation shows that technological development leads to new actors and models on the e-learning market [35]. Technical and economic developments pose new challenges for science and practice. The discussion of the constellation of value networks with cloud computing for online learning platforms is still pending. The fields of application for cloud computing are diverse. Education networks were examined and are already supported by cloud computing. A representative example is WINFOLine, which provides individuals with educational content in the field of information systems.

Please note that the educational network depicted in Figure 2 does not entail all possible interrelations between the actors.

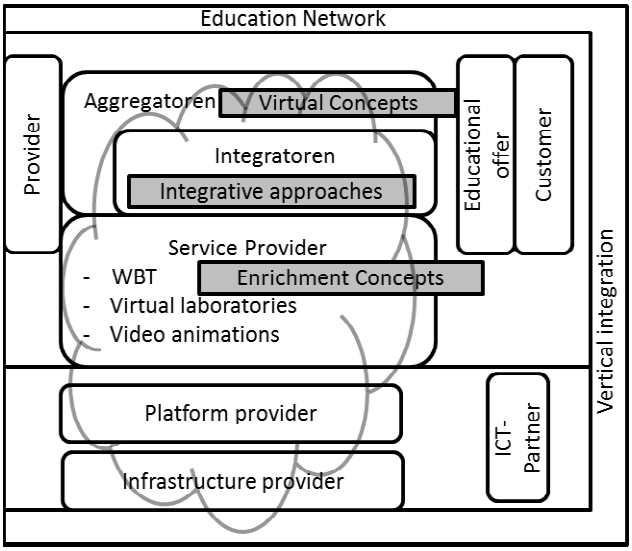


Fig. 2. Educational Cloud Network Approach

Service providers offer content, services or systems as an online cloud learning platform service. Customers access such services, which can already include training facilities, as well as central catalogue- or market-based cloud platforms. *Aggregators* combine existing services to service products in education. This new value-added service is offered through the platform. Further actors (*Integrator*) bring together information across all layers of existing systems. With the increased focus on core competences and the continuing technical standardization, the layer model of cloud computing leads to a variety of providers, who break the model of "single-provider" [3]. Purely technical aspects are provided by cloud partners. Thereby, the platform operators are also providers of infrastructure. The roles of the service providers and customers are, for example, taken on by universities and companies.

4 Conclusion for Research and Application

Requirements have to be met in several steps in order to make the potential of cloud computing in an university environment usable.

4.1 Practice Contribution

It is important for the application layer to note that the provided services, responsibilities and host structures are to be determined systematically. Within the consolidation of offered services there will be an attempt to highlight overlaps and convergences between the overall offers. For this purpose documented features can be extracted from services, compared, and examined under a coverage analysis of divergence,

convergence and completeness. Identified services and gaps in offerings are derived as artifact from the coverage analysis. Services and core functions are to be designed as interoperable as possible with the cloud approach. A technical infrastructure is required to ensure the service. Basic services are provided by the university IT infrastructure. An encompassing network infrastructure is needed to ensure the cloud offering, which is in an university environment operated largely by its own computing centers. Some individual cases will have to be investigated to find out whether and to what extent the technological suited XaaS can be operated by the university itself. Hereby, the use of technology by is determined through the reflection of the services based on the enrichment approach, integration approach and virtualization approach. The use and development of e-learning solutions is operated based on the interfaces of a cloud provider and the development environments with increasing intensity.

The distributed learning options for high school cross-training network can be brought together, enriched by the idea of networking and focusing on core services of actors.

4.2 Perspective on Research

Cloud offers are often classified on the level on which the individual services are provided [11]. Typical levels therefore are systematized and presented based on literature. Furthermore, explicit learning platform components are described in a concept-decomposition-diagram. The view of educational networks has been extended by the addition of technical actors towards a value network. Hence, the following requirements for research and application areas can be derived.

From the technical point of view it has to be considered in the next steps how established and future learning platform services can be provided in the cloud. Previous publications in this regard considered the feasibility and proposed first concepts. Although actors have already been examined in an e-learning cloud as well, the exploration of technological use of cloud computing in learning platforms is still in an early stage. Some publications already exist in related approaches to service-oriented architectures and education networks, which seem to be promising when transferred to cloud computing.

The contributions to cloud computing can also contribute to the acquisition of knowledge in the sense of this work. A comparison of these approaches and transferability are still due. This and other circumstances can be taken as an indication of the maturity lack in the available approaches.

5 Conclusion and Outlook

In the article at hand learning platform concepts were analyzed and examined with regard to their technical and organizational implementation of cloud computing. The use of cloud computing in the study of learning platforms leads to considerable advantages for operators, teachers and learners compared to the traditional use of learning concepts, and can do more than only reduce costs. The potential and requirements

derived here provide a basis, which is particularly important for those organizations that previously had difficulties with the operation and use of learning platforms and with the creation and delivery of content. Cloud computing enables bundling effects in the design of hybrid teaching forms in learning platforms. Thus derived requirements enhance the seamless use of new technologies in many scenarios.

In a *technical* respect *online learning platform* services can be shifted to complex systems in order to meet requirements of the chosen teaching concept. The different levels encompass the technological infrastructure, the platform for the operation of e-learning applications and services. The architecture layers have the potential to create a mixture of service-oriented learning platforms in the sense of educational networks [35]. In an *organizational* respect the young and dynamic cloud market can be described as value network, in which customers can make use of services such as e-learning and IT infrastructure from a provider network of related educational partners [11].

The next step is to investigate how the technical and structural conditions can be created to bundle previous offerings in the cloud. For this purpose a study of current cloud-operator could be provided in order to find out to which extend the proposed requirements are supported. In a following step a specific value network with the actions between the actors will be modeled.

References

1. ACM Digital Library, http://portal.acm.org/
2. AIS Electronic Library (AISeL) - Association for Information Systems Research, http://aisel.aisnet.org/
3. Bächle, M.: Web 2.0 viel mehr als ein Hype. Wirtschaftsinformatik & Management (January 2009)
4. Barros, A.P., Dumas, M.: The Rise of Web Service Ecosystems. IT Professional 8(5), 31–37 (2006)
5. Becker, A.: Nutzenpotenziale Service-orientierter Architekturen – Ergebnisse einer Expertenbefragung, Darmstadt (2008)
6. Bensch, S., et al.: Beschaffungsmanagement für hybride Leistungsbündel in Wertschöpfungsnetzwerken - Status Quo und Gestaltungsperspektiven. In: Wirtschafts informatik Proceedings 2011, Zürich, Schweiz, pp. 1–11 (2011)
7. Bensch, S.: Technical and Organizational Potentials of Value Networks for Ubiquitous Information Products and Services: Exploring the Role of Cloud Computing. In: 4th International Conference on Ubi-Media Computing (U-Media), pp. 71–76. IEEE (2011)
8. Bensch, S.: Technische und organisatorische Potentiale eines E-Learning Wertschöpfungsnetzwerks in der Cloud. In: MKWI Proceedings 2012, Braunschweig, Germany, pp. 1–12 (2012)
9. Bett, K., Wedekind, J.: Entscheidungsprozesse und Begleitmassnahmen bei der Auswahl und Einführung von Lernplattformen. Lernplattformen in der Praxis, Waxmann (2003)
10. Bohl, et al.: Nachhaltige Geschäftsmodelle für e-Bildungsdienstleistungen deutscher Hochschulen, E-Learning: Einsatzkonzepte und Geschäftsmodelle, pp. 241–254. Physica Verlag, Heidelberg (2005)
11. Böhm, M., et al.: Cloud Computing: Outsourcing 2.0 oder ein neues Geschäftsmodell zur Bereitstellung von IT-Ressourcen? Information Management Consulting 24(2), 6–14 (2009)

12. Vom Brocke, J., Campus, L.: Organisations gestaltung im E-Learning: Konzeption und Anwendung für die integrierte Prozessgestaltung an Großuniversitäten. In: Marktplatz Internet: Von e-Learning bis e-Payment. LNI, vol. P-72, pp. 157–164. Springer, Heidelberg (2005)
13. Chandra, D.G., Borah, M.D.: Cost benefit analysis of cloud computing in education. In: International Conference on Computing, Communication and Applications (ICCCA), Dindigul, Tamilnadu, pp. 1–6 (2012)
14. Cisco: Designing Secure Multi-Tenancy into Virtualized Data Centers. Cisco Systems, Inc., 170 West Tasman Drive, San Jose, CA 95134-1706, USA (2009)
15. The CloudCycle Project, http://www.cloudcycle.org/
16. e-Learning in "the Cloud", http://www.syberworks.com/articles/elearning-in-the-cloud-article.htm
17. Dong, B., et al.: An E-learning Ecosystem Based on Cloud Computing Infrastructure. In: Proc. of the 2009 Ninth IEEE Int. Conf. on Advanced Learning Technologies, pp. 125–127. IEEE, Washington, DC (2009)
18. Dong, B., Zheng, Q., Qiao, M., Shu, J., Yang, J.: BlueSky Cloud Framework: An E-Learning Framework Embracing Cloud Computing. In: Jaatun, M.G., Zhao, G., Rong, C., et al. (eds.) Cloud Computing. LNCS, vol. 5931, pp. 577–582. Springer, Heidelberg (2009)
19. Durst, S.M., Sucky, E.: Lieferantenentwicklung: Stand der empirischen Forschung. In: Bogaschewsky, R., et al. (eds.) Supply Management Research, Gabler, pp. 37–71 (2010)
20. Dykman, C.A., Davis, C.K.: Part One–The Shift toward Online Education. Journal of Information Systems Education 19(1), 6 (2008)
21. EBSCO - EBSCOhost Online Research Databases, http://www.ebscohost.com/
22. Frankfurth, A., Schellhase, J.: Potentiale serviceorientierter Architekturen für E-Learning-Infrastrukturen an Hochschulen. DeLFI, pp. 351-362 (2006)
23. Gad, S.H.: Cloud Computing and MapReduce for Reliability and Scalability of Ubiquitous Learning Systems. In: Proceedings of SPLASH Workshops, pp. 273–278. ACM, New York (2011)
24. Ghazizadeh, A.: Cloud Computing Benefits and Architecture in E-Learning. In: 7th Wireless, Mobile and Ubiquitous Technology in Education Proc., pp. 199–201. IEEE (2012)
25. GoogleApp: Google App Engine - Google Code, http://code.google.com/appengine/
26. Grohmann, M.G.: Ansatzpunkte zur Organisation virtueller Lernszenarien am Beispiel des Bildungsnetzwerkes WINFOLine (2002)
27. Guangzuo, C., et al.: A Web Services Oriented Framework for Dynamic E-Learning Systems. In: GCCCE 2004 International Conference, Hong Kong (2004)
28. Gupta, S., Bostrom, R.P.: Technology-Mediated Learning: A Comprehensive Theoretical Model. Journal of the Association for Information Systems 10(9), 686–714 (2009)
29. Hevner, A., et al.: Design Science in Information Systems Research. MIS Quarterly 28(1), 75–105 (2004)
30. IEEE Xplore - Home, http://ieeexplore.ieee.org/Xplore/dynhome.jsp
31. Katzmarzik, A.: Produktdifferenzierung für Software-as-a-Service-Anbieter. Wirtschafts informatik 53(1), 21–35 (2011)
32. Kerres, M., et al.: Technische und organisatorische Rahmenbedingungen für eine erfolgreiche Einführung und nachhaltige Nutzung von E-Learning an Hochschulen. DINI Schriften (2005)
33. Leimeister, S.L., et al.: The Business Perspective of Cloud Computing: Actors, Roles, and Value Networks. In: Proceedings of 18th European Conf. on Info. Systems, pp. 1–12 (2010)

34. Ma, Z.: The Applied Research of Cloud Computing in the Construction of Collaborative Learning Platform under E-learning Environment. In: 2010 International Conference on System Science, Engineering Design and Manufacturing. IEEE (2010)
35. Martin, D.O., et al.: Konfiguration WEB-Basierter Bildungsangebote in Private-Public-Partnership-Netzwerken. Arbeitsbericht des Fachgebiets Wirtschaftsinformatik der Universität Kassel (2005)
36. Napier, A.P., Smith, S.: Assessing Blended Learning: Student Outcomes and Perceptions. In: Proc. AMCIS, USA (2009)
37. Peffers, K., et al.: A Design Science Research Methodology for Information Systems Research. J. Manage. Inf. Syst. 24(3), 45–77 (2007)
38. Pibernik, R.: Flexibilitätsplanung in Wertschöpfungsnetzen. Zeitschrift für Betriebswirtschaft 71(8), 893–913 (2001)
39. Pocatilu, P., et al.: Measuring the Efficiency of Cloud Computing for e-learning Systems. WSEAS Transactions on Computers 9(1), 42–51 (2010)
40. Pueschel, T., Neumann, D.: Management of Cloud Infrastructures: Policy-Based Revenue Optimization. In: ICIS 2009 Proceedings, pp. 1–17 (2009)
41. Rautenstrauch, C., Turowski, K.: Common Business Component Model (COBCOM): Generelles Modell komponentenbasierter Anwendungssysteme. In: Internationale Tagung Wirtschaftsinformatik, Augsburg, pp. 681–695 (2001)
42. Schulmeister, R.: Virtuelle Universität. Virtuelles Lernen, Oldenbourg (2001)
43. Schulmeister, R.: Zur Didaktik des Einsatzes von Lernplattformen. In: Lernplattformen: Optimierung der Ausbildung oder didaktischer Rückschritt?, Dübendorf, pp. 11–19 (2005)
44. Seufert, S., Euler, D.: Nachhaltigkeit von eLearning-Innovationen: Fallstudien zu Implementierungsstrategien von eLearning als Innovationen an Hochschulen (2005)
45. Stake, R.E.: Case Studies. In: Denzin, N.K., Lincoln, Y.S. (eds.) Handbook of Qualitative Research, pp. 435–454. Sage, Thousand Oaks (2000)
46. VDA: Auto Jahresbericht 2008. VDA (Verband der Automobilindustrie), Frankfurt am Main (2008)
47. Weitz, B.O.: Möglichkeiten und Grenzen der Einzelfallstudie als Forschungsstrategie im Rahmen qualitativ orientierter Bildungsforschung. Universität-GHS Paderborn, Essen (1994)
48. Winand, M.H., et al.: Grundlagen zur Gestaltung von Geschäftsmodellen für akademische eBildungsdienstleistungen. E-Learning Geschäftsmodelle und Einsatzkonzepte, pp. 62–84. Gabler Verlag (2006)
49. Woelk, D.: E-learning, semantic web services and competency ontologies. In: World Conf. on Educational Multi-, Hypermedia & Telecommunications. pp. 2077–2078 (2002).
50. Yin, R.K.: Case Study Research – Design and Methods. Sage, Thousand Oaks (1994)
51. Youseff, S.: Toward a Unified Ontology of Cloud Computing. In: Grid Computing Environments Workshop (2008)
52. Zaha, J.M., Kelch, S.: Komponentenfindung in monolithischen objektorientierten Anwendungssystemen. In: Gesellschaft für Informatik. LNI, vol. P-57, pp. 151–162 (2004)
53. Zhang, J., et al.: A Framework of User-Driven Data Analytics in the Cloud for Course Management. In: Proceedings of the 18th International Conference on Computers in Education, Malaysia, pp. 1–5 (2010)
54. Zhang, Q., et al.: Cloud computing: state-of-the-art and research challenges. J. Internet Serv. Appl. 1(1), 7–18 (2010)

Serious and Simulation Games – A Definition Approach

Bastian Kurbjuhn

Otto von Guericke University Magdeburg
Faculty of Computer Science
Department of Technical and Business Information Systems
Very Large Business Applications Lab
Universitätsplatz 2, 39106 Magdeburg, Germany
BKurbjuhn@ovgu.de

Abstract. Information systems used in modern business become more complex because the holistic approach of process implementation turns increasingly into the focus of corporate's demands. To handle such Very Large Business Applications (VLBA), the recruiting of tomorrow's new experts is an important educational task. Business simulation games are an effective teaching method: On the one hand, the participants learn the theoretical concepts. On the other hand, they can adapt their new knowledge. But there are lots of different serious games, even in literature multiple expressions exist. In this paper, a literature review is performed to mark out different types of games with an educational character. It is found out, that the most common expressions are serious gaming and (business) simulation games.

Keywords: Serious Games, Simulation Games, definition.

1 Introduction

In relation to market globalisation, business applications grow because information technology enables competitive advantages. Topics like holistic process implementation or workflow support services are one reason why the complexity of industrial information systems raises. Besides the importance of theoretical background, a substantial demand on experts with practical skills who can handle such complex systems exists.

In addition, life includes – but is not limited to – making decisions, whether in private or business matter. Much of them are made intuitively; but critical situations, where the impact of decisions is elusive, are hardly to manage. Otherwise when a choice is made it is difficult to roll it back.

Educational games are a helpful instrument to train the handling of such situations while they simulate a special environment where the decisions could be analysed and evaluated. Once developed for military to train strategic and tactical thinking, they have broken into almost every field of life, e. g. as flight/driving simulators, for security education, for defence purposes, in corporate management or for handling information systems; either with more intensity on entertainment or with regarding to reality.

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 177–185, 2012.

According to their application, target group and focus, different terms have been developed in literature to define or describe such educational games. In order to presend an overview about the meanings of such terms, the paper presents the definition and shows how the meanings of the words differ from each other.

2 Definition Approaches

"Serious games" or "serious gaming" and "simulation games" are very common terms for games with a focus on education. But before being able to deal with different types of educational games, the term "game" itself has to be explained. This is to be done in this section.

2.1 Game

In 1939, Johan Huizinga delivered an extensive and approved definition of the term game (Ger.: Spiel) within his work "Homo Ludens". There, he compares it as a voluntary action or engagement within a specified dimension of time and space. This engagement follows voluntarily accepted but stringent rules. The participants undergo a feeling of excitement, pleasure and an awareness which differs from "normal life".[1]

Consequently, forced or appropriated activities are not captured by the term *game*, they would be covered by *seriousness* [2].

In English literature, there are three different comprehensions between the words *play, game* and *gambling*:

- *Play* could be put on a level with the above definition of Huizinga, respectively his description of *game* (Spiel).
- *Game* emphasises the rule-based concept. So, developing a rule-following strategy or tactic will be supported or is possible to dominate or win the game. Huizinga's broad definition fits here, too.
- *Gambling* consists of games of pure chance. In such games luck plays an enormous role. Although these sort of games obey a set of rules. But due to the extensive part of luck, a strategy could hardly influence the result.[2] Known games of chance are roulette and blackjack.

Words like video game or computer game are attempts to create a definition of the word game – detached by the word play – to underline its special rule-based character [3].

Salen and Zimmerman propose the following definition:

---

[1] Huizinga, J.: *"Spiel ist eine freiwillige Handlung oder Beschäftigung, die innerhalb gewisser festgesetzter Grenzen von Zeit und Raum nach freiwillig angenommenen, aber unbedingt bindenden Regeln verrichtet wird, ihr Ziel in sich selber hat und begleitet wird von einem Gefühl der Spannung und Freude und einem Bewusstsein des 'Andersseins' als das 'gewöhnliche Leben'"* [1].

[2] If the percentage of luck consists of almost 100 percent in a game, a strategy performs any effect on the result and is therefore ineffective.

"A game is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome" [4].

But many computer games give their players the opportunity to let them decide in with way they might perform, so this sort of video games does not target on a concrete aim, they offer a sand box environment, wherein the player could act freely as he or she might to. Therefore, they could not be summarized under the narrowly defined term game. They belong to free games. Examples of such free (video) games are SimCity[3], The Sims[4] or the Myst and Uru Universe[5]. Free games have been performed by children, even by adolescents and adults for centuries [3].

However, the meaning of a game varies from the age of human life: During infancy, gaming helps to promote cognitive skills and to develop social identity, e. g. by imitation of other people or by training/repetition of motion sequences. Gaming is predominantly a single activity until school readiness. Later, with the group activities, the competition character of games rises under an agreement of rules. Adults consider games on the one hand as a possibility to relax and to escape from pressure in reality. On the other hand, competitive aspects still play an important role [2].

In general, a game describes the activities in a specified dimension of space and time, which should be executed in a way to fulfil given aims without injuring predefined rules. The gaming process itself surrounds the participant in an atmosphere of tension, joy and challenge [2]. However, this atmosphere can also turn into frustration and sadness when the result of the game does not match expectations.

2.2 Serious Games

In the American-speaking world, Serious Games form the interface between entertainment technologies and applications in the institutional and educational sector.

Serious Games are used in many branches, e. g. in the medical sector, in military and in business sector to recruit or train employees. Nevertheless, they also take part in public sector in order to sensitize people for a special or relevant topic. Learning within a secure and realistic environment is a crucial advantage of serious gaming (e. g. flight simulators) [5]. Enriched with funny activities in a particular context, they are often used for education, planning and advertising purposes. Their playful and/or challenging design focusses on user motivation. In order to ensure good results, their development needs a "careful strategy" [19].

This genre of games was established in conjunction with the release of *America's Army* and the foundation of the Serious Games Initiative by Ben Sawyer and David Rejeski in 2002 (at least in the English-speaking world) [2]. But this expression has already been used before era of Information Technology: In 1968, Clark C. Abt used the expression for board and card games and defined it as follows:

"Reduced to its formal essence, a game is an activity among two or more independent decision-makers seeking to achieve their objectives in some limiting context. A more conventional definition would say that a game is a context with rules among

[3] http://www.simcity.com
[4] http://thesims.com
[5] http://www.cyanworlds.com

adversaries trying to win objectives. We are concerned with serious games in the sense that these games have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement" [6].

But a general definition has not been established until now. It is only evident, that serious games describe games or similar applications which are developed by instruments or technology of the entertainment software sector but do not serve only for playful belongings [5].

The parts of game-aspects (entertainment value) and informative or educational aspects should be balanced. So, the user could adapt his or her experiences for life. Serious Games are not limited on one genre[6] and can use every game technology or be conceived for every platform. The main effect of these games lies in learning experience [2].

An expanded definition is formulated by Ben Sawyer, co-founder of Serious Games Initiative:

"Any meaningful use of computerized game/game industry resources whose chief mission is not entertainment" [7].

Sawyer points out that a clear division between entertainment games and serious games could not be performed and is either not necessary. In accordance to the game America's Army, this game simulates a training situation for one person whether another person plays the game only for entertainment purposes. So, in general, every game, which is not intended only for entertainment, could be seen as a serious game [2].

Table 1 categorises serious games into the following subdivisions [5]:

Table 1. Classification of Serious Games [5]

Classification	Description	Examples
Corporate Games	Industrial applications for corporate training of education	Innov8, Sharkworld, KoCUA, Ultimate Team Play
Educational Games	Serious Games for use in scholar/university qualification with concrete learning objectives. Even games especially developed for children belong to that category.	DiaboloVR, Genius Politik, Luka und das geheimnisvolle Silberpferd
Health Games	Health games simulate concrete medicinal scenarios and encourage general health care	Glucoboy, Human Sim, Mind Habits, Pulse!!, Re-Mission
Persuasive Games	Political or social motivated games for the broad public which transport arguments and persuasion through edutainment	Darfur is Dying, Fatworld, Food Force, Global Conflicts – Palestine, PowerUp, September 12th, Serious Policy

[6] Simulation, adventure, strategy or shooter games are examples for game genres.

Furthermore, it is not clear if serious games could be put on a level with learning games or e-learning. For example, Hawlitschek subsumes all video games of the category "playing with educational content" or "learning with playful content" as learning games, whilst in serious games, both aspects are combined in a way that the user learns playfully [8].

E-learning refers to computational and online-aided learning. The knowledge transfer happens with the help of software. Otherwise, trainer and participants meet in virtual class rooms.

Marc Prensky defines the (digital) game-based learning as "[…] any learning game on a computer or online" [9]. So, the educational content has to be integrated in a way that the user should feel like a gamer and not like a learner. As described above, entertainment and education have to be balanced. Otherwise it is a learning program or a usual entertainment game [10].

2.3 Simulation Games

Originally used for tactical simulation and planning of war operations in military, simulation games have been adopted to civil training since 1950s because their application area has been extended as a result of the combination of war games, operations management and computer science [20].[7] In present times, simulation games now hold a special part of serious games. But not all serious games are simulation games. Serious gaming is a more general term for games, whilst simulation games have a special relation to reality because they concentrate on simplification (simulation) of existent problems in reality.

Geilhardt/Mühlbradt compare a simulation game (Ger.: Planspiel) with a constructed situation, in which one or more person(s) act(s) in a model of predefined rules. The behaviour will then be systematically observed and analysed [11].[8]

Kriz disassembles the term into three components [12]:

- **Simulation** of a closed (sub-)system. Simulation means the technical transformation of a given model. Separate elements could only influence or interact each other within simulation environment.
- **Game** refers to a bundle (set) of rules for structuring relevant and predefined processes
- **Role** is defined as function which the participants take over in dynamic systems.

From this decomposition, the characteristics of simulation games could be derived: Planning (description) and Game [13]. In addition to that, Adelsberger et al.

[7] Business simulation games supporting corporate management and leadership in enterprises (decision making) were the first civil games. As a first approach, the famous "BEER Game", developed in 1985, enriched those games with manufacturing content and engineering related topics [20].

[8] Geilhardt, T./Mühlbradt, T.: "konstruierte Situation, in der sich eine oder mehrere Person(en) in oder an einem […] Modell nach vorgegebenen Regeln verhalten, wobei das gezeigte Verhalten systematisch festgehalten und nach einem explizierbaren Kalkül bewertet werden kann" [11].

emphasise the importance of evaluation apart from description of scenario and simulation [14].

Even Léger adds the evaluation process as the final component of simulation games because the behaviour and decisions of participants/teams has to be reflected in order to be able to learn from the feed-back of the game leader [15].

By using simulation games, important processes and hierarchical structures existing in real life could be simulated and simplified for comprehension [11].

Comparison of Simulation Games and Simulations

In literature, there are partially contrary approaches of diversification between simulation games and simulations themselves. One possibility to divide is delivered by Kern. He cuts the dynamic role process into two sections [16]:

1. In the *action section* (Ger.: Aktionsbereich) the person acts in the sense of its represented role and makes decisions.
2. In the *reaction section* (Ger.: Reaktionsbereich) a simulation processes the decisions of the activities done in the action section. Results are presented as output.

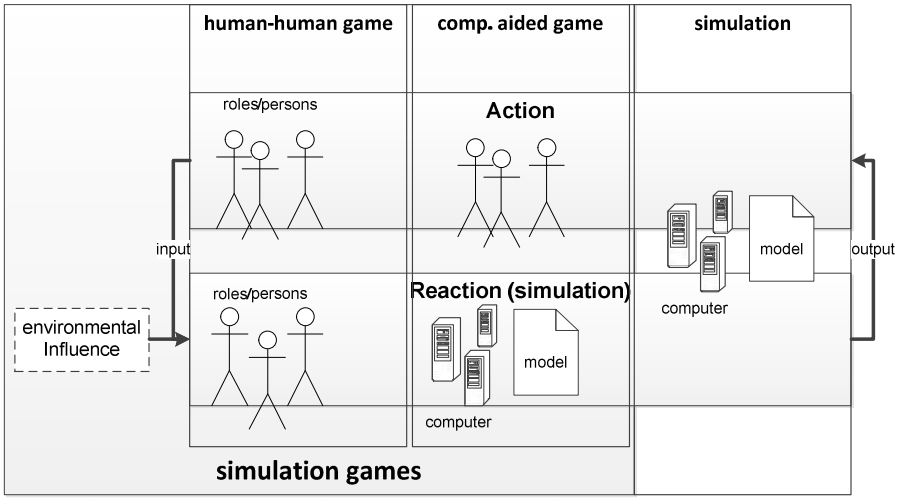


Fig. 1. Action and reaction section of simulation games [ref. to 13]

Fig. 1 illustrates action and reaction section of simulation games referred to Kern's work. Simulation games could be computer-aided designed or a human-human game where e. g. the game leader performs the output by processing the teams' decisions with the use of formulas or other given instruments. While serious games span both areas, simulations only fill out the reaction section [9]. Therefore, a simulation could be part of a simulation game but not the other way round. An additional system component representing extraneous environmental disturbance could influence the results of simulation games.

Rigid-Rule and Free-Form Games

Simulation games consists of two different gaming concepts: either the rigid-rule game (Ger.: geschlossenes Planspiel) or the free-form game (Ger.: offenes Planspiel).

The most common types of simulation games are the rigid-rule games. These games follow strict rules [17]. The participants will be set into a preconfigured situation. A set of rules and a game instruction exist. The task is to play the game in a most successful way.

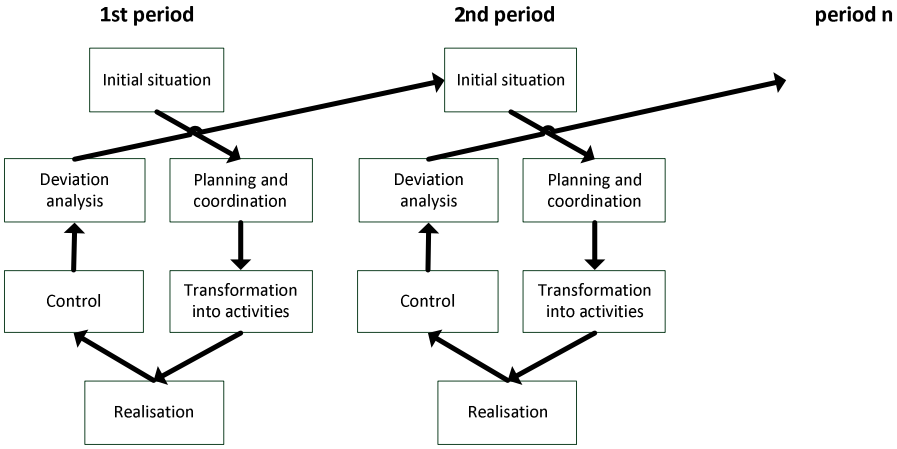


Fig. 2. Typical management cycle of ridge-based games [ref. to 17]

Due to the strict set of rules, the gaming process consists of the same phases during each round. Fig. 2 shows a typical management process cycle of rigid-rule games. Every step repeats in every new business period.

In comparison to that, free-form games deliver an organisation to produce or evaluate a simulation. Instructions are widely avoided [17]. Although free-form games have rules how they are created, they do not offer a predefined system of rules which restricts the action opportunities of the participants. They will be defined during the gaming process or during the development process of the simulation. Nevertheless, such games also follow a process structure but the problems are not predetermined by the game leader [17].

3 Summary and Future Work

It is found out that among the prevalent terms serious games and simulation games, a general approved definition still does not exists. In some cases, it is also difficult to delimit the terms.

While serious games are a sort of combination of entertainment and education (“edutainment”), simulation games mostly build on reality scenarios and strict processes to transfer knowledge.

Further research work could be concentrate of the construction of a methodological box in order to structure the high amount of serious and simulation games on the market. An interesting point of view would be – by modelling the box – to find out criteria, in which serious and simulation games have the same characteristics and in which points they differ from each other. Also a new approach in simulation games are system-integrated games, such as ERPsim[9] from HEC Montéal where the participants directly enter their decisions into the information system (here: SAP ECC 6.0) or have to get their results directly from the system. How far have teaching scenarios of classic simulation games[10] to be adapted for system-integrated simulation games?

References

1. Huizinga, J.: Homo Ludens. Vom Ursprung der Kultur im Spiel. Rowohlt, Reinbek, Hamburg (2004)
2. Marr, A.C.: Serious Games für die Informations- und Wissensvermittlung. Bibliotheken auf neuen Wegen. In: Fuhlrott, R., Krauß-Leichert, U., Schütte, C.-H. (eds.) B.I.T.online – Innovativ., vol. 28, Dinges & Frick, Wiesbaden (2010)
3. Bremer, T., Busch, C.: SpielZeit. Meilensteine der Spielentwicklung, ein Abriss. In: Sieck, J., Herzog, M.A. (eds.) Kultur und Informatik: Serious Games, pp. 8–9. Verlag Werner Hülsbusch, Boizenburg (2009)
4. Salen, K., Zimmerman, E.: Rules of Play. Game Design Fundamentals, p. 90. MIT Press (2003)
5. Göbel, S.: Über Serious Games. Serious Game Conference (2012), http://www.nordmedia.de/content/digitale_medien/digital_media_cluster/serious_games_conference/netzwerk/ueber_serious_games/index.html (March 25, 2012)
6. Abt, C.C.: Serious Games. Viking Press, New York (1970)
7. Sawyer, B.: The "Serious Games" Landscape. Digital Mill (2007), http://internet2.rutgers.edu/pres/speaker6-sawyer-final.ppt (March 25, 2012)
8. Hawlitschek, A.: Spielend lernen in der Schule? Ein Serious Game für den Geschichtsunterricht (2009), http://www.spielbar.de/neu/wp-content/uploads/2009/06/hawlitschek_spielend_lernen.pdf (March 25, 2012)
9. Prensky, M.: Digital Game-Based Learning. McGraw-Hill Inc., New York (2001)
10. Lampert, C., Schwinge, C., Tolks, D.: Der gespielte Ernst des Lebens: Bestandsaufnahme und Potenziale von Serious Games (for Health). In: MedienPädagogik – Zeitschrift für Theorie und Praxis der Medienbildung, Nr. 15/16 (2009), http://www.medienpaed.com/15/lampert0903.pdf (March 25, 2012)
11. Geilhardt, T., Mühlbradt, T.: Planspiele im Personal und Organisationsmanagement. Hogrefe, Göttingen (1995)
12. Kritz, W.-C.: Die Planspielmethode als Lernumgebung. In: Mandl, H., Keller, C., Reiserer, M., Geier, G. (eds.) Planspiele im Internet – Konzepte und Praxisbeispiele für den Einsatz in Aus- und Weiterbildung, wbv, Bielefeld, vol. 26, pp. 78–95 (2001)

[9] http://www.erpsim.ca/

[10] E. g. X enterprises can produce x products and deliver x markets.

13. Konstantinitis, C., Kienegger, H., Wittges, H., Krcmar, H.: Planspiele in der ERP-Lehre. Eine empirische Untersuchung deutscher Bildungseinrichtungen. In: Schumann, M., Kolbe, L.M., Breitner, M.H., Frerichs, A. (eds.) Multikonferenz Wirtschaftsinformatik 2010, pp. 1709–1721. Universitätsverlag Göttingen, Göttingen (2010) ISBN: 978-3-941875-31-9
14. Adelsberger, H.H., Bick, M.H., Kraus, U., Pawlowski, J.M.: A simulation game approach for efficient education in enterprise resource planning systems. In: European Simulation Multiconference, Warsaw (1999)
15. Léger, P.-M.: Using a simulation game approach to teach enterprise resource planning concepts. Journal of Information Systems Education 17(4), 441–447 (2007)
16. Kern, M.: Planspiele im Internet – Netzbasierte Lernarrangements zur Vermittlung betriebswirtschaftlicher Komponenten. DUV, Wiesbaden (2003)
17. Gust, M., Klabbers, J.H.G.: Unterschiede zwischen offenen und geschlossenen Planspielen. In: Blötz, U., et al. (eds.) Planspiele in der beruflichen Bildung. Auswahl, Konzepte, Lernarrangements, Erfahrungen. Aktueller Planspielkatalog 2005, pp. 58–67. W. Bertelsmann, Bielefeld (2005)
18. Breuer, J., Bente, G.: Why so serious? On the Relation of Serious Games and Learning. Eludamos. Journal of Computer Game Culture 4(1), 7–24 (2004)
19. Morares, R.M.: de: Serious Games and Virtual Reality for Education, Training and Health. In: Cruz-Cunha, M.M. (ed.) Handbook on Research on Serious Games as Educational, Business and Research Tools, pp. 315–336. IGI Global, Hershey (2012)
20. Riedel, J.C.K.H., Hauge, J.B.: Evaluation of Simulation Games for Teaching Production (Engineering). In: Cruz-Cunha, M.M., Carvalho, V.H., Tavares, P. (eds.) Computer Games as Educational and Management Tools – Uses and Approaches, pp. 263–279. Information Science Reference, Hershey (2011)

Repository-Based ERP Case Studies: A Study about Chances and Benefits of Agile Case Study Development

Hans-Jürgen Scheruhn[1], Sandro Sicorello[1], and Stefan Weidner[2]

[1] Business Informatics
Hochschule Harz, Friedrichstr. 57-59,
38855 Wernigerode, Germany
hscheruhn@hs-harz.de, sandro.sicorello@googlemail.com
[2] SAP UCC Magdeburg | School of Computer Science
Otto von Guericke University Magdeburg
Universitätsplatz 2, 39106 Magdeburg, Germany
weidner@ucc.uni-magdeburg.de

Abstract. ERP case studies have an important impact on the transfer of knowledge between software vendor, lecturer and user. This paper describes to which extent various ERP case studies are comparable. In the light of this paper, business processes with focus on procurement represented in different ERP systems and datasets are considered, e.g. Global Bike Inc., ALMIKA Holding and IDES AG. The study is based on information models derived from the GBI model repository created with the ARIS Business Designer by Software AG and synchronized with the SAP Solution Manager repository.

Keywords: ERP case study, ARIS Business Designer, Business Process Management life cycle, SAP Solution Manager, Agility.

1 Problem Identification

Companies must adapt their ERP systems to changing circumstances as well as their products [1]. ERP case studies are used for the transfer of knowledge in both the academic and the non-academic sector [20]. They are considered in different publications. ERP case studies deal either with the system development, process execution or with the implementation and the handling of ERP systems [5; 8; 9; 10; 11; 22; 27; 34]. Leyh concludes the need of more than one ERP system for teaching at universities and Winkelmann et al. describe the comparing usage of two different ERP systems respectively ERP case studies. Hufgard et al. compare a B2B cloud teaching scenario of SAP Business ByDesign and SAP ERP [7].

A study about the similarities of different ERP case studies with the intention to create a common framework as a repository for structured transfer of knowledge and more efficient development and maintenance of ERP case studies (e.g. release management of SAP Business ByDesign from rel. 2.6. to rel. 3.5) is yet missing. These objectives are explored in this paper.

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 186–197, 2012.

2 Objective of a Solution

ERP case studies support students learning the handling of different functional areas in a practical way [31]. Several ERP case studies exist, for example Global Bike Inc. [33], SSB Inc. [10], SAP TERP10 exercises at IDES AG [31], SAP Business ByDesign´s ALMIKA Holding [8; 17; 18; 19] and OPM´s IDES AG [22] which provide an introduction to different ERP systems of SAP. This paper aims to investigate the information model repository-based teaching and design respectively maintenance of ERP case studies in approach to further benefits. One target is to teach students educational approaches of business administration and/ or business informatics due to a structured integration of transfer of knowledge based on information models.

Similarities of different ERP case studies need to be analyzed and synchronized in a common ARIS/ SAP Solution Manager [23] core model repository as a basis for a possible implementation and maintenance of the entire process life cycle.

The paper compares information models based on an evaluation matrix created by the authors which can be used as a framework for further case study comparisons (e.g. SSB Inc.) or case study developments (e.g. GBI to-be-status) as well [24]. The integration of aspects of business and information technologies, of strategic, tactical and operational points of views as well as the integration of "user roles", "process and data flow" up to the ERP implementation in the context of the entire case study process life cycle are also possible [24].

3 Design and Development

3.1 Constructs

The term *agility* describes the characteristics to proceed flexibly, actively, adaptively and initiatively in times of changes and uncertainty [36].

Classical and agile software development is implementing solutions in reaction to changes. Agile software development focuses on short-term planning in order to better react to occurring events and understands change as a chance. On the other hand the classical software development defines change as a risk. This implies that agility is not just accepting the change; it is approving it [29].

Five central targets of agility are considered in literature:

(1) Speed: Quick deployment of the solution
(2) Flexibility: Configuration based on change requirements
(3) Proactive innovation: Change needs to be supported actively not only to react to it
(4) Quality: Increase performance, structure, process quality and output
(5) Profitability: Focus on economic targets [35]

Software developers can increase individualization of a new product. Furthermore, the configuration for specific customer needs can be done faster, which makes implementation projects less expensive [3]. However, the quantitative measurement

of agility and the transformation into monetary resources is difficult to achieve for the user. Agility is criticized because of that [3]. *Reference models* should provide support for construction processes of an individual enterprise model [4; 32]. In order to ensure a consistent model creation principles of construction are necessary. This enables an efficient model construction [14].

The information models developed and evaluated in this paper are considered as software reference models which depict typical company processes of the engineering industry based on software, e.g. ERP systems. These software reference models together with their model type and object type instances uniquely refer to one of the three model companies considered and are called *enterprise models* in this investigation.

According to Houy, the *Business Process Management (BPM)* life cycle encompasses six phases [6]. These correspond with the five phases of "Online Process Management" by Scheruhn [22]. Houy's sixth phase "optimization" complies with the phase "strategic planning" initiating the second loop by Scheruhn.

Phases of process life cycle by Houy	**Phases of process life cycle by Scheruhn**
1.Strategy development	1. Strategic planning
2.Definition and modeling	2. Process design
3. Implementation	3. Implementation
4. Execution	4. Process execution
5. Monitoring and controlling	5. Process controlling (monitoring)
6. Optimization and improvement	1. Strategic planning

Fig. 1. Business Process Management life cycle [6; 22]

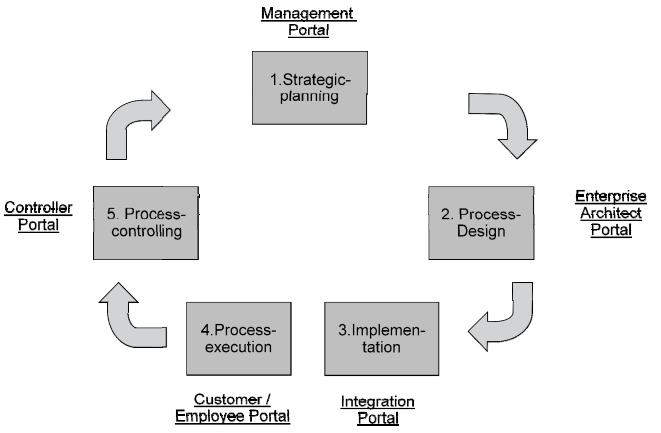


Fig. 2. Process life cycle with five phases by Scheruhn [22]

The *information pyramid* [12], in the context of the evaluation matrix, is divided into three levels. The tactical level connects the strategic and the operational level. The considered ERP case studies mainly cover the operational level in the context of the horizontal integration. Regarding possible future case study development (to-be-status) the tactical and strategic level on top also need to be considered.

3.2 Four IT Integration Layers

In this research business processes of all case studies are regarded as sub models of a model system, which includes

- Subject-oriented modeling using process-specific role assignments [26] based on descriptive case studies,
- process-oriented modeling,
- service-oriented modeling and
- Enterprise data modeling [6].

This leads to four IT integration layers (1.Presentation, 2.Processes, 3.Functions and 4.Data) in the enterprise model (see Fig. 3) describing an IT system according to user interaction in one specific case study (subject-adverb-predicate-object). At the same time it includes the four typical components of an ERP system and also corresponds to four views of the ARIS House of Business [21]. A vertical decomposition of each of the four IT integration layers leads to a model hierarchy of four levels shown in figure 3 [23]. Hierarchical levels can be mapped with the four process levels of the SCOR reference model [30] as well as partly with the three levels of the SAP Solution Manager Business Blueprint Architecture [15]. For the evaluation matrix, the information models including all model objects are considered on the type as well as on the instance level (see 3.4).

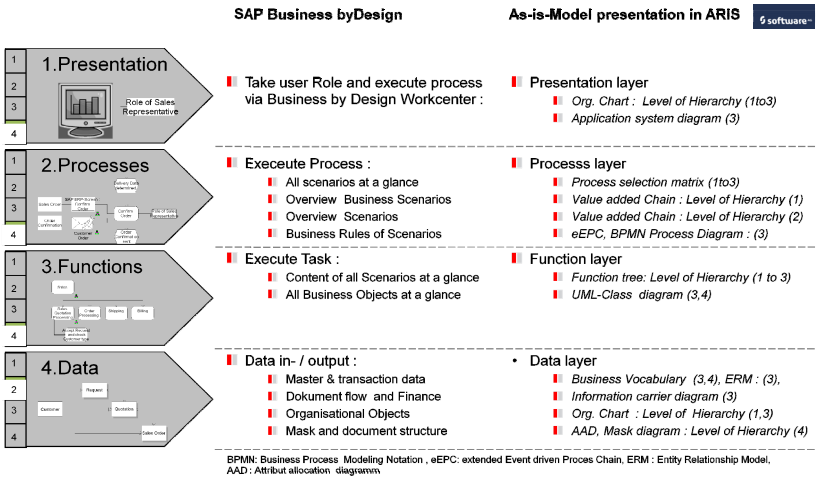


Fig. 3. ALMIKA enterprise model with four IT integration layers and four hierarchical levels [23]

3.3 Methods

First a hypothesis is created which has to be reviewed in the course of this investigation [13] based on the enterprise models. These models are created with ARIS Business Designer by Software AG [28] derived from the GBI model repository as an extended and synchronized subset of the SAP Solution Manager Business Blueprint repository [15].

Hypothesis: The above-mentioned case studies are similar to each other and according to this they can be implemented in a common model repository. Thereby, learners find similar structures and focus more on the ERP learning process and/or company-specific content.

Thus, it has to be proven to which extent the three model companies

- work with the same employee roles and employee competencies,
- perform the same core, overview and detailed functions (tasks) and business rules,
- communicate to each other using the same "business vocabulary",
- access the same (paper-based) document types and data represented by them,
- work on the same "masks" (screens) – structures as well as primary keys and foreign keys for data input and output,
- post the same incoming and outgoing payment types,
- work with the same system structures,
- pursue the same visions, strategic targets and KPI's (to-be-status),
- can use SAP Business ByDesign (rel. 2.6 or 3.5) as well as SAP ERP or other ERP systems due to changes planned in the future (merger and acquisition and/ or joint venture).

3.4 Instantiation

The investigation elaborates the integration of three different enterprise models; for the Global Bike Inc. scenario documentation, the SAP TERP10 exercises in SAP ERP, and the ALMIKA Holding case study in SAP Business ByDesign.

The model companies are situated in the manufacturing industry and have different sizes and structures. Global Bike Inc., for example, is a multinational enterprise producing and distributing bicycles. In contrast, ALIMKA Holding is an independent mid-sized company which is established in Germany and distributes heating systems as well as services. IDES AG [16] is the model company used in all SAP Education courses for end-user training. It is an international group with subsidiaries in several countries and manufactures pumps and other products.

The three companies use different ERP systems. Global Bike Inc. has implemented SAP ERP and ALMIKA Holding works with SAP´s OnDemand solution for mid-sized companies – SAP Business ByDesign. IDES AG - is working with SAP ERP.

4 Demonstration

This section presents results of the first implementation of above-mentioned ideas to construct an enterprise model repository based on the GBI model company. The SAP University Alliances curriculum "Introduction to SAP ERP using GBI 2.1" currently consists of eleven chapters:

- Introduction to SAP
- Navigation in SAP ERP
- Introduction to Global Bike Inc. (GBI)
- Sales and Distribution
- Materials Management
- Production Planning and Execution
- Financial Accounting
- Managerial Accounting
- Human Capital Management
- Warehouse Management
- Project Management

In the initial phase of this evaluation the project focused on the procurement process. In addition to the Materials Management chapter including presentations, exercises and a case study, the Introduction to GBI material was analyzed.

First, from the detailed description of the global company, an organizational chart was developed which displays the complete organizational hierarchy from the global enterprise structure to department level down to individual positions and employees' names.

Second, the Materials Management case study was decomposed into its learning steps (individual tasks within the SAP ERP system). These tasks were replicated in function trees in the enterprise model using the same task names as in the case study. These functions representing ERP transactions were recomposed to value-added chains that give the student not only the possibility to display the chronological order of the process but also facilitates horizontal and vertical navigation through the model. Students can drill into individual process steps modelled in extended event-driven process chains (eEPC). All eEPC's of the procurement process are strung together so that a user can follow the process in the model in parallel to the case study tasks in the ERP system.

Third, for advanced users the model was enhanced with Entity Relationship Models (ERM), Screen Diagrams and a Business Vocabulary. All models are accessible online for all SAP University Alliances members. Direct feedback shows that the enterprise model is used in classrooms around the world. In order to get more detailed end-user feedback a global SAP UA questionnaire is planned on the GBI curriculum including the enterprise model.

5 The Evaluation Matrix

A comparison of enterprise models considered in this paper shows much similarity in the model structure: four IT integration layers, the operational level of the information pyramid as well as the second and fourth phase of the process life cycle. The Global Bike Inc. enterprise model synchronized with the SAP Solution Manager repository additionally supports the third and fifth phase of the process life cycle. However, only three of four hierarchical levels (document level is covered but not made available yet) are implemented in the SAP Solution Manager. The first phase of the process life cycle determines the possible future repository based to-be-status of the model [24].

About 50% of the model types respectively 75% of the object types of the GBI starting model in ARIS (first column in the evaluation matrix) exist in the SAP Solution Manager repository.

Enterprise models		**GBI [ARIS Repository]**	**GBI [SAP SolManager Repository]**	**ALMIKA [strategic procurement/ ARIS]**	**IDES [TERP10/ ARIS]**
Model structure	Operational level of information pyramid[1]	●	●	●	●
	Process life cycle phases by Scheruhn	2,4	2-5	2,4	2,4
	4 IT-integration layers	●	●	●	●
	Levels of hierarchy	4	3	4	4
Model types	Shown in Fig. 3	●	◑	●	●
Object types	Ref. to: Process Governance Matrix [25]	●	◕	●	●
Model type instances/ 1.Presentation layer	Org chart	●	◕	○	○
2.Process layer	Value chain at core business level	●	●	○	○
	Value chain at overview level	●	●	○	◕
	Business rules as eEPCs as well as BPMN process diagrams	●	◑	○	●
3.Function layer	Function tree, UML class diagram	●	◕	◔	●

Fig. 4. Evaluation matrix

[1] These models mainly exist on the operational "third" level. Possible to-be-models are dealing also with the two levels above because of the need of strategic ideas.

4.Data layer	Business vocabulary	●	○	●	●
	Data models	●	◑	●	●
	Document and financial flow	●	○	●	●
	System organization	●	◑	◑	●
	Mask and document structure	●	◑	○	●
Object type instances/ 1.Presentation layer	Roles	●	●	●	●
2.Process layer	Core process step	●	●	○	○
	Overview process step	●	●	○	◕
3.Function layer	Process step	●	●	◔	●
4.Data layer	Master and transaction data	●	●	●	●
	Documents and forms	●	◔	●	●
	System organizational units	●	◕	○	◕
	Keys, foreign keys, describing attributes	●	◑	◕	●
Agility of Development	Speed	●	●	◑	◕
	Flexibility	●	◑	●	●
	Quality	●	●	◕	◕

Fig. 4. (*Continued*)

The model and object types of the ALMIKA Holding in ARIS entirely comply with the GBI starting model but additionally include the model type "Process Selection Matrix rel. 2.6 vs. 3.5" due to the scenario structure of SAP Business ByDesign. Furthermore, additional object types, especially "Scenario", "Work Center" and so called "Step" within the windows (GBI mask) are covered. The IDES AG/TERP10 model and object types encompass the entire content of the Global Bike Inc. enterprise model in ARIS.

Not surprisingly, the model type instances differ to some degree. Three individual org charts describe the three companies. Compared to GBI, similarity of the ALMIKA Holding enterprise model on the process layer is very low because of the usage of different terms for same processes (e.g. "Procurement" vs. "Materials Management", because of a restricted range of functions (e.g. "Controlling") on the core business level and because of an additional scenario structure (e.g. "Strategic sourcing") on the overview level as well as different masks and windows (e.g. "Sourcing") on the detailed level. Because of different chapter structures on the core business level (e.g. "chapter 5: business process purchase to pay") the IDES AG/ TERP10 model is different from the Global Bike Inc. starting model. Additional terms for the same subchapters on the overview level (e.g. "exercise 21/1 create

material") as well as a small superset and subset of processes on the detailed level appear as almost identical model type instances.

"Business rules" which are depicted on the detailed level as eEPC's –or as Business Process Modelling Notation (BPMN) models cannot be depicted completely in the SAP Solution Manager Repository because of missing process splits and joins.

The UML class diagram on the function layer is not available in the SAP Solution Manager repository directly but can be connected to corresponding SAP ERP System business objects with their methods and attributes. However, due to different terms used in the description of the ALMIKA Holding on the detailed level only few similarities respectively same functions exist (e.g. "create vendor" or "create material").

Small differences on the data layer between the ALMIKA Holding and Global Bike Inc. ARIS model are based on an incomplete mapping of ALMIKA Holding organizational objects with the Global Bike Inc. system organizational units respectively on the sole existence of "Work Center" in the ALMIKA Holding (e.g. "Work Center: sourcing and contracting").

The remaining model type instances correspond with the GBI source model respectively include a common extension of the business vocabulary (e.g. "accounts payable number"="supplier number"), of the data models (e.g. "sourcing request") and document flow (e.g. "sourcing"). The SAP Solution Manager repository is only capable of depicting a smaller amount of the model type instances of the GBI ARIS starting model on the data layer. The mask structure documented as a transaction by request is automatically connected to the corresponding transaction of an existing SAP ERP System with its input and output fields. These fields can also link into the data type and attribute type instances of the SAP ERP data model (see object type instances).

A comparison of the object type instances reflects the high similarity between the different models. On the presentation layer same roles indicate same business competencies of the employees working for these three different companies. Differences on the process and functional layer are caused by similar reasons affecting also the model type instances. Master and transactional data as well as documents and forms are almost similar in all of the three companies. The differences regarding the system organizational units can be traced back to the introduction of so called organizational objects of Business ByDesign (e.g. "company" or "functional unit") in the ALMIKA Holding respectively to other locations (e.g. "company code 1000" or "plant 1000") in the IDES AG. Key and foreign key attributes are identical to a high extent of all three companies considered (e.g. "accounts payable number" or "material number"). The system organizational units documented in the SAP Solution Manager are semi-automatically connected to the corresponding configuration transaction of an existing SAP ERP system for displaying locations (e.g. "company code US00") and system structures (e.g. "Sales Area GBI"). Finally, differences are caused by the describing attributes due to other field names (e.g. "target value of the sourcing") or structures.

The evaluation matrix also reflects to which extend three sub targets of agility are obtained. Assuming that the design and development of the GBI model repository is a

"proactive innovation" (100%) the remaining sub targets "speed", "flexibility" and "quality" are achieved at different levels. The "profitability" is to be evaluated in further investigations. Once the scenario structure of ALMIKA is mapped to the GBI model repository it takes only one day to adapt to a new release. The synchronisation with the Solution Manager only takes a few minutes if its model structure has been created carefully in advance. The enterprise models of the three different companies are created at high flexibility except of the Solution Manger due to significant limits of its model and object types structure. Finally, the formal quality of all three enterprise models is high because of the strict compliance with the model conventions based up to 75% on the SAP Solution Manager repository.

6 Conclusions

The user can work with similar employee roles and competences during the execution of the case studies. Additionally, he accesses similar core and overview processes as well as detailed functions. The business vocabulary helps communicate similar documents or data objects. Data input and output can be made with similar screens. The external financial flow can be represented by similar ingoing and outgoing payment types processed within the same system structures.

The Global Bike Inc. model repository can not only be used for the agile development of the ALMIKA Holding information model but also for TERP10. In additional investigations the synchronized subset of the GBI SAP Solution Manager repository has to be tested for possible extensions of the case studies to all phases of the entire Business Process Management life cycle.

Model structures and object types are similar to a large extent; however in particular the similarities of the process terms are low. This circumstance possibly impedes the students learning of facts and coherences.

Due to the high similarity of the considered case study models one common case study could be applicable on different ERP systems in the future. This could lead to both a structured transfer of knowledge - and to a simplified maintenance of the case study which are to be evaluated in further investigations. An agile derivation of new case studies, of new releases of case studies like SAP Business by Design rel. 3.5 or of a possible to-be-status (e.g. GBI) from the repository of an existing case study could happen fast and flexibly at a high quality.

Further investigations need to prove whether the creation of the ERP case study documentation and controlling used for the teaching of ERP case studies can be created model-driven as well.

7 Summary

In this paper three different ERP case studies were compared with each other regarding the structures, model types and object types as well as the model type and object type instances of the corresponding enterprise models.

The comparison based on the GBI model repository results in many similarities which are worked out structurally in an evaluation matrix. On the other hand, the findings provide differences which result from the fact that different terms of different case study descriptions are used for describing the same content. Thus, learning is possibly more difficult or less efficient for the users.

A resulting advantage from this could be on the one hand a decrease of the effort for the creation and maintenance of ERP case studies. On the other hand it could be possible to improve the integration of the transfer of knowledge in business administration and/or business informatics.

References

1. Baggio, M.: ERP-System ohne Ablaufdatum in. In: Gronau, N., Lämmer, A., Andresen, K. (eds.) Wandlungsfähige ERP-Systeme-Entwicklung Auswahl und Methoden, pp. 151–162. GITO-Verlag, Berlin (2010)
2. Becker, J., Rosemann, M., Schütte, R.: Entwicklungsstand und Entwicklungsperspektiven der Referenzmodellierung (1998)
3. Becker, A.: Nutzenpotentiale und Herausforderungen Serviceorientierter Architekturen. Gabler Verlag, Wiesbaden (2011)
4. Fettke, P., Loos, P.: Der Referenzmodellkatalog als Instrument des Wissensmanagement - Methodik und Anwendung. In: Becker, J., Knackstedt, R. (eds.) Wissensmanagement mit Referenzmodellen: Konzepte für die Anwendungssystem und Organisationsgestaltung, pp. 3–23. Physica-Verlag, Berlin (2002)
5. Friedrichmeier, H., Mair, M.: Fallstudien: Entwicklung und Einsatz von Fallstudien - Erfahrung und Best-Practice-Beispiel. Linde Verlag, Wien (2007)
6. Houy, C., Fettke, P., Loos, P., van der Aalst, W., Krogstie, J.: Geschäftsprozessmanagement im Großen in Wirtschaftsinformatik 06/11. Gabler Verlag (2011)
7. Hufgard, A., Legner, C., Winkelmann, A.: B2B-Geschäftsszenarien mit der Cloud-Lösung SAP Business ByDesign. In: Proceedings of the Multi-Konferenz Wirtschaftsinformatik (2012a)
8. Hufgard, A., Krüger, S.: SAP Business by Design: Geschäftsprozesse, Technologie und Implementierung anschaulich erklärt: Galileo PRESS, Bonn (2012b)
9. Leyh, C.: ERP-System-Einsatz in der Lehre Ergebnisse einer Umfrage an deutschen Universitäten und Fachhochschulen. In: Proceeding of the Multi-Konferenz (2012)
10. Magal, S., Word, J.: Essentials of Business Processes and Information Systems. Wiley & Sons (2009)
11. Magal, S., Word, J.: Integrated Business Processes with ERP Systems. Wiley & Sons (2011)
12. Mertens, P., Bodendorf, F., König, W., Picot, A., Schumann, M.: Grundzüge der Wirtschaftsinformatik. Springer, Berlin (2005)
13. Peffers, K.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24(3), 45–77 (2007)
14. Rosemann, M.: Grundsätze ordnungsmäßiger Modellierung: Intention, Entwicklung, Architektur und Multiperspektivität. In: Maicher, M., Scheruhn, H.-J. (eds.) Informationsmodellierung - Referenzmodelle und Werkzeuge, pp. 1–21. Deutscher Universitäts Verlag, Wiesbaden (1998)

15. SAP Bibliothek: Business Blueprint (2012a), http://help.sap.com/saphelp_sm32/helpdata/de/2a/62c33af63ae93ae10000000a11402f/content.htm
16. SAP Bibliothek: IDES – das SAP Modellunternehmen (2012b), http://help.sap.com/saphelp_46c/helpdata/de/af/fc4f35dfe82578e10000009b38f839/frameset.htm
17. SAP: Business ByDesign – Referenzsysteme - Szenario-Überblick Strategische Bezugsquellenfindung, Rel. 3.5 (2012a)
18. SAP: Business ByDesign: Unternehmensprofil für ALMIKA Vers.3.5 (2012b)
19. SAP: Strategische Bezugsquellenfindung - Referenzsystem für SAP Business ByDesign, Rel. 3.5 (2012c)
20. SAP: TERP10 (2012d), http://www.sap.com/germany/services/education/globaltabbedcourse.epx?context=[[|terp10|||||]]|
21. Scheer, A.-W.: ARIS-Modellierungsmethoden, Metamodelle, Anwendungen. Springer, Berlin (2001)
22. Scheruhn, H.-J.: Online-Prozessmanagement mit SAP ERP. In: Scheruhn, H.-J., Papenfuß, D., Funk, B., Niemeyer, P. (eds.) Modellierung und Implementierung von Geschäftsprozessen in verteilten Systemen, pp. 141–177. Dr. Kovac (2010)
23. Scheruhn, H.-J., Thielert, B., Sicorello, S.: ALMIKA-Unternehmensmodell/ Business Process Analysis (2012a), http://hscheruhn.hs-harz.de/lehre/alle/Vergleich_ALMIKA_GBI_auf_Basis_BPA1_de_final.pdf
24. Scheruhn, H.-J., Thielert, B., Weidner, S.: GBI-Unternehmensmodell/ Business Process Analysis/ Soll-Zustand
25. Scheruhn, H.-J.: Model Governance Matrix: ALMIKA (2012c), http://hscheruhn.hs-harz.de/lehre/alle/Modell-Matrix_horizontale_Navigation.pdf
26. Schmidt, W., Fleischmann, A., Gilbert, O.: Subjektorientierter Ansatz in Prozessmanagement: Praxis der Wirtschaftsinformatik. Heft 266 4/09 (2009)
27. Schubert, P.: Williams. S.: Erwartungen und Nutzen beim Einsatz von Business Software - Empirische Ergebnisse einer Analyse von Fallstudien (2010)
28. Software AG: Software AG (2012), http://www.softwareag.com/de/
29. Stein: Emergenz in der Softwareentwicklung - bereits verwirklicht oder Chance? (2004)
30. Supply Chain Council: SCOR-Referenzmodell (2012), http://supply-chain.org/scor
31. TERP10: SAP ERP - Integration of Business Processes (2012), http://www.terp10.de/pages/deutsch/kurs.php
32. vom Brocke Referenzmodellierung - Gestaltung und Verteilung von Konstruktionsprozessen (2003)
33. Weidner, S., Magal, S., Word, J.: Global Bike Inc. Version 2.1 (2011)
34. Winkelmann, A., Heide, T., Holler, J., Püster, J.: Lernerfolg durch System-Vergleich – Einsatz von SAP Business ByDesign und SAP ERP in der Lehre. In: Proceedings of the Multi-Konferenz Wirtschaftsinformatik, Braunschweig (2012)
35. Winter, R., Aier, S., Fischer, B.: Management von Integrationsprojekten: Konzeptionelle Grundlagen und Fallstudien aus fachlicher und IT-Sicht (2009)
36. Wolfgang Martin Team: Business Intelligence trifft Business Process Management (2011)

Interactive Learning – Teaching IT Project Management Using an Explorative Role Play

Stefan Weidner

SAP UCC Magdeburg, School of Computer Science,
Otto von Guericke University Magdeburg
Universitätsplatz 2, 39106 Magdeburg, Germany
weidner@ucc.uni-magdeburg.de

Abstract. The awareness of companywide (as well as cross-company) process and data integration and the importance of application systems effect education and advanced training in a way that the concept of integration needs to be taught to students. Today, business applications have an increasing degree of complexity and functional integration. Whereas the design of Very Large Business Applications is a main part in business informatics lectures, their configuration – practiced especially in consulting projects – is not yet part of most curricula. This paper introduces an explorative role play that may be used to impart soft skills in communication, teamwork or consultancy as well as hard skills like IT project management, system analysis and configuration.

Keywords: Learning methods, teaching philosophy, interactive learning, project management, explorative role play.

1 Motivation

Business informatics students should, first, acquire *knowledge* about how information systems are constructed, how they work and why they work like this. On this basis, students need to gain *competence* on how to (re)construct and configure information systems [8].

Since the signing of the Bologna declaration in 1999, almost 50 European countries have been implementing standards for academic degrees and quality assurance. Especially the promotion of European citizens' employability and international competitiveness [2] has led to introduction of job-qualifying Bachelor and Master programs throughout Europe [9].

According to a study of the German Chamber of Industry and Commerce about expectations on university graduates in 2011, students are still lacking practical, social and communication *skills* [5].

The author argues in this paper that a combination of the teaching methods *project* and *role play* can support a variety of lectures imparting knowledge, competence, and/or skills.

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 198–207, 2012.

2 Curriculum Objectives

Being the operational head of the SAP University Competence Center at Otto von Guericke University Magdeburg and working closely together with SAP's University Alliances program has provided the author with multiple opportunities to meet with professors from a variety of disciplines and universities world-wide as well as with practitioners in IT and consultancy.

At the net's work 2007 conference, the author led a discussion on the current implementation progress of Bachelor and Master programs at German universities with representatives from academia, politics and industry. While government and universities were coping with the transformation of their degree programs and trying to implement the Bologna declaration, the industry criticized the modifications for not being profound and successful enough. Three perspectives (academic, educational, and professional) on the same observation object (the student) revealed surprising assessments on the skill level of current university graduates. In addition to the lack of soft skills, consultancy firms pointed out the students' limited **understanding of systems integration and configuration in real-world scenarios**.

A report from the German Rectors' Conference [14] summarizing the reform progress as of February 2008 showed similar findings. More than 60 percent of all study programs at German universities and nearly 90 percent at universities of applied sciences had already been converted. The transition, however, was more organizational and structural than dedicated to a curriculum reform. In addition to *research-oriented knowledge*, university graduates need to gain *interdisciplinary key qualification*, *soft skills* and *academic personal development*. Zervakis specifies – among others – the following skills and activities [14].

Interdisciplinary key qualification

- internships with companies
- thesis papers on relevant issues
- business skills
- project management skills
- experience abroad
- foreign language skills

Soft skills

- communication skills
- team playing
- flexibility
- persistence

Academic personal development

- authenticity
- credibility
- character

The solution needed to meet above-mentioned objectives ("hard" and "soft" skills).

3 Theoretical Foundation

All times a society's quality of knowledge transfer has been one of the most important indicators for social capability and technical progress. The term education in this context stands for acquisition and development of knowledge and competence, skills and abilities [7].

The increasing demand for knowledge made it necessary and the invention of printing possible to reach readers time- and place-independently. Through technical progress such as the invention of radio, television, and the Internet more and more knowledge sources were accessible for the learner. However, these achievements would have been useless without parallel enhancements of teaching methods [13].

Today, knowledge is defined as a complex, net-worked and dynamic system, which structures we need to explore and understand in order to arrive at conclusions for structuring and organizing the learning process [.

Besides classic learning theories like behaviorism, cognitivism, and constructivism, modern learning approaches such as situated, case- and role-based learning become widely used in classrooms. To an increasing degree these techniques realize social aspects in learning. It is attempted to let the student experience the learning matter in most realistic and complex situations in a complicated, interrelated context. Instead of and additional to theoretical factual knowledge, rather individual and multiapplicable problem solving strategies should be trained the student. Social contact and interactive teamwork come to the fore (Bruner 1996, p. 167 et seq. and 173 et sqq.).

Learning targets, content and methods are essential factors of the educational system and, thus, are organizational and executive forms of learning and teaching (Döring 1971, p. 94).

Baumgartner and Payr developed a more complex learning model with similar dimensions: learning targets per learning type, content of teaching and teaching strategies. Their heuristic cube model facilitates an enormous configuration scope from novice to expert (Baumgartner et al. 1999, p. 95 et sqq.).

In today's curricula learning targets are usually dimensioned in cognitive, affective and psychosomatic (Pätzold 1996, p. 31). Due to the fact that most modern teaching methods focus on thinking, understanding and problem solving, the following cognitive learning targets of Bloom's taxonomy will be used: knowledge, comprehension, application, analysis, synthesis, and evaluation (Bloom 1976).

Figure 1 links Baumgartner/Payr's cube model with Bloom's cognitive learning targets.

In the cube model above the learning path can be represented by the space diagonal P1P2 or any space curve between these two points. Accordingly, factual knowledge is explained to novices whereas advanced learners and experts in complex learning situations are merely advised by a coach. All following conclusions refer to this learning path.

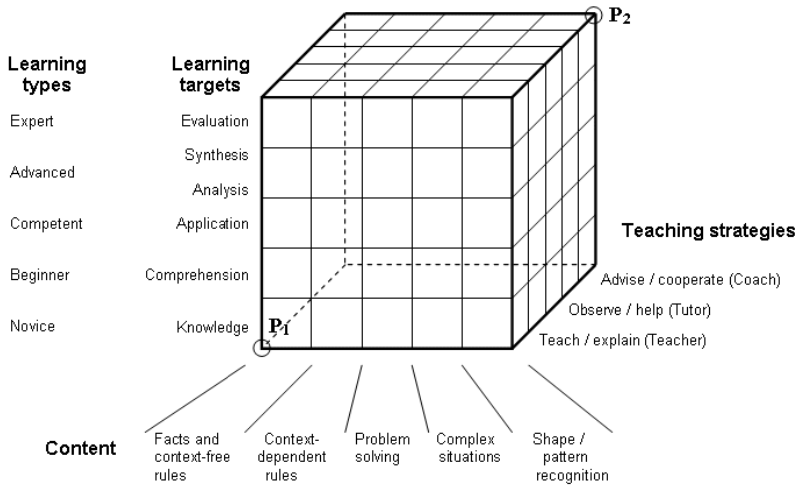


Fig. 1. Heuristic learning model (adapted from Baumgartner et al. 1999, p. 95 et sqq.)

Common competence fields can be added to the learning target/learning type dimension in the enhanced cube model (see figure 2).

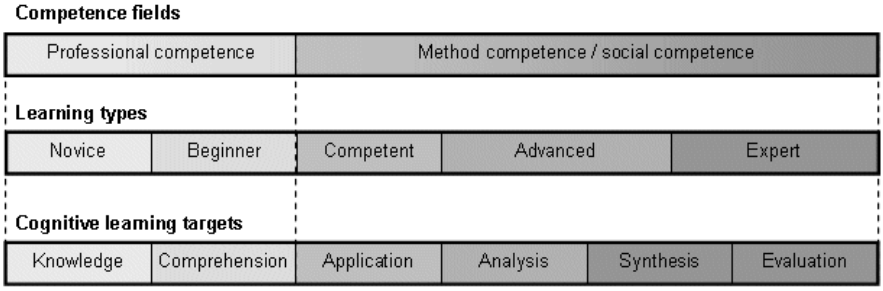


Fig. 2. Competence fields – learning targets (Weidner 2003, p. 15)

Furthermore, the following combinations of cognitive learning targets and teaching content result from the learning path from P1 to P2.

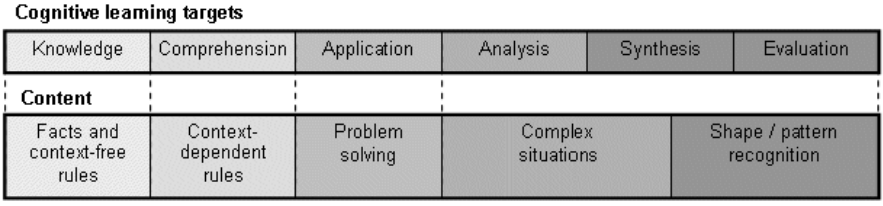


Fig. 3. Learning targets – content (Weidner 2003, p. 16)

Finally, specific teaching methods can be mapped to combinations of learning targets and content.

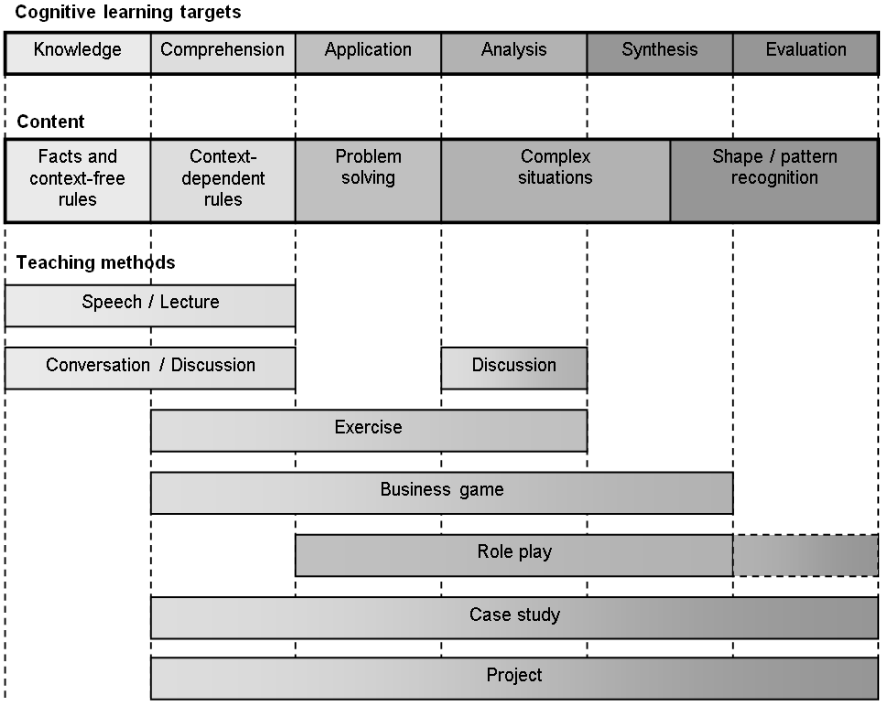


Fig. 4. Targets – content – methods (Weidner 2003, p. 17)

According to the learning path in figure 1, above-mentioned teaching methods reflect possible teaching strategies. While lectures explain, case studies are handled in cooperation with the learner. Cognitivistic exercises in higher education teaching – mostly advised by a tutor – provide an opportunity for students to recapitulate acquired theoretical knowledge and to apply scientific methods and cognitions. Role plays and projects, in addition to the above, emphasize working in a team on real or realistic problems.

4 Design and Demonstration of Interactive Role Play

Given that business applications exist in and are constructed for reality, business informatics is an applied science with practical reference. A key requirement for business informatics education and their teaching methods is the practical focus, but based on theory, thus, a theory-driven practical orientation (Heinrich 1993, p. 47 et seq.).

Following political directives as well as industry demand for practical, social and communication skills – in addition and in parallel to knowledge and

competence – learning methods such as business games, role plays, projects, case studies etc. have been widely adopted in higher education teaching (Leyh 2012, p. 6).

For business informatics graduates, IT consultant is a viable career option. However, most lectures – although attempting to provide practical cases and interactive learning environments – tend to fall short in creating authentic situations and initiating interactive engagement, or even exploratory spirit.

At the net's work conference 2007, industry experts presented how they support universities in imparting professional know-how and social skills. While the SAP University Alliances Program focuses on a two-level train-the-trainer approach to equip lecturers with the IT and didactic skill set for practical courses, the consulting firm itelligence sends consultants to university lectures in order to give students a realistic presentation of competences and skills needed in consulting projects. Both approaches achieve the objective of introducing some of the above-mentioned skills and qualifications. However, based on industry experience they are neither well scalable nor do they engage students measurably better than traditional approaches.

From the academic standpoint taken by the author in the panel discussion teaching should not be occupied by training or even vocational demands to an extent that academic programs or large parts of them become industry-driven trainee programs[1].

The project team in 2007 attempted to solve the problem of lacking skills with a far different and unusual approach. Brainstorming led to the idea of an *interactive* and *explorative role play* performed live on stage and recorded for later use in classes.

The role play was planned to be performed on stage at SAP's University Alliances User Group Meeting 2007 in St.Leon-Rot. In four months, a group of researchers at Otto von Guericke University Magdeburg and itelligence consultants developed the initial idea further.

The role play was created in German language around an IT implementation project at a fictitious medium-size company, Motorsport AG. itelligence consultants representing nine different roles acted in nine typical project situations.

Roles (consulting company)

- sales person
- project manager
- junior consultant
- senior consultant

Roles (customer)

- Chief Executive Officer
- Chief Financial Officer
- head IT department / project manager
- key user (logistics)
- logistics specialist

[1] Today, such ideas are considered and implemented in dual systems of vocational education and academic studies in German higher education.

Scenes

1	First customer meeting
2a	Internal preparation
2b	External preparation
3	Kick-off meeting
4-1	Analysis workshop - Take 1
4-2	Analysis workshop - Take 2
5	Prototyping
6	Steering committee
7	Go-live

Researchers at VLBA Lab Magdeburg drafted the storyboards including background information, scene description, didactic intention and dialogues as well as setting and seating plan. These initial drafts were sent to itelligence consultants and to lecturers for review. Although already rich in detail, practitioners enhanced the story based on their experience.

The final scripts illustrating an ideal IT implementation project were sent to the author for approval. Having experimented with didactic mistakes in other teaching methods such as case studies (Weidner 2003, p. 39 et sqq.) and business games (Rautenstrauch et al. 2001), the author decided to modify the stories by incorporating both obvious and hidden mistakes.

The video footage taken by two cameras during the performance was cut and enriched with images of presentations and demo systems used in the play. Twelve German educational institutions of different academia types like research universities, universities of applied sciences, and vocational schools were selected. Questionnaires for both lecturers and students were sent out together with the video DVD. During winter semester 2007/2008, lecturers exposed the video scenes to their students and collected feedback. At CeBIT 2008, all student groups presented their results in form of curriculum material and feedback.

When the final material was released to the SAP University Alliances community at the beginning of summer semester 2008, the learning environment comprised the videos, role descriptions, lists of mistakes as well as case studies and additional background material.

5 Evaluation and Lessons Learned

Feedback collected at the SAP UA User Group Meeting 2008 showed that more than 20 universities had used the role play during the summer term. Surprisingly many (40 percent) used the videos in social skill seminars. Only one third took the material to enrich their project management lectures. Two lecturers stated that they had used the material in ERP configuration courses.

The author integrated the role play in an *IT project management* lecture at Otto von Guericke University Magdeburg and in an advanced *ERP configuration* course at University of Applied Sciences Magdeburg; both in Bachelor degree programs.

In the university course, the theoretical lecture on project management principles and methods in IT is interrupted three times for 90-minute sessions where the videos are shown to students and discussed together with the lecturer. As no lab time is assigned to this lecture, students work on the videos in class and at home. Groups of four to six students are asked to critically observe each actor's behavior and replay the scenes with improvements and additional material.

At the university of applied sciences, the author teaches a three-semester series of lectures on ERP systems. In the last lecture, students are taking the role of consultants implementing SAP ERP at a fictitious mid-size company. The role play was the ideal material to have the students explore a 'real case' and find mistakes before they were applying what they had learned in other classes and what they had discovered from replaying the itelligence videos in their project groups.

Self-assessments at both institutions from 2008 to 2011 before and after the introduction of the role play showed a significant increase in the students' **project management skills** (from 46% to 75% at university and 50% to 82% at university of applied sciences) and a better **understanding of roles in IT projects** (24% to 70% respectively 38% to 74%). In addition, students' evaluation on their **communication**, **teamwork** and **presentation skills** indicated measurable improvement – on average increasing by nearly forty per cent.

The role play has thus already proven to positively affect hard and soft skills. The full data set of student evaluations and assessments from 2002 until 2012 is currently being analyzed against Zervakis' skill specifications and will be published in a separate paper.

Four years after the release of the German role play, the material was implemented into curricula at more than 40 universities. In addition, SAP and itelligence are using the videos in their trainee programs and internal skill workshops. In German-spoken countries, the role play has proven to be a tool box that is not as precise as some lecturers from different disciplines would like, but due to this characteristic it is easy to adopt for various purposes and target groups. In most cases, it showed its interactive and explorative nature of design. Its wider use was limited by one defect: The role play was available in German language only.

6 Outlook

At various meetings in 2011, the idea of an English version of the role play developed. itelligence, SAP and the author took the opportunity and started an initiative together with itelligence The Netherlands.

The new project team wanted the replay to meet many more of Zervakis' needs for key qualifications and skills. International actors from the US, England, Germany, The Netherlands, France, and Belgium agreed on the challenge of a studio production of nine, very similar scenes.

In November 2011, nine itelligence consultants met in Eindhoven for rehearsal and taping of the new version. This edition was based on the new global model company in the SAP UA program, a mid-size bicycle manufacturer named Global Bike Inc.

The parallel implementation of SAP's OnDemand solution SAP Business ByDesign at foreign plants outside Germany and the US was another shift towards a more complex case.

After post production, on March 16, 2012, the new international role play was presented to lecturers during a premiere event in Eindhoven. Early in the process of collecting feedback and defining the next steps the new material has the potential to demonstrate **international experiences**, **foreign languages and dialects**, tensions in a global team and **authenticity** of near-realistic situations.

7 Conclusion

After creating and releasing an explorative role play in German language the applicability of the material has proven to be manifold. Assessments already analyzed have shown that many student skills were affected positively.

Depending on the teaching method and utilization mode chosen, students evaluated the improvement of hard skills (system analysis, implementation, configuration, project management), soft skills (communication, teamwork, consulting skills), and personal skills (authenticity, self-confidence) to be much higher in classes where the role play was applied compared to classes without the interactive material. In addition, participants evaluated the relevance of the lecture with the role play higher (around 80% compared to 26 to 71% for other lectures).

With the English version the project team is trying to address and affect – in addition to the ones above – other skills such as internationality and foreign languages/dialects. The author envisions this new material to be used to show cultural differences in project teams as well.

References

1. Baumgartner, P., Payr, S.: Lernen mit Software. Studien-Verlag, Innsbruck (1999)
2. Bloom, B.S.: Taxonomie von Lernzielen im kognitiven Bereich. Beltz, Weinheim et al. (1976)
3. Bruner, J.: The Culture of Education. Harvard University Press, Cambridge (1996)
4. DIHK: Erwartungen der Wirtschaft an Hochschulabsolventen. Kevin Heidenreich, Deutscher Industrie-und Handelskammertag e.V., Berlin (March 26, 2012), http://www.dihk.de/ressourcen/downloads/hochschulumfrage-2011
5. Döring, K.W.: Lehr- und Lernmittelforschung. Beltz u. Gelbert, Weinheim et al. (1971)
6. Gudjons, H.: Pädagogisches Grundwissen, 6th edn. Verlag Julius Klinkhardt, Bad Heilbronn (1999)
7. Heinrich, L.J.: Wirtschaftsinformatik – Einführung und Grundlegung. R. Oldenbourg Verlag, Munich (1993)
8. KMK: Kultusministerkonferenz. Qualifikationsrahmen für Deutsche Hochschulabschlüsse (March 30, 2012), http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2005/2005_04_21-Qualifikationsrahmen-HS-Abschluesse.pdf

9. Leyh, C.: ERP-System-Einsatz in der Lehre Ergebnisse einer Umfrage an deutschen Universitäten und Fachhochschulen. In: Proceeding of Multikonferenz Wirtschaftsinformatik 2012, Braunschweig (2012)
10. Pätzold, G.: Lehrmethoden in der beruflichen Bildung, 2nd, enhanced edition, Heidelberg (1996)
11. Rautenstrauch, C., Schrader, H.: iDECOR – Das Unternahmensplanspiel, Magdeburg (2001)
12. The Bologna Declaration of 19 June 1999 (March 26, 2012), http://www.ehea.info/Uploads/Declarations/bologna_declaration1.pdf
13. Weidner, S.: Fallstudiendesign mit Hilfe der Case Study Methode am Beispiel der Integration von Unternehmensstrukturen in ein ERP-System. Master thesis. Magdeburg. University Magdeburg Press (2003)
14. Zervakis, P.: Implementing Bologna in Germany: The universities in search for new business alliances. In: German Rectors' Conference, Education and Culture DG – Lifelong Learning Programme, Brussels (2008)

Ontology for Semantic Description of Enterprise Architectures

Liviu Gabriel Cretu

Alexandru Ioan Cuza University, B-dul Carol 1 nr.22, 700505, Iasi, Romania
liviugabriel.cretu@gmail.com

Abstract. The level of complexity involved in real world enterprise systems development is increasing nowadays above the usual tolerance thresholds with the proliferation of open Cloud API's. These come on top of the usual spaghetti effect of inter-dependencies generated by on-premises components and applications. Within this landscape the Enterprise Architecture (EA) approach may provide the answer to manage the complexity. However, EA is itself a complex process having to accommodate and inter-link multiple conceptual models, languages, people and disciplines. In this paper we develop a methodology to transform EA descriptions made in plain natural language into machine readable EA models using Semantic Web technologies. We do this first by choosing an EA framework, then developing the ontology for this framework and finally showing the usage of semantic annotations in transforming HTML pages into well defined machine-readable EA content.

Keywords: Enterprise Architecture, Semantic Web, semantic annotations, HTML Microdata.

1 Introduction

One approach in managing the complexity of enterprise systems is to define the Enterprise Architecture (EA). The Massachusetts Institute of Technology defines the enterprise architecture as the organizing logic for business processes and IT infrastructure reflecting the integration and standardization requirements of the company's operating model [15].

However, even by taking the EA approach, companies find themselves struggling to develop communication channels between all the stakeholders of the enterprise systems: business owners and users, business analysts, software architects, developers, testers, implementers etc. The main issue with EA is that it represents complex systems from multiple perspectives. As a consequence multiple languages and knowledge sources are aggregated into one single EA model. Building a shared understanding of these models is really a challenge and often seen as time-consuming process having no directly usable results.

We consider ontologies and Semantic Web technologies as of special value for EA. The most important one is that they can transform human readable content into

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 208–219, 2012.

machine readable models having well defined meaning and available for further reasoning and automation.

As a consequence, the main goal of this work is to develop a formal representation of EA models such that to become possible to describe the EA in plain natural human language and then to derive the EA framework in the form of a machine readable version. Based on this version, the EA can then be transformed into multiple visualizations using XSLT techniques.

In order to do this we will make use of the Zachman framework (widely known as the first EA framework), the ISO/IEC 42010:2007 standard, ontologies and RDF technologies, and semantic annotation techniques. The remainder of this paper is structured in five sections. In Section 2 we explain why we have chosen the Zachman framework and the ISO/IEC 42010 standard to represent EA models. Section 3 is dedicated to the usability of the SW approach in EA development while in Section 4 we define the EA ontology. Then, in Section 5 we apply the HTML Microdata techniques to annotate HTML pages describing EA models in plain natural language with the EA ontology elements. Thus, we will finally be able to derive a machine readable EA model from the web page that describes the architecture. We conclude with related work (section 6) and future work (section 7).

2 Formal Models for Enterprise Architecture

In the 1987 article "A Framework for Information Systems Architecture" [16] Zachman introduced a classification schema for organizing the architecture of information systems. He started from the observation that the term "architecture" was used loosely by information systems professionals, and meant different things to planners, designers, programmers, communication specialists, and others. Thus he defined the EA matrix which provides a synoptic view of the models needed for enterprise architecture . One row represents a complete functional (sub-) system only regarded from the specific stakeholder's perspective. Each column represents a complete part of the system and it exists a translation mechanism between the models shown on each row.

In 1992, J. F. Sowa and J. A. Zachman extended that framework to include more formal representation of the model using conceptual graphs [12]. The goal of this conceptualization was to enable the machines to read the EA model by means of the CASE tools. The final version of the Zachamn framework indicates six columns (*what, how, who, when, where, why*) and five rows (*contextual, conceptual, logical, physical, detailed*).

Twenty years later, the proposed EA schema is still perfectly valid for the organization of IT resources of modern companies. In 2007 the ISO standardization body adopted ISO/IEC 42010:2007 standard (Recommended Practice for Architectural Description of Software-intensive Systems) formerly an IEEE standard [11], which comes to formally define an architecture description (AD): a document, repository or collection of artifacts used to define and document architectures. According to this standard, every system is considered in the context of its

environment. The environment of a system is understood through the identification of the *stakeholders* (e.g. client for the system, users, operators, developers, suppliers, regulators) of the system and their system concerns (e.g. data structure, behavior, data access, control, cost, safety, security). In order to take into consideration both the stakeholders and the many concerns of a system, the standard introduces two fundamental constructs of the system's architecture: *viewpoints* and *views*. They can be shortly described as follows:

- A *viewpoint* captures the conventions for constructing, interpreting and analyzing a particular kind of view such as languages, notations, model types, modeling methods, analysis techniques, design rules and any associated methods. Examples of viewpoints include: business, conceptual, technical, physical:
- A *view* is a collection of models representing the architecture of the whole system relative to a set of architectural concerns. A view is part of a particular architecture description for a system of interest. Examples of views: data, functions, events, roles

From the above definitions we can easily deduct that the Zachman's rows correspond to ISO/IEC 42010 viewpoints and the columns to views.

More formal models of enterprise architecture are based on reference architectures and are most effective when enforced by tools. Reference models are usually implemented as controlled vocabularies and/or taxonomies of terms. The most advanced one is the Open Group Architecture Framework (TOGAF[1]) with its Technical and Integrated Information Infrastructure reference models. The framework has been inspired by Zachamn and proposes three main views and viewpoints. The U.S. Government's Federal Enterprise Architecture (FEA) (OMB 2004) partitions the EA space into five reference models describing standard hierarchies of business functions, performance measures, application (system components), technology, and data. Other well defined EA frameworks include: eTOM[2] and SAFE[3]

All these frameworks have two things in common: (1) repositories of models for model reuse and (2) multiple models to describe different aspects of the same architecture.

3 The Relationship Between Enterprise Architecture and Semantic Web Technologies

Semantic Web (SW) has reached the maturity level nowadays with the proliferation of the semantic technologies based on Resource Description Framework[4] (RDF). RDF provides an infrastructure for uniquely identifying and merging both distributed data and metadata. RDF Schema (RDFS) and Web Ontology Language (OWL) are W3C

[1] http://www.opengroup.org/togaf/
[2] http://www.tmforum.org/InDepth/6637/home.html
[3] http://mike2.openmethodology.org/wiki/SAFE_Architecture
[4] http://www.w3.org/RDF/

standards for representing semantic models. RDFS offers a simple vocabulary for describing schemas or metadata. OWL provides a richer vocabulary (on top of RDFS) with a set of pre-built formalisms for expressing logical definitions and constraints. Ontologies and controlled vocabularies have been increasingly applied in many domains within the last years, such as in Medicine, Biology, eGovernment, Web Services, Blogs, Social Web etc. This trend is becoming even more prominent as more vocabularies (RDFS vocabularies or OWL ontologies) are being defined for and used by datasets in the Linked Open Data Cloud (see Linked Open Vocabularies - http://labs.mondeca.com/dataset/lov/index.html).

Following the advancement of SW, we have been motivated to explore the applicability of Semantic Web technologies to EA. Ontologies built using RDF and OWL enable integration of distributed data without assuming a single, monolithic, centrally controlled knowledge base. They also enable progressive capturing of new insights, shared understanding and formal structures. Additionally, SPARQL supports key scenarios for using RDF that are critical to EA, such as semantic interoperability, data integration and meaningful searching. Enterprises and/or departments within organizations can share and exchange EA models as well as query each other's data using RDF and SPARQL.

Based on these strengths, we believe that the application of the Semantic Web technologies and techniques to EA descriptions may lead to a range of new opportunities such as information discovery and aggregation, categorization of EA models, model equivalencies and further reasoning based on EA models. Moreover, semantic annotation techniques, meaning adding meta-data (or structured data) to existing (unstructured) data on the web, may provide very powerful for transforming natural language text provided in HTML into machine readable content. And this is one of the main goals of this paper: to allow business analysts to describe the EA in plain natural language and then to add metadata about the models used in such descriptions.

4 The Enterprise Architecture Ontology

We base our EA ontology on the concepts of the Zachman framework and the ISO/IEC 42010 standard as the most suitable combination for highly flexible Enterprise Architectures. The key advantage in using these two is they can be seen as providers of the building blocks for the other framework while not enforcing reference architecture as we believe there will never be a standard to cover all aspects important for one enterprise or the other.

Both the Zachman framework and the ISO/IEC 42010:2007 standard provide only general tools to organize and control the complexity of enterprise information systems. For example the content of each cell of the Zachman matrix is left entirely to the choice of the enterprise architect. This makes very difficult to create other tools to transform the matrix into machine readable content and vice-versa. As a matter of fact, we have identified 2 main issues when trying to apply the Zachman framework to model real enterprise architectures. First, it does not specify any granularity for the

concepts used to create a matrix. Should this matrix be defined to describe systems, processes, activities, services? Secondly the primary source of EA is the system's description using plain natural language, as the requirements of that system. Transforming this description into the EA matrix may lead to loss of details.

To be able to derive machine readable formal description of the EA from natural language description we first define the EA Ontology. To create this model, we add the following rules as an extension to the ones already defined by the Zachman framework:

1) The granularity for the Zachman matrix is the *Enterprise Resource* (ER). There can be only one matrix for each ER. This leads to redefining the EA as the architecture of an ER, showing all the components used by the ER across the enterprise system, where each ER may be itself a composition of other ERs.
2) We adopt the terms views and viewpoints as defined by the ISO/IEC 42010 instead of rows and columns for the EA matrix.
3) Each model is described in a web page (usually in a wiki environment).
4) Each entry (cell) in the matrix contains a link to the description of that model.

Listing 1 shows our proposed EA ontology, using RDF Schema. This model is created with the intent of being easily extendable and it does not have to be complete. It is expected to be extended by each enterprise architect with the required views and viewpoints.

Listing 1. The Enterprise Architecture Ontology

```
@prefix rdf: <http://www.w3.org/1999/02/22rdfsyntaxns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdfschema#> .
@prefix sea: <http://www.feaa.uaic.ro/ontologies/sea#> .
sea:EntepriseResource a rdfs:Class .
sea:View a rdfs:Class .
sea:Viewpoint a rdfs:Class .

sea:hasView a rdf:Property ;
      rdfs:domain sea: EntepriseResource;
      rdfs:range sea:View .
sea:hasViewpoint a rdf:Property ;
      rdfs:domain sea:View ;
      rdfs:range  sea:Viewpoint .
#fundamental views
sea:WhenView a sea:View .
sea:HowView  a sea:View .
sea:WhatView a sea:View .
sea:WhoView  a sea:View .
#fundamental viewpoints
sea:LogicalVP a sea:Viewpoint
sea:PhysicalVP a sea:Viewpoint
```

```
#view properties
sea:how a sea:hasView ;
      rdfs:domain sea: EntepriseResource;
      rdfs:range sea:HowView .
sea:when a sea:hasView ;
      rdfs:domain sea: EntepriseResource;
      rdfs:range sea:WhenView .
sea:what a sea:hasView ;
      rdfs:domain sea: EntepriseResource;
      rdfs:range sea:WhatView .
sea:who a sea:hasView ;
      rdfs:domain sea: EntepriseResource;
      rdfs:range sea:WhoView .
#viewpoint properties
sea:logical a sea:hasViewpoint ;
      rdfs:domain sea:View ;
      rdfs:range sea:LogicalVP .
sea:physical a sea:hasViewpoint ;
      rdfs:domain sea:View ;
      rdfs:range sea:PhysicalVP .
```

The EA ontology model is a straightforward representation of the Zachman framework enhanced with the rules defined earlier. In this model, we have the following terms:

- `EntepriseResource` (class) – it is the formal representation of the Zachman matrix concept. It is intended to be a dynamic structure with as many Views and Viewpoints as needed in particular scenarios.
- `View` (class) – it represents a column in the Zachaman framework and the View concept defined by ISO/IEC 42010. It specifies the aspects covered by the related Viewpoint models. The EA ontology also proposes four types of fundamental subtypes of the View class with well-defined meaning for each:
    1. `WhenView` - the event to which the Enterprise Resource responds to;
    2. `HowView` - the service/ application/task that handles the event;
    3. `WhatView` - the domain model used by the Enterprise Resource;
    4. `WhoView` – the organization role in charge of the resource;
- `hasView` (property) – specifies that an `EnterpriseResource` may have one or more associated `Views`;
- `Viewpoint`(class) – it represents the viewpoint concept in ISO/IEC 42010. Two types of viewpoints are defined by this ontology: `LogicalVP` and `PhysicalVP`
- `hasViewpoint` (property) – specifies that a `View` may be described using one or more `Viewpoints`;

- `when, how, what, who` (properties) – extend the `hasView` property and restrict the range to the corresponding type of View;
- `logical`, `physical` (properties) – extend the `hasViewpoint` property and restrict the range to the corresponding Viewpoints.

However, the proposed ontology is intentionally incomplete. For example, there is no Conceptual Viewpoint defined. As specified earlier, the main difference between the Zachman model and other EA frameworks is that it does not enforce any fixed (universal) combination of Views and Viewpoints. As this is considered a real advantage, allowing for a gradual evolution of a system's overall architecture, our EA ontology does not enforce either a certain model for the EA matrix. As we have said earlier, the ontology has been defined in order to be easily extendable by adding new Views and Viewpoints. So, an enterprise architect may extend the model depending on the needs, with Views such as Rules, Objectives, and Motivation or Viewpoints such as Conceptual, Deployment etc. Also, each Enterprise Resource may have a different EA ontological description than the others. Our ER example may also be an activity (one of the steps) in a business process. So it will be described using all the four Views above. On the other hand, a service may be described using only `HowView` and `WhatView` while an organizational role or employee may be described only using the `WhoView`.

We can see there is no Business Viewpoint either in our ontolgy. This is mainly due to the fact that we consider this Viewpoint as part of the natural language description upon which the other perspectives will build. In other words, it is the Business Viewpoint the root of the ER formal description and that is part of the plain natural language descriptions intended for humans, not for machines. These elements will be the identifier for business analysts to match the associated machine readable translation as described in the next section.

5 Using HTML Microdata to Annotate Enterprise Architecture Descriptions with EA Ontology Elements

As we have discussed in the introduction section, the main goal of our methodology is to end up with a plain natural language description of an Enterprise Resource while still preserving the capability of extracting the machine readable model based on this resource description. In other words, we would like the regular users to work only with the first line of the Zachman matrix (the business context viewpoint).

Our approach is based on the semantic annotation (SA) mechanism which means attaching meta-data to an XML document in order to add semantics otherwise unavailable by means of the original document's elements. Metadata, literally meaning data about data, is the language of the enterprise. Metadata should provide contextual information about every term or verb used in EA descriptions. The main goal is to greatly simplify the EA specification by using wikis (simple wikis not their semantic version) to create EA formal descriptions based on plain text for the whole enterprise system. To achieve this, there are three main techniques available today

from which we can choose to annotate HTML documents: HTML microdata [5], RDFa (XHTML annotations using attributes) [1] and microformats[5] (reuse XHTML elements). We have selected the first one as it provides the *finest granularity needed for our goal*. It is also the most recent W3C standard for semantic annotation (work still in progress) and has been conceived to provide a mechanism which allows machine-readable data to be embedded in HTML documents in an easy-to-write manner, with an unambiguous parsing model. It is built to be suitable for JavaScript tools as it is fully integrated within the Document Object Model of the host page.

The microdata model consists of groups of name-value pairs known as *items*. Each group is known as an item. Each item can have an item type, a global identifier (if the item type supports global identifiers for its items), and a list of name-value pairs. Each name in the name-value pair is known as a property, and each property has one or more values. Each value is either a string or itself a group of name-value pairs (an item).

To *exemplify* our EA ontology, and also to *validate* our model, we take the *Approve Order* system, and see how can we transform the description of this system into an Enterprise Resource. This resource may be described in plain English language as follows: "*When the Order message is received, the head of Financial department has to approve it using the Google Docs Spreadsheet service*". Based on this description we can derive the simplified version of the Zachman matrix shown in table 1. It is simplified because we present only three views and three viewpoints in order to point out the EA mechanism. In this table, we have *replaced the model with the URL that points to web page* describing each view-viewpoint combination.

Table 1. Simplified EA matrix for the Approve Order Enterprise Resource

		Event (When)	**Service (How)**	**Role (Who)**
Approve Order	Business	Order Received	Google Docs Service	Head of Financial Department
	Logical	http://company.com/OrderEventUML	https://developers.google.com/google-apps/spreadsheets/#adding_a_worksheet	http://company.com/OrgStructUML
	Physical	http://company.com/Event#	https://spreadsheets.google.com/feeds/worksheets/key/private/full	http://company.com/OrgStruct#

As already mentioned, we have introduced the ER concept as the building block of EAs and thus we have transformed the EA from a simple and universal matrix into an organized collection of formal descriptions of ERs.

Applying the rules defined by the new W3C HTML Microdata standard, and according to our EA ontology, one may annotate the HTML description of the Approve Order ER as it is shown in listing 2. One may notice we use the link element

[5] `http://microformats.org/`

extensively, on one hand because it is not a part of the presentation semantics and on the other hand it is suitable as a child element of the enclosing tag used to reference a certain View.

Listing 2. HTML description for the Approve Order annotated with EA ontology elements

```
  <div itemscope itemtype="http://.../sea#EnterpriseResource"
itemid="#ApproveOrder">   When the
  <div itemscope itemprop="when">
<link itemprop="logical"  href="http://.../OrderEventUML"/>
<link itemprop="physical" href="http://.../Event#"/>
      <b> Order message</b>     </div> is received,
  <div itemscope itemprop="who">
<link itemprop="logical"  href="http://.../OrgStructUML"/>
<link itemprop="physical" href="http://.../OrgStruct#"/>
  the head of the Financial department
  </div>
  has to approve it using the
  <div itemscope itemprop="how">
<link itemprop="logical"  href="https://.../google-
apps/spreadsheets/#adding_a_worksheet"/>
<link       itemprop="physical"       href="https://.../google-
apps/spreadsheets/#adding_a_worksheet"/>
      <a href="...">Google Docs service</a>
      </div></div>
```

The listing above may be translated into EA matrix and obtain the result shown earlier in Table 1. Also an XSLT template which applies the algorithms and rules defined by HTML microdata would be able to extract the RDF shown in listing 3 (Turtle serialization)

Listing 3. The RDF description of the Approve Order enterprise resource

```
  @prefix rdfs: <http://www.w3.org/2000/01/rdfschema#> .
  @prefix sea: <http://www.feaa.uaic.ro/ontologies/sea#> .
  @prefix sawsdl: <http://schema.org/sawsdl#> .

  <http://company.com/resources#approveOrder>
  sea:when [ sea:logical <http://company.com/OrderEventUML>
      sea:physical <http://company.com/Event#">]
  sea:who [  sea:logical < http://company.com/OrgStructUML >
      sea:physical < http://company.com/OrgStruct# >]
  sea:how  [sea:logical  <  https://developers.google.com/google-
apps/spreadsheets/#adding_a_worksheet >
      sea:physical    <    https://developers.google.com/google-
apps/spreadsheets/#adding_a_worksheet >]
```

6 Related Work

To the best of our knowledge, there is no similar work to formalize the Enterprise Architecture semantics using ontolgies and RDF and to apply such a model using semantic annotations of plain natural language descriptions.

Regarding more formal organization of the Zachman framework we can mention the work of Scott Ambler [2] which relates EA with the Rational Unified Process (with a focus on UML) or the Method for an Integrated Knowledge Environment (MIKE2.0[6]) which is an open source delivery framework for Enterprise Information Management. It provides a comprehensive methodology (861 significant articles so far) from which the Layered Semantic Enterprise Architecture[7] gives a layered organization for the SAFE architecture. The ontology layer plays the role of the common knowledge base where all models are described in the form of linked data and then translated to access the underlying enterprise assets. We give particular attention to this approach and we plan to evaluate its value for future extensions of our model towards executable enterprise architectures.

In what concerns the modeling languages for EA, we can mention ArchiMate, an Open Group Standard open and independent modelling language for enterprise architecture that is supported by different tool vendors and consulting firms. ArchiMate provides instruments to enable enterprise architects to describe, analyze and visualize the relationships among business domains in an unambiguous way. However, ArchiMate is specifically designed for TOGAF tools vendors. In contrast, we propose a language that is easily extendable and may be used to integrate with any RDF vocabulary available on the web such as the Linked Open Data vocabularies.

Regarding the semantic description of resources (which are just another type of enterprise resource in our model), most of the efforts are heading to semantically describe services on the web. Ontologies, like OWL-S[8], WSDL-S[9] and WSMO[10], are being used to semantically annotate Web Services with descriptions as instances of that model, and which expose, share, and connect these descriptions with other existing relevant data, information, and knowledge in the so called Web Data Space. Another example is the hRESTS microformat (HTML for RESTful Services)[9] designed to obtain machine-readable descriptions of Web APIs [Kopecky, 2008] described in the form of HTML service documentation for developers. All these techniques use SA mechanism to enhance the XML (or XHTML) description of services. However, there is one important limitation with SA, identified also in [4]: the first question before creating an annotation would be "what aspect of the content must be represented". So far, the great majority of WS annotation models can represent only one aspect (viewpoint in the EA terminology). The same idea is presented by the authors of

[6] http://mike2.openmethodology.org/

[7] http://mike2.openmethodology.org/wiki/Layered_Semantic_Enterprise_Architecture

[8] http://www.ai.sri.com/daml/services/owl-s/1.2/

[9] http://www.w3.org/Submission/WSDL-S/

[10] http://www.wsmo.org/TR/d2/v1.3/

hREST. They show that in order to obtain some level of automation one needs to capture four aspects of service semantics: information model (a domain ontology) represents data, especially in input and output messages; functional semantics specifies what the service does, by means of functionality classification or through preconditions and effects; behavioral semantics defines the sequencing of operation invocations when invoking the service; and nonfunctional descriptions represent service policies or other details specific to the implementation or running environment of a service. The authors propose an extension to hREST using WSMO-Lite [14] which defines a lightweight ontology for the four kinds of semantics, and uses SAWSDL (Semantic Annotations for WSDL and XML Schema [8] to annotate WSDL documents with instances of that ontology. This makes WSDL-based Web services amenable to SWS automation.

Our approach is different with respect to the existing efforts towards semantic description of services. First, the EA ontology builds on well defined concepts instead of arbitrary chosen aspects (like the ones presented above). Secondly, the proposed EA ontology promotes dynamic models where the number of aspects needed to be represented is not restricted by the ontology model. It allows the EA descriptions to evolve gradually, adapting to the growing complexity of the system. Thirdly, our target is not the service but any enterprise resource. The service (RESTful or plain old Web Service) might be just another enterprise resource described using as many views and viewpoints as needed.

7 Conclusions and Future Work

In this paper we have shown how SW technologies may be used to formally describe system's Enterprise Architecture expressed in plain natural language in order to obtain machine readable content. The paper is the first attempt in the literature to define an ontology for EAs based on a combination of Zachman and ISO /IEC 42010 state of the art models. The main research contribution of this new ontology is that it encourages diversity in EA models instead of enforcing a (closed) reference model. On the other hand, maybe the most important practical value of our model is the well defined extensibility mechanism, one aspect usually widely appreciated by IT professionals.

The paper also shows a methodology to transform plain text descriptions of EAs into machine readable content by annotating various parts of a HTML page, containing contextual description of the EA, with the EA ontology elements by using HTML microdata. Applying the proposed technique, organizations may use simple wikis (instead of sophisticated semantic wikis) to gradually build the EA models, more and more complex and to finally obtain the EA of the whole enterprise system.

As the main orientation of our future work we plan to extend this model with the final goal to obtain executable Enterprise Architecture. Such a goal is rather achievable since each model in the EA matrix may have a well defined RDF description. For example the conceptual viewpoint may link to UML diagrams serialized in RDF using the Standford's vocabulary to map UML concepts to RDF[11].

[11] `http://www-db.stanford.edu/~melnik/rdf/uml/uml-state-20000507.rdf`

On the other hand, the detailed (application) viewpoint models may link to the RDF descriptions based on well defined ontologies such as WSMO (when the HowView points to a Web Service) or hREST (when HowView points to a RESTful services) or FOAF (for WhoView).

The second direction of research will head towards investigating the usage of Natural Language Processing tools in order to automate the extraction of relevant items out of an EA statement written in plain text. Such an approach should boost the productivity of software specification while providing guaranteed consistency between different EA descriptions.

References

1. Adida, B., Birbeck, M.: RDFa Primer. Bridging the Human and Data Webs (2008), http://www.w3.org/TR/xhtml-rdfa-primer/
2. Ambler, S.: Extending the RUP with the Zachman Framework (2004), http://www.enterpriseunifiedprocess.com/essays/zachmanFramework.html
3. Booth, D., et al.: Web Services Architecture, W3C Working Group Note (2004), http://www.w3.org/TR/ws-arch/
4. Euzenat, J.: Eight questions about Semantic Web annotations. Intelligent Systems 17(2) (2005)
5. Hickson, I.: HTML Microdata - W3C Working Draft (2012), http://www.w3.org/TR/microdata/
6. IEEE Standard 1471-2000: Recommended Practice for Architectural Description of
7. Kopecky, J., Gomadam, K., Vitvar, T.: hRESTS: An HTML Microformat for Describing RESTful Web Services. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT, vol. 1, pp. 619–625 (2008)
8. Kopecky, J., et al.: SAWSDL: Semantic Annotations for WSDL and XML Schema. IEEE Internet Computing 11(6), 60–67 (2007)
9. Kopecky, J., Gomadam, K., Vitvar, T.: hRESTS: An HTML Microformat for Describing RESTful Web Services. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, pp. 619–625 (2008)
10. Sessions, R.: A Comparison of the Top Four Enterprise-Architecture Methodologies (2007), http://msdn.microsoft.com/en-us/library/bb466232.aspx
11. IEEE: IEEE Standard 1471-2000, Recommended Practice for Architectural Description of Software-Intensive Systems. IEEE Computer Society, New York (2000), http://www.iso-architecture.org/ieee-1471/
12. Sowa, J., Zachman, J.: Extending and Formalizing the Framework for Information Systems Architecture. IBM Systems Journal 31(3), 590–616 (1992)
13. Udrea, O., Recupero, D.R., Subrahmanian, V.S.: Annotated RDF. ACM Transactions on Computational Logic, 1–40 (August 2008)
14. Vitvar, T., Kopecký, J., Viskova, J., Fensel, D.: WSMO-Lite Annotations for Web Services. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds.) ESWC 2008. LNCS, vol. 5021, pp. 674–689. Springer, Heidelberg (2008)
15. Weill, P.: Innovating with information systems. What do the most agile firms in the world do? In: Sixth e-Business Conference Pwc&IESE, Barcelona, Spain (2007)
16. Zachman, J.A.: A Framework for Information Systems Architecture. IBM Systems Journal 26(3) (1987)

Two-Phased Protocol for Providing Data Confidentiality in Cloud Storage Environments

Alexandru Butoi, Gabriela Andreea Morar, and Andreea Ilea

Faculty of Economics and Business Administration, Babes-Bolyai University, Cluj-Napoca, Romania
{alexandru.butoi,gabriela.morar,andreea.ilea}@econ.ubbcluj.ro

Abstract. In expanding fields, such as cloud computing, one of the greatest concerns is related to data security. Transferring and storing data to a cloud computing infrastructure raises many issues related to data privacy and data integrity. In this context, we propose[1] a protocol designed to provide data storage confidentiality in the cloud. The protocol implies splitting the data into small encrypted data chunks, dispersed into data storage volumes. It is a two phased protocol that handles separately the operations related to writing the data to the storage volumes and the operations of retrieving data from the corresponding volumes.

Keywords: cloud computing security, data confidentiality, cloud storage, protocol, middleware.

1 Introduction

Cloud Computing is a rapidly evolving technology paradigm, which facilitates the provisioning of computational resources and services and as a consequence, the cloud's secure storage services are rapidly gaining popularity. On the other hand, data being moved to a public or community cloud, there, new threats and risks appear. Therefore, data confidentiality issue becomes very important, as the data is stored and processed on the cloud provider's hardware at the provider's data center. In this context, encryption and digital signature schemes replace physical location, as means of protecting data confidentiality and integrity.

In order to prevent unauthorized disclosure of data stored in public or hybrid cloud storage a great deal of attention must be paid to assuring data confidentiality by SLA enforcement or by developing secure interaction mechanisms with the cloud infrastructure. As a result, some of the most important aspects to consider from the point of view of a cloud service provider, cloud consumer, and third-party authorities as well, are: virtualization infrastructure, software platform, identity management and access control, data integrity, confidentiality and

[1] G. A. Morar acknowledges support from the: Investing in people! Ph.D. scholarship, Project cofinanced by the Sectoral Operational Program for Human Resources Development 2007 - 2013, contract nr. POSDRU/88/1.5/S/60185 "Innovative Doctoral Studies in Knowledge Based Society"İ. The results presented in this paper were partly supported by the CNCSIS TE_316 Grant.

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 220–230, 2012.

privacy, physical and process security aspects and none the less, legal compliance in cloud.

Nowadays businesses are producing more and more data and the costs of storing and administering them are raising. In this context, a public cloud infrastructure is an alternative to cut down the costs with data storage because building and administering a private data center or deploying a private cloud implies higher costs. One of the main barriers for real businesses, in choosing a public cloud storage system for storing their business data, is the lack of trust that their confidential data will be kept confidential. In order to tackle with this aspect some secure solutions for interacting with public cloud storage are needed. In respect with this approach, we propose a protocol for storing confidential data in a public cloud. It will allow storing critical data in public cloud by protecting it for unauthorized disclosure of data. The aim of this article is to describe the proposed protocol and to demonstrate its applicability and performance in terms of computational costs. We propose a middleware that will enable clients to safely store data in the Cloud and protect their high confidential data for unauthorized access.

The paper is structured as follows. In Section 2 we present the work accomplished so far in data security in the Cloud. The following section, Section 3, we present a detailed analysis of our protocol for securely sending data to the cloud. In Section 4 we describe our experiments, and present the results in Section 5. In the final part of the paper we state our conclusions and future work.

2 Related Work

This section presents part of the related work that has been conducted so far in the Cloud security area with focus on confidentiality issues. [1] takes a holistic view of cloud computing security by spanning across the possible issues and vulnerabilities connected with the use of cloud services. Some of these are: virtualization infrastructure, software platform, identity management and access control, data integrity, confidentiality and privacy, physical and process security aspects, and legal compliance in the cloud. The authors also outline the importance of future development in regards of Trusted Computing, Information Centric Security and Privacy Preserving Models. Moreover, in order to protect data confidentiality and privacy, the cloud must prevent certain attacks and give users the ability to assess whether the necessary mechanisms are in place, instead of simply trusting the cloud providers.

[2] uses encryption mechanisms and proposes a model for improving database confidentiality in the cloud by classifying data according to their confidentiality level and by using a set of tools named Silverline. These tools provide key management and encryption mechanisms in meant to reduce the complexity of the cloud database confidentiality protection and are designed in a manner that does not affect the performances of the applications which exploit these databases.

In the same context of encryption techniques, [3] improves the Key-Policy Attribute Based Encryption for cloud computing using a system which divides

files in header and body and selectively delegates decryption through a Type-Based Proxy re-encryption mechanism.

Another approach to data confidentiality protection in the cloud is presented in [4] which introduces a set of protocols named Secure Multiparty Cloud Computing (SMCC). They are using a scenario in which several number of entities are sending their data into the cloud. SMCC protocols provide data anonymity by introducing a transfer layer between the clients and the providers. By using the concept of Deep Cloud Confidentiality it creates a secure transmission and storage environment using encryption techniques. Similar to our approach, these protocols provide specific mechanisms which randomly select the storage resources and providers from trusted resource pools and infrastructures. These protocols are designed in the context of specific cloud architecture, the Cloud Cube Model, while our approach is designed to interact with any cloud provider which provides virtual storage volumes.

[5] introduces a new model for protecting sensitive data in cloud environments through a clear separation between Software as a Service providers and Infrastructure as a Service providers, and hiding the information regarding the data ownership combined with data masking algorithms.

Similar to our approach and related to the cloud confidentiality issue but not limited to it, [6] presents Depsky, a cloud system architecture for storing critical patient data in medical domain. The project solves the confidentiality, integrity and availability of data in cloud computing by scattering it in storage cloud volumes selected from four different storage providers and using encryption based protocols. Another relevant project is Belisarius [7] which tackles the problem of fault-tolerance combined with secure storage and verifiable secret sharing scheme in order to reduce the additional infrastructure and complexity implied by the confidentiality issues.

Data confidentiality is seen as a mandatory premise for the future development of cloud technologies by [8] and solves the problem of confidential data processing in the cloud by introducing Oblivious Data Processing (ODP). This consists in a set of algorithms and systems capable to process data in a manner that guarantees the confidentiality. ODP is based on partitioning, distribution and independent processing of the partitions in Residue Domains.

Moreover, [9] identifies five cloud data patterns used for assuring confidentiality in such environments: Confidentiality Level Data Splitter which stores data into a location by confidentiality level, Confidentiality Level Data Aggregator which assures the reverse process of the aggregator, Pseudonimyzer of Critical Data passes critical data from private cloud to a public cloud by using pseudonymization process of high confidential data and Anonymizer of Critical Data. Some common points between our protocol's mechanisms and the above approach are: interaction is made only through the middleware solution which implements the protocol that splits, encrypts and sends a file into the cloud storage, each chunk being a smaller file that is pseudonymized by renaming it with its own hash signature before storing it into the cloud.

Comparing the existing approaches with ours we can conclude that most of them are based on largely accepted solutions like encryption schemes and algorithms, public key infrastructure, advanced techniques of data hiding, or secure standardized protocols like SSH, SSL/TLS. For example DepSky [6] is an approach aiming to achieve confidentiality, integrity and availability by storing data on four different cloud providers while our approach uses one cloud provider and multiple virtual volumes which can be less costly, and we tackle only with the aspect of confidentiality. Moreover Belysarius [7] proposes a fault-tolerance combined with assuring data confidentiality approach. Our approach deals only with confidentiality aspect and currently has the disadvantage of the single point of failure, aspect which will be considered in the future work. In ODP paradigm the data is partitioned in Residue Domains and distributed in the form of partitions while we distribute one file (split into smaller chunks) in several different cloud storage volumes in such a way that when an unauthorized person tries to recompose the original file the reverse engineering should be hard to achieve. Like SMCC [4] we propose a protocol that will guarantee the secret of the data stored in a public cloud infrastructure by combining an encryption scheme with data dispersion and temporary destroying data semantic integrity in a reversible process.

3 Protocol

Our protocol is designed to provide data storage confidentiality in the cloud through a middleware component. It has the role of splitting the files into small encrypted data chunks and disperse them into the data storage volumes, thus making the data hard to find and read by unauthorized persons.

The protocol requires for the client to have more than two data storage volumes (for example Amazon EBS volumes) at their disposal, such that the middleware running on the client side can securely distribute the data on the cloud infrastructure. The protocol has two phases: the Send phase, which is associated to the operation of writing a file to the data storage volume, and the Retrieve phase, which is associated to the operation of retrieving a file from the data storage volume. The Send phase is depicted in table 1. This phase shows how data is split into chunks and sent to the cloud storage volumes. The file is first sent to the middleware which splits it into small files (data chunks) of a given size. Using a symmetric encryption algorithm, the resulted data chunks are encrypted with a secret key known only by the user. Next the protocol gets the list of available storage volumes from the cloud provider and uses a weighted undirected graph to distribute data chunks to the available cloud storage volumes. The graph used is defined as follows: each available storage volume is considered a vertex of the graph, and the vertices of the graph are connected with weighted edges in order to form a connected graph. Each data chunk is sent to a cloud storage volume determined as follows:

Considering $S = V_1, V_2, V_3, \ldots, V_k, \ldots V_n,\ k = 1, ..., n$ the set of the available cloud storage volumes and $C = C_1, C_2, C_3, \ldots, C_i, \ldots, C_m,\ i = 1, ..., m$ the set

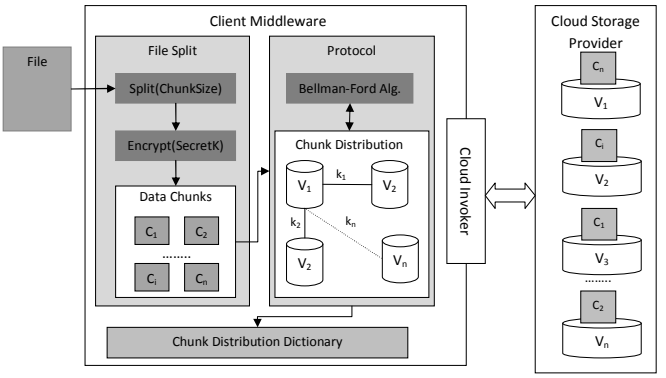


Fig. 1. File sending phase

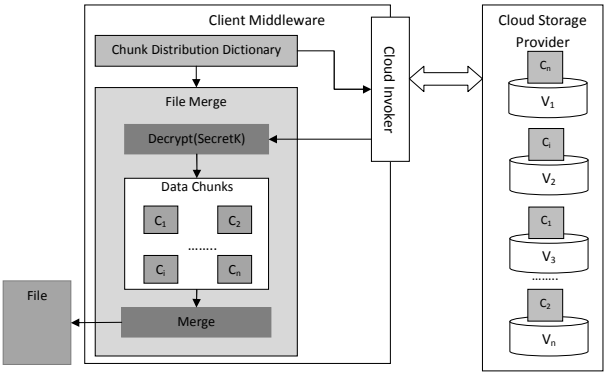


Fig. 2. File retrieving phase

of the encrypted data chunks to be sent to the cloud storage. $G(S, E(V_x, V_y, w))$, is the connected graph used for data chunk distribution. S is defined above and E is a set of edges between V_x and V_y, with $x, y = 1, ..., n$, and w is the corresponding weight of the edge. The data volume to where the first data chunk will be sent is randomly chosen from the available volumes. Based on this ,V_k, the current volume is selected from S applying the Bellman-Ford shortest path algorithm to G. $Sig(C_i)$ is the unique check sum calculated of the C_i data chunk file, I_i is the index of C_i in the source file (the position of the data chunk in the original file). As 1 describes, first a random volume $startNode$ is chosen. Then for each data chunk the volume V_k on which it will be stored is the volume that has enough available space and is located on the shortest path in the graph from the $startNode$. In order to determine the shortest path in G, we used the Bellman-Ford algorithm which computes the shortest paths from a starting node to others. When the destination volume V_k is determined, the chunk is sent to the cloud and stored on the V_k volume which becomes the starting point for the

Table 1. Send phase of the protocol

$startNode = RandomChooseVolume(G)$;
for each C_i in C
    $V_k = GetShortestPathVolumeWithAvailableSpace(G, startNode)$;
    $sendChunkToCloud(V_k, C_i)$;
    $startNode = V_k$;
    $AddToDictionary(Sig(C_i), I_i, V_k)$;
    $ReorganizeGraph(G)$;

next volume search. The operation is journalized into the Chunk Distribution Dictionary. Each record from the dictionary is composed of the signature of the data chunk (which is the checksum of the file and uniquely identifies it), the index of the data chunk in the original file and the destination volume where the data chunk will be stored. In order to have variety between the selected volumes, each weight of the graph is randomly regenerated after every volume selection.

The Chunk Distribution Dictionary of the middleware is the corner stone of the protocol because it permits to identify all the data chunks of a file dispersed in the cloud storage. It records the storage volume on which every data chunk is stored and the location of the data chunk in the original file. The file that contains a data chunk is renamed with its signature before being sent to the cloud storage, thus improving the confidentiality. Moreover the signature is used by the protocol for the file integrity check.

The dictionary is not sent to the cloud but remains on the user side. Therefore only the user who has the correct dictionary and holds the secret key used for encryption will be able to recompose the file, preserving the confidentiality of the data stored in that file. In addition the file dispersion mechanism to different storage volumes assures that files cannot be recomposed by an intruder that will not be able to identify all data chunks of a file.

The Retrieve phase is described in table 2 and is less complex than the first one. It starts when a user wants to retrieve a file previously stored on the cloud using the Send option of the protocol. The file retrieval is based on the records stored in the Chunk Distribution Dictionary and can be done only by the user who possesses the right dictionary for the file. Every record in the dictionary provides the unique name of a data chunk, the location of the data chunk in the original file and the cloud storage volume where the data chunk is stored. If we consider C_i the current data chunk, the dictionary record associated to it will be $r_j = (Sig(C_i), I_i, V_k)$ where $Sig(C_i)$ is the unique checksum of the file that contains the data chunk, I_i is the index of the data chunk in the original file and V_k is the cloud storage volume where the data chunk is stored. A file F will be described in the dictionary by the record set $RF = \{r_j = (Sig(C_i), I_i, V_k) \mid j = 1, c\}$, where c is the total number of data chunks that originate from the file F. The data retrieval is done using the algorithm: For each record associated with the desired file, the data chunks from the volume V_k are retrieved. They must have the signature provided by the record and a valid check for their integrity. The integrity of a chunk is examined by

Table 2. Retrieve phase of the protocol

```
for each r_j in RF
    r_j = (Sig(C_i), I_i, V_k);
    C_i = getDataChunkFileFromCloud(Sig(C_i), V_k);
    if checkForIntegrity(C_i, Sig(C_i))
      Decrypt(scretKey);
      addChunkToFile(F[I_i], C_i);
```

comparing its signature with its actual checksum. If the integrity check succeeds the data chunk will be decrypted with the secret key and it will be added to the original file at the location given by the recorded index I_i.

The purpose of the splitting files is to temporary and reversible destroy the semantic integrity of the data stored in a specific file and then dispersing the data chunks in different cloud storage volumes in order to make the information unreadable by unauthorized entities. The encryption of each data chunk is an extra measure of security so if someone wants to have access to a file he has to posses both: the secret key and the distribution dictionary of the file. The problem of the single point of failure can occur if the secret key or the dictionary is lost or corrupted but this issue can be addressed with extra security policies.

4 Experiments

In order to test the proposed protocol several experiments are conducted. For the relevancy and comparability of the results we looked at the Amazon Storage Services and tuned our simulation environment according to their service specifications. The parameters taken into account are: number of volumes, dimension of a data chunk, the size of the file, transfer rate and topologies based on which the data chunks are distributed among data volumes. The number of volumes can take values from the set VN={25, 50, 75, 100}. The considered dimensions for the files are 5MB, 50MB, 250MB, 500MB. Based on these dimensions the size of the data chunks can vary from 0.1 MB to 300MB per chunk. For the encryption of the data chunks the DES algorithm is used. The considered transfer rate selected, based on the description given in [10], is 120MB/sec and 70MB/sec. The data chunks can be distributed to the available volumes based on 4 different topologies that imply a connected graph as a starting point: Random Graph, Complete Graph, Cycle Graph, Binary Tree. For the implementation of the graphs JGraphT Open Source [11] library was used. It allowed us to implement the weighted graphs used by the protocol. The library provides also an implementation of the Bellman-Ford algorithm.

In order to simulate the storage Cloud environment and storage volumes (volumes similar as a concept to Amazon EBS [12]) the CloudSim Simulator [13] is used. It simulates the behaviour of a typical hard drive storage in a cloud infrastructure.We used the class *org.cloudbus.cloudsim.File* for simulating a file in the cloud environment and *org.cloudbus.cloudsim.HarddriveStorage* for simulating the cloud storage volumes. The middleware contains the implementation

of the protocol and it is integrated with CloudSim. The default values for this storage are those of a Maxtor DiamonMax 10 ATA hard-disk with the following parameters: latency = 4.17 ms, avg seek time = 9 ms max transfer rate = 133 MB/sec. The data volume dimension is randomly chosen from the set VS={50GB, 100GB, 250GB, 500GB, 1000GB} when the topology is generated.

5 Results

This section summarizes the results obtained by running the experiments with the setup described above. Due to variation of multiple parameters that can influence our protocol we focused our efforts on testing its performance in relation to them.

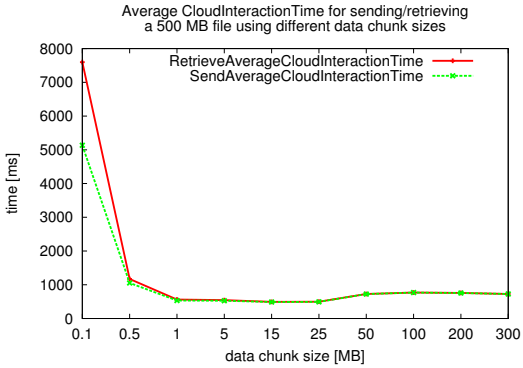


Fig. 3. Both protocol phases execution time evolution for different chunk sizes

First we analyzed time evolution of sending and retrieving for different data chunk sizes into the cloud. Results are presented in figure 3. The Cloud Interaction Time when retrieving a file is the total time elapsed from the moment when the cloud instance receives the first request until it serves the last chunk of data.The Cloud Interaction Time when sending a file is the total elapsed time from the moment when the cloud instance receives the first data chunk to be stored until the last data chunks of the file is stored. We ran this experiments set for the following parameters: number of volumes 50, maximum transfer rate 120 MB/s for 500 MB files with different data chunk sizes. The horizontal axis contains the data chunk sizes and the vertical axis contains the cloud interaction time. Both phases of the protocol follow the same pattern: for chunk sizes smaller than 1 MB it decreases exponentially as the chunk size increases but for the chunks larger than 1 MB it evolves relatively constant. For smaller chunk sizes than 1 MB the send phase is faster than the retrieve phase. This can be explained by the extra computational effort for processing a larger number of data chunk files and distribution dictionary records (increases as the chunk size decreases).

Next we analyzed how the graph type used by the send phase of the protocol affects the complexity of the algorithm through measuring the average total send time when data chunk dimension is kept to 0.5. The total send time respects the same evolution pattern of slowly and linear increase for file sizes lower than 50 MB then increases more rapidly for file sizes greater than 50 MB. When the complete graph is used the protocol produces the biggest average execution time because the shortest paths are searched in the maximum number of edges sets. In case of the random connected graph the number of edge is lower than in the complete graph and the search of the shortest paths is faster in terms of execution time. The cycle graph and the binary tree usage increases linearly as file size grows and follow the same evolution due to the well structured nature of these graph types.

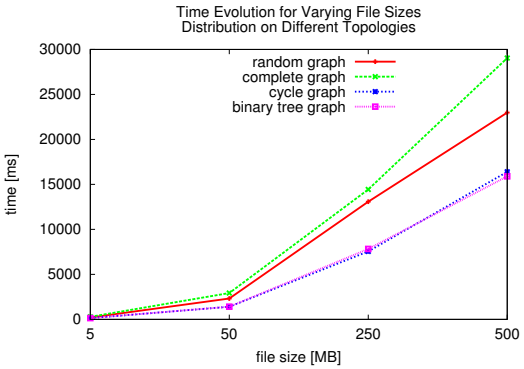


Fig. 4. Performance comparison of different topologies. Cycle Graph is always the best except for the largest file size tested.

In the next set of experiments we analyzed how splitting/merging, encryption/decryption and the distribution of the data chunks using weighted graph algorithms affects the performance of our protocol. We implemented a simplified version (named here Raw Protocol)of the protocol which in the Send phase only splits the file into chunks but does not encrypts the data chunks and distributes them into the cloud storage randomly without using the weighted graph algorithm as normally the protocol does. The Retrieve phase of the Raw protocol is still using the data chunk distribution dictionary. We ran the two versions of the protocol on the same set of experiments with the following parameters: number of volumes 50, maximum transfer rate 120 MB/s and for 500 MB files with different data chunk sizes. Then we computed Total Execution Time for both phases of each protocol version which is the sum of the elapsed time for split/merge process and execution time of the destination storage volume identification. For each data chunk size we have calculated the Average execution time in both cases and the results are comparatively presented in figure 5. We

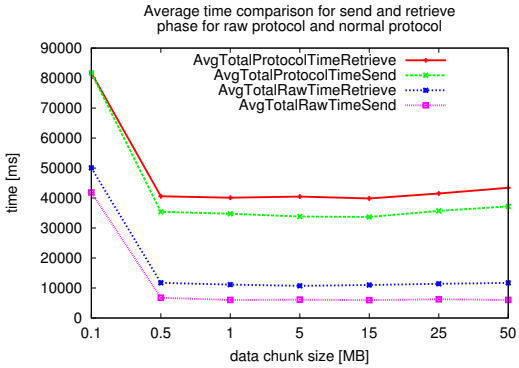


Fig. 5. Performance comparison for sending a 500 MB file using different data chunks sizes

observe that for both versions of the protocol the maximum Average Total Time is achieved when the data chunk is smaller than 0.5 and for data chunk larger than 0.5 it evolves relatively constant in both phases of the protocol. Compared with the Raw Protocol version, our protocol has an average increase in Total Execution Time of 3000 ms (3 seconds). The cost in terms of execution time of a higher level of confidentiality assured by encryption and advanced graph based volume selection is 3 seconds which is acceptable.

6 Conclusions

In order to tackle the problem of cloud data storage confidentiality we proposed a two-phase protocol that interacts with cloud storage by splitting and encrypting the original file into data chunks. Then it stores the data chunks separately into different available storage volumes.

The protocol is implemented similar to a client-side middleware that has a dictionary which records the exact location of every data chunk in the original file and the exact location where every data chunk can be found in the cloud environment.

The confidentiality is provided by the encryption of the data chunks and by the dispersion mechanism of them in several cloud storage volumes. Moreover, only the user which has the secret key and the right dictionary can retrieve all the data chunks and recompose them into the original file. The protocol generates an average extra cost of 3 seconds in terms of execution time which is acceptable.

Future research will tackle the elimination of the single point-failure or defining a relation between data protection and the utility obtained by the user.

References

1. Sengupta, S., Kaulgud, V., Sharma, V.S.: Cloud computing security–trends and research directions. IEEE Congress on Services, 524–531 (2011)
2. Puttaswamy, K.P.N., Kruegel, C., Zhao, B.Y.: Silverline: toward data confidentiality in storage-intensive cloud applications. In: Proceedings of the 2nd ACM Symposium on Cloud Computing, SOCC 2011, pp. 10:1–10:13. ACM, New York (2011)
3. Do, J.M., Song, Y.J., Park, N.: Attribute based proxy re-encryption for data confidentiality in cloud computing environments. In: ACIS/JNU International Conference on Computers, Networks, Systems and Industrial Engineering, pp. 248–251 (2011)
4. Maheshwari, N., Kiyawat, K.: Structural framing of protocol for secure multiparty cloud computation. In: Asia International Conference on Modelling & Simulation, pp. 187–192 (2011)
5. Yau, S.S., An, H.G.: Protection of users' data confidentiality in cloud computing. In: Proceedings of the Second Asia-Pacific Symposium on Internetware, Internetware 2010, pp. 11:1–11:6. ACM, New York (2010)
6. Bessani, A., Correia, M., Quaresma, B., André, F., Sousa, P.: Depsky: dependable and secure storage in a cloud-of-clouds. In: Proceedings of the Sixth Conference on Computer Systems, EuroSys 2011, pp. 31–46. ACM, New York (2011)
7. Padilha, R., Pedone, F.: Belisarius: Bft storage with confidentiality. In: NCA, pp. 9–16. IEEE Computer Society (2011)
8. Phatak, D.S., Sherman, A.T., Pinkston, J.: A new paradigm to approximate oblivious data processing (odp) for data confidentiality in cloud computing. IEEE Congress on Services, 391–398 (2011)
9. Strauch, S., Breitenbuecher, U., Kopp, O., Leymann, F., Unger, T.: Cloud data patterns for confidentiality. In: Proceedings of the 2nd International Conference on Cloud Computing and Service Science, CLOSER 2012, Porto, Portugal, April 18-21, pp. 387–394. SciTe Press (2012)
10. Amazon's elastic block store explained, `http://blog.rightscale.com/2008/08/20/amazon-ebs-explained/`
11. Jgrapht open source library, `http://jgrapht.org/`
12. Amazon elastic block store (ebs), `http://aws.amazon.com/ebs/`
13. Cloudsim, `http://www.cloudbus.org/cloudsim/`

Exploring the Meaning behind Twitter Hashtags through Clustering

Cristina Ioana Muntean, Gabriela Andreea Morar, and Darie Moldovan

Faculty of Economics and Business Administration, Babes-Bolyai University, Cluj-Napoca, Romania
{cristina.muntean,gabriela.morar,darie.moldovan}@econ.ubbcluj.ro

Abstract. Social networks are generators of large amount of data produced by users, who are not limited with respect to the content of the information they exchange. The data generated can be a good indicator of trends and topic preferences among users. In our paper we focus on analyzing and representing hashtags by the corpus in which they appear. We cluster a large set of hashtags using K-means on map reduce in order to process data in a distributed manner. Our intention is to retrieve connections that might exist between different hashtags and their textual representation, and grasp their semantics through the main topics they occur with.

Keywords: k-means, clustering, hashtag, twitter.

1 Introduction

Twitter has been a prolific environment for analysis allowing research to dive into real worldwide large-scale phenomena. There have been various studies on its content and structure. Twitter is a micro-blogging platform that allows users to make *tweets*, messages no longer that 140 characters, resembling SMS (Short Message Service). Tweets are synthetic messages containing different kinds of information: links, media attachments, mentions (@) and hashtags (#).

The user has no limitations regarding the content of the text they can write in a tweet. This freedom and lack of formalism generates issues when it comes to analyzing the text and classic NLP tools seem almost powerless. The text can hold acronyms like "tfb", concatenated phrases like "ilikeitwhen" or it can contain spelling mistakes. Due to Twitter slang particularities, even the most popular terms can be cryptic to users, and even more so to automatic text processing applications. In our research[1] we attempt to make a first step towards finding structure and meaning in hashtags. In this preliminary study we wish to cluster hashtags in order to decipher their meaning, with the help of their

[1] Acknowledge support from the: Investing in people! Ph.D. scholarship, Project co-financed by the Sectoral Operational Program for Human Resources Development 2007 2013, contract nr. POSDRU/88/1.5/S/60185 Innovative Doctoral Studies in Knowledge Based Societyand and the CNCSIS TE_316 Grant.

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 231–242, 2012.

unique content, and group them into semantically interconnected groups. This can be a very useful task for disambiguating the meaning of hashtags such as "1thingiwant4christmas" without the need for a human assessor. The complex way of creating tags on Twitter can be revealed, and they can be structured according to the degree of granularity desired.

On the other hand one can find significant utility in hashtags. They can help Twitter users to go beyond the friendship level and follow topics, discussions or tweets that might not appear in their timeline, but are of interest to them. Just like following certain users [1] can help increase information gain, a similar effect can be achieved with following hashtags.

Application of machine learning techniques on Twitter, like classification, clustering or recommendation, can become useful for better fitting the interest of the user. In [2, 3] the authors study the applications of topic modeling algorithms, including LDA. The clustering task can be seen as a preliminary task to several more complex tasks like recommendations, filtering or ranking [4]. Clustering tasks on Twitter include tweet clustering or user clustering, but to the best of our knowledge hashtags clustering has not been thoroughly studied yet.

In our analysis we assume that each hashtag has a unique representation in our dataset, composed of the concatenation of all tweets which include it. It is quite improbable for two different hashtags to have the same so-called virtual document. This would mean they would have to co-occur each time either of them is mentioned. We present the proposed hashtag representation model, which forms the basis of our experiments, namely the dataset, and the results obtained after various clustering options.

In our experiments we cluster approximately 280.000 distinct hashtags from approximately 900.000 daily tweets per dataset, using K-means while varying the number of clusters k. According to the granularity of k we can obtain a more general grouping, if k is large, e.g. 500, it means we have 500 categories covering different major topics, the hashtags are more segregated and thus refer to more specific pieces of information. By lowering k, the groups become more generic. We carry out tests with different values of k, on datasets from three consecutive days.

The paper is structured as follows. In Section 2 we present a variety of studies conducted on Twitter, studies similar to ours regarding clustering, and some references to running large-scale data experiments in a distributed way using map-reduce. The following Section 3 presents a detailed analysis of our dataset. In Section 4 we describe our experiments, from preprocessing to algorithm tuning for the results shown and explained in Section 5. In the final part of the paper we present our conclusions, future work and possible applications.

2 Related Work

The overall popularity of Twitter has created various research themes. Several studies with regard to the content, the dynamics and the structural characteristics of Twitter have appeared in recent years. There are papers that concentrate

their attention on information diffusion throughout the network, on the discovering of communities, and on the analysis of user intents [5, 6, 7]. The detection of spam is also an important topic and has been studied in [8, 9]. Papers that characterize Twitter as a news media offer solutions to recommendation tasks like news and contents [10, 3, 11] or users [1]. In [12] the authors make a state-of-the-art survey on research on Twitter and try to define possible topics and open problems regarding the matter.

Several papers on machine learning techniques applied to Twitter tackle subjects like summarization and topic detection (LDA) [13], clustering [14] and disambiguation of topics or classification [15, 16]. Most studies on clustering regarding Twitter include topic modeling algorithms. In [15] the authors use LDA in order to classify short and sparse text using hidden topics from large-scale data. Recommendation systems use clustering as a prior step to offering suggestions. In [11] the authors suggest tweets based on a user's history and topic model. They transform text according to VSM and assign TFIDF weights to vectors. Similarly, TwitterRank [4] is based on tweet topics and the authors' attempt to find influential users. They use LDA to build topic models for each users according to their tweets.

In [2] the authors tackle the Twitter dataset from a NLP point of view and observe the esoteric nature of language and grammar, the fact that short text contains less stop words and word redundancy. In their analysis of disaster-related Twitter data they use probabilistic topic models and treat data as bags of words – due to the lack of fluency withing tweets – while also inferring latent relationships between data.

A comparison between K-means, SVD and affinity propagation, a graph based approach, has been made in [17], in which authors test various clustering techniques on short text documents, namely tweets. As in the case of our study the biggest challenge in handling short text is the problem of sparsity. TF will be very small in most cases so the vectors are basically represented by the IDF. In order to tackle this problem, they propose building the vectors with value 1 when the word occurs and 0 for the contrary case. They use two distance measures, one based on Jaccard coefficient and the other Cosine. Their experiments are run on a set of 661 tweets with a vocabulary consisting of 1678 distinct words. For evaluation they use cluster density techniques.

The above cited papers rely heavily on annotated data and small sample datasets on predefined topics. They mostly focus on clustering tweets, while in our case we are more preoccupied with clustering hashtags on a wide data sample in which also the rare hashtags are considered.

In what regards large scale data processing, experiments can be optimized by taking advantage of distributed computing. For our experiments we use the MapReduce paradigm [18], which is designed to simplify the concepts around large scale distributed computing and allows dealing with large datasets. It is divided into two steps: map and reduce. The map function takes a single instance of the type key/value pair as an input. The output of the function are key/value pairs that are grouped by key and are used as an input for the reduce function.

Based on the key value and the list of values outputted by the map function, the reduce function performs some computations over that list and outputs key/value pairs.

Map Reduce has been implemented by projects like Hadoop [19] and Disco [20]. Hadoop is an open-source implementation of MapReduce and it was chosen because it is currently the most feature-complete system and widely used in industry.

The Hadoop framework is composed of the MapReduce functionality and a distributed file system (HDFS) [21]. The distributed file system has the role of distributing input data across the cluster. Hadoop tries to allocate map tasks based on the physical location of each piece of data on HDFS. There are several distributed machine learning libraries that use MapReduce. The most famous ones are Apache Mahout [22] and Weka [23]. Apache Mahout has implemented algorithms described by Chu et al. [24].

The methods mentioned above are ideal for simple distribution, as partitioning the datasets used across multiple machines will not change the end result.

3 Dataset

For our experiments we use various datasets collected through the Twitter Streaming API for a period of three days, starting 14.12.2011 until 16.12.2011 The resulting dataset represents a random sample of 10% of the entire daily activity, thanks to the Gargenhose API account. The data used for the conducted experiments was retrieved from non-protected public accounts. The public accounts post public statuses as candidates for the streaming API, thus we do no need to tackle privacy issues.

After parsing the datastream obtained through the API, we succeed in building dataset as shown in Table 1. The raw datasets adds up to almost 20 million tweets per day and around 60 million tweets for the entire dataset. In order to be able to process the data, we first clean the dataset and kept just the English

Table 1. Data description

Datasets		
Dataset14	Tweets	20.184.280
	Tweets with hashtags	947.815
	Hashtags	1.293.470
	Hashtags distinct	287.091
Dataset15	Tweets	18.543.703
	Tweets with hashtags	877.760
	Hashtags	1.195.910
	Hashtags distinct	267.680
Dataset16	Tweets	20.928.904
	Tweets with hashtags	1.011.717
	Hashtags	1.385.771
	Hashtags distinct	288.542

tweets containing hashtags. Moreover we remove tweets that have been retweeted or which consist of dialogues.

Since our intention is to cluster hashtags we take a look at the structure of hashtags and try to identify certain patterns. According to the presented hashtags we can see that Twitter slang is quite poignant. We can see a top 10 of most frequent hashtags per dataset and their corresponding frequency.

- **Dataset14:** teamfollowback (18910), oomf (16966), np (16938), nowplaying (12303), 2011regrets (11274), idislike (10329), mygoalfor2012 (8837), thingsthataredead (8203), myfavoritethings (7568), jobs (6827).
- **Dataset15:** thingsweallhate (27415), teamfollowback (17091), np (15784), oomf (15265), thingsthatpissmeoff (11451), nowplaying (11350), 1thingiwant4christmas (8476), jobs (6012), mylastwordswillbe(5689), nf(5244).
- **Dataset16:** ff (61469), teamfollowback (20031), np (17202), oomf (15334), nowplaying (12266), ilikeitwhen (10083), jobs (6787), answer (6493), onlyifyouknew (6130), nf (5544).

Almost none of the hashtags follow the classical pattern of tagging with terms. We can see some represent abbreviations like "oomf", while others, entire phrases of concatenated words "mylastwordswillbe".

There are several challenges that we need to tackle in order to analyze properly such a distinct dataset. We first observe how the hashtags are structured.

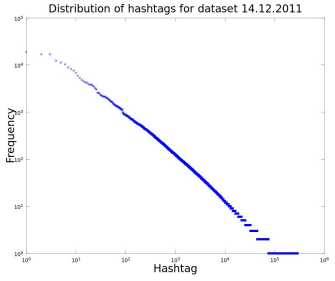


Fig. 1. The distribution of hashtags

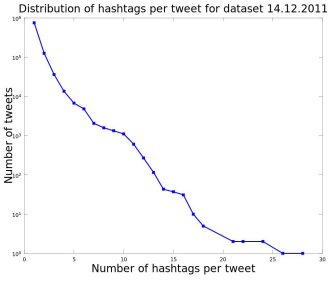


Fig. 2. Hashtags per tweet

Figure 1 represents the distribution of hashtags in the collection. The type of distribution is maintained throughout all daily datasets. We can observe that hashtags follow a *power law*, while a few popular hashtags repeat themselves in the collection a great number of times, a large number of hashatgs have a small frequency. Translated in our dataset, popular hashtags are represented as rich documents, while very less frequent ones have poor documents. The fact that many hashtags have small documents, creates a very sparse vector representation of the dataset.

Figure 2 represents the distribution of hashtags per tweet. We can see that most tweets have a small number of hashtags, and just a few, a large number. After manually inspecting those tweets we come to the conclusion that these last type of tweets represent spam. Users put together several popular hashtags and a shortened URL in order to drive traffic to a web page.

We follow several preprocessing steps. Starting from the JSON set, we parse the tweet text and hashtags within that tweet. We clean the text from: mentions and urls, while also trying to segment hashtags. The resulting dataset is then processed in order to build the inversed associations, namely from hashtag to tweet. For each hashtag we have build a virtual document, consisting of the concatenation of all the tweets in which it was mentioned. This task is solved using the Cascading [25] library over Hadoop. We also assume that each hashtag has an unique representation through its virtual document.

We define $T = \{t_1, t_2, ...t_n\}$ as the set where each t_i is a tweet document and $H = \{h_1, h_2, ...h_m\}$ as the set of all the hashtags in the dataset. A virtual document for one hashtag h_j is a concatenation of tweets as follows:

$$d_j = \sum_{i \in [1,n],\, h_j \in t_i} t_i \quad , \qquad \forall j \in [1, m] \tag{1}$$

4 Experiments

4.1 Preprocessing

For running our experiments we structure our data into files named after each hashtag h_j and containing the text of the corresponding virtual document d_j, according to the definition in (1). We order the files into directories (10.000 files per directory). Dataset14 for example has 287.091 files, namely hashtags represented as virtual documents. In order to proceed to the clustering step we need to represent the text documents as vectors. A common way to do this is to use the vectors space model VSM. The vectorized representation of text in the case of hashtag virtual documents is sparse, mainly in the case of rarer hashtags.

In order to unify hashtags written in different ways, e.g "christmasgift" or "ChristmasGift", they are lowercased and then considered the same hashtag. In the case of "christmas" and "xmas", we expect their virtual documents to be quite similar and the two hashtags grouped together in the same cluster, as variations of the same concept.

For building the vectors, we make feature selection and build a dictionary that better allows us to represent significant words into vectors. We eliminate stop words and some spelling mistakes. Into the dictionary, we put tokens with a minimum document frequency value of 10, which represents the minimum number of documents the term should appear in. These parameter setting helps us prune the dataset from spelling mistakes and very rare words.

Other several preprocessing tasks include tokenization, stop words removal and stemming for the words building the vocabulary. The tasks are accomplished by customizing the Lucene Analyzer class [26]. Using the analyzer we succeed to

remove stop words and words that contain certain patterns, for example three identical characters in a row within a token, such as "aaa", which we treat as spelling mistakes. All tokens passing the filters are lowercased and stemmed according the Porter Stem algorithm [27]. The preprocessing phase is important because it helps to reduce sparsity. If in the beginning without any preprocessing our dictionary had around 500.000 terms, after preprocessing the dictionary for Dataset14 had around 60.000 terms.

The virtual documents corresponding to hashtags are transformed into weighted vectors. We use TFIDF for weighting the terms and 2-norm form normalizing vectors. Thus d_j becomes $\overrightarrow{d_j}$.

4.2 K-Means

K-means [28] is a rather simple but well known unsupervised learning algorithm for clustering. Given a dataset, the algorithm partitions data into a number of clusters. This number of clusters, k, is fixed a priori. The algorithm is divided into the following steps:

1. initialize k points, also known as centroids, randomly chosen from the dataset;
2. assign each virtual document $\overrightarrow{d_j}$ to the cluster having the closest centroid;
3. after all vectors have been assigned, recalculate the position of the centroids as the mean of the points in the relative cluster;
4. repeat steps 2 and 3 until a stop condition is reached or the centroids no longer change.

K-means algorithm minimizes an objective function, we use Jaccard distance measure, which is found to be a suited measure for text in [17]. The Jaccard distance is calculated as follows:

$$DIST_J(\overrightarrow{d_a}, \overrightarrow{d_b}) = 1 - \frac{\overrightarrow{d_a} \cdot \overrightarrow{d_b}}{|\overrightarrow{d_a}|^2 + |\overrightarrow{d_b}|^2 - \overrightarrow{d_a} \cdot \overrightarrow{d_b}}. \quad (2)$$

We also vary the number of clusters, so for each dataset we experiment with k equal to 20, 100 and 500.

The experiments were conducted using the Mahout library over a Hadoop single node cluster installation. This setup allows K-means to run 4 tasks in parallel, 2 map and 2 reduce jobs. Execution time is highly influenced by the number of clusters we wish to produce and can vary from tens of minutes to several hours. We wish to tackle efficiency and speed up problems on multi-node cluster architecture in future work.

A method for determining the best number of k for K-means clustering is by using Canopy clustering [29]. On the other hand, in order to construct canopies we need to set up a minimum and a maximum threshold, values that can themselves become a shortcoming if not set properly. Considering the sparsity of the data and the fact that clusters may not have clear boundaries, there is no need to divide the data in a precise number of clusters. K can be regarded as a measure of granularity of the clustered results, according to which we group hashtags into more specific or more general collections.

5 Results

The results of the clustering show that it is possible to identify semantically related hashtags. For each cluster we extract the top terms, i.e. the most frequent terms in the virtual documents of the cluster. These top terms are the most representative for the cluster, and fulfill their role as explanatory terms. We also extract top hashtags within a cluster, they are obtained by ranking all the hashtags in the cluster by an importance score. This score is computed multiplying the centrality of the hashtag, i.e. the distance from the centroid, by the dimension of its virtual document, that is proportional to the popularity of that hashtag.

For exemplification purposes we will show partial results from Dataset15. We show some sample clusters, for each cluster the top terms and the top hashtags are visualized:

Table 2. Cluster example for Dataset15 with $k = 100$

top terms	december, weather, light, red, degree, middle, warm, blue, green, rain.
top hashtags	buylightmeup, globalwarming, wdisplay, december, wiki, earthquake, climatechange, wheresthesnow, iwantsnow, die.

In the example in Table 2 hashtags and terms are mainly about the weather in middle December 2011, which resulted to be quite warm. We can notice some noise (like "die" or "buylightmeup"), this is due to the small number of clusters, with respect to the high number of topics in the dataset. This is even more notable for $k = 20$, when a large number of different topics are aggregated in a few unique clusters.

Table 3. Cluster example for Dataset15 with $k = 500$

top terms	occupy, ows, wall, street, protest, ndaa, movement, afghanistan, noccupy, st.
top hashtags	ndaa, ows, occupy, occupywallstreet, china, peace, yyc, economy, kpop, washington.

In Table 3 we can see that the tokens of which the top hashtags are composed are often present in the top terms. Some hashtags with ambiguous meaning can be understood reading the relative top terms. For example, in this cluster, clearly related to the popular Wall Street protest, "ows" can be explained with the top terms "occupy", "wall", "street". It is often evident that top hashtags are understandable themselves, this is because they are the most popular hashtags, the easiest to compose or read.

Figure 3 shows evaluation results for clustering Dataset14 using different values for k. The evaluation measures are the average *Inter-cluster* distance and the

average *Intra-cluster* distance, calculated according to Jaccard distance measure presented in (2). The first measure describes how well data is separated, computing the average of the distances between the centroids, higher values mean better separation. The second measure defines the average distance between points in the same cluster, lower values mean higher density in the cluster and a better separation. We can see that both measures tend to stabilize themselves for $k > 100$ at an average of 0.95 for Inter-cluster distance and 0.75 for Intra-cluster distance, and maintain this trend up until $k = 1000$. For $k = 20$, $k = 40$ and $k = 60$, the Inter-cluster distance has lower values: 0.600, 0.621 respectively 0.716, meaning clusters are not very well defined, but it increases as the number of clusters grows. The Intra-cluster distance rises slowly from 0.673 to a maximum of 0.757, meaning the sparsity of cluster points within each cluster has little variation.

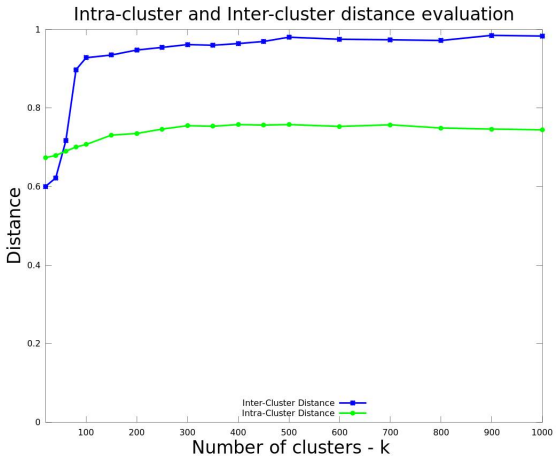


Fig. 3. Evaluation of K-means for Dataset14 by varying k

For a small k, i.e. $k = 20$, the size of the clusters varies quite a lot, some are larger while others are smaller, whereas for a larger number of k, i.e $k = 500$ or $k = 1000$, we notice a uniform distribution of cluster sizes.

Another conclusion that can be drawn from our experiments is that, for such a large dataset, results improve as the k increases, i.e. $k = 100$ or $k = 500$. The precision of semantic association between top hashtags and top terms increases because clusters becomes more specialized. Exploring and tuning the k parameter will bring to a better comprehension of the hashtags distribution.

6 Conclusions and Future Work

In this paper we experiment with clustering and Twitter hashtags. We describe a hashtag as the concatenation of tweets in which it appears. Working on large

datasets and with distributed clustering algorithms we have obtained interesting results about the semantic association of hashtags. We can see a clear connection between hashtags and the top terms of the cluster. These results can bring future improvements and applications of our idea, refining the tuning of the algorithm and experimenting with even more options and number of clusters. We are also planning to use a larger dataset, that we have already extracted, which consist of one week of tweets.

By clustering hashtags we succeed in making an unsupervised classification into flexible groups that are not constrained by a target class. In the authors' opinion a traditional classification task into predetermined topics would not suffice in respect with the variety of constantly changing daily chatter. Topics in Twitter are divided into long term topics and short term topics, usually generated by important news. A topic can be described with the help of several hashtags and viral terms grouped together according to similarity.

The main applications of the obtained research results are hashtag prediction and recommendation tasks. Clustering is useful for restricting the search base for the recommendation candidates. Searching through all the dataset can be a time consuming task, while searching within clusters can reduce this workload. If we wish to have a small group of candidates to suggest from, the granularity of the clustering task must be high (large k). If we wish to recommend more general hashtags, we can cluster the dataset into somewhat bigger groups (small k). Hashtags in clusters are ranked according to their frequency.

The results presented above can be later refined through NLP techniques in order to discover synonyms, antonyms etc. As previously mentioned, the smaller the cluster size, the more concise and specific the topic, the stronger the connection between hashtags. Clustering captures the co-occurrence of terms and hashtags. Once the grouping is accomplished, one can apply named entity recognition models in order to discover related entities or sentiment analysis techniques for finding opinions.

In future works we wish to perform hierarchical clustering on hashtags with the purpose of bringing them closer to a taxonomy. Several algorithms and machine learning techniques, like classification or recommendation of hashtags, could be applied to hashtag virtual documents in order to extract or suggest useful information. One possible application could be a tool for generating a human readable explanation of the meaning of a hashtag, using the top terms in the clusters. Comparing hashtags can be useful for suggesting the most popular hashtags to a user, in order to help increase the popularity of his tweets.

References

1. Hannon, J., Bennett, M., Smyth, B.: Recommending twitter users to follow using content and collaborative filtering approaches. In: Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys 2010, pp. 199–206. ACM, New York (2010)
2. Kireyev, K., Palen, L., Anderson, K.: Applications of Topics Models to Analysis of Disaster-Related Twitter Data (December 2009)

3. Phelan, O., McCarthy, K., Smyth, B.: Using twitter to recommend real-time topical news. In: Proceedings of the Third ACM Conference on Recommender Systems, RecSys 2009, pp. 385–388. ACM, New York (2009)
4. Weng, J., Lim, E.P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, WSDM 2010, pp. 261–270. ACM, New York (2010)
5. Romero, D.M., Meeder, B., Kleinberg, J.: Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In: Proceedings of the 20th International Conference on World Wide Web, WWW 2011, pp. 695–704. ACM, New York (2011)
6. Gruhl, D., Guha, R., Liben-Nowell, D., Tomkins, A.: Information diffusion through blogspace. In: Proceedings of the 13th International Conference on World Wide Web, WWW 2004, pp. 491–501. ACM, New York (2004)
7. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter: understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, WebKDD/SNA-KDD 2007, pp. 56–65. ACM, New York (2007)
8. Thomas, K., Grier, C., Song, D., Paxson, V.: Suspended accounts in retrospect: an analysis of twitter spam. In: Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, IMC 2011, pp. 243–258. ACM, New York (2011)
9. Wang, A.H.: Dont't Follow me: Spam Detection in Twitter. In: Proceedings of the International Conference on Security and Cryptography (SECRYPT) (July 2010)
10. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World Wide Web, pp. 591–600. ACM (2010)
11. Chen, J., Nairn, R., Nelson, L., Bernstein, M., Chi, E.: Short and tweet: experiments on recommending content from information streams. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, pp. 1185–1194. ACM, New York (2010)
12. Ellen, J.: All about microtext - a working definition and a survey of current microtext research within artificial intelligence and natural language processing. In: ICAART (1) 2011, pp. 329–336 (2011)
13. O'Connor, B., Krieger, M., Ahn, D.: TweetMotif: Exploratory Search and Topic Summarization for Twitter. In: Cohen, W.W., Gosling, S., Cohen, W.W., Gosling, S. (eds.) ICWSM. The AAAI Press (2010)
14. Xu, T., Oard, D.W.: Wikipedia-based topic clustering for microblogs. Proceedings of the American Society for Information Science and Technology 48(1), 1–10 (2011)
15. Phan, X.H., Nguyen, L.M., Horiguchi, S.: Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In: WWW 2008: Proceeding of the 17th International Conference on World Wide Web, pp. 91–100. ACM, New York (2008)
16. Pennacchiotti, M., Popescu, A.M.: A machine learning approach to twitter user classification (2011)
17. Rangrej, A., Kulkarni, S., Tendulkar, A.V.: Comparative study of clustering techniques for short text documents. In: Proceedings of the 20th International Conference Companion on World Wide Web, WWW 2011, pp. 111–112. ACM, New York (2011)
18. Dean, J., Ghemawat, S.: Mapreduce: simplified data processing on large clusters. Commun. ACM 51(1), 107–113 (2008)

19. Hadoop, http://hadoop.apache.org
20. Papadimitriou, S., Sun, J.: Disco: Distributed co-clustering with map-reduce: A case study towards petabyte-scale end-to-end mining. In: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, ICDM 2008, pp. 512–521. IEEE Computer Society, Washington, DC (2008)
21. Shvachko, K., Kuang, H., Radia, S., Chansler, R.: The hadoop distributed file system. In: IEEE/NASA Goddard Conference on Mass Storage Systems and Technologies, pp. 1–10 (2010)
22. Apache Mahout, http://hadoop.apache.or
23. Wegener, D., Mock, M., Adranale, D., Wrobel, S.: Toolkit-based high-performance data mining of large data on mapreduce clusters. In: Proceedings of the 2009 IEEE International Conference on Data Mining Workshops, ICDMW 2009, pp. 296–301. IEEE Computer Society, Washington, DC (2009)
24. Chu, C.T., Kim, S.K., Lin, Y.A., Yu, Y., Bradski, G.R., Ng, A.Y., Olukotun, K.: Map-reduce for machine learning on multicore. In: Schölkopf, B., Platt, J.C., Hoffman, T. (eds.) NIPS, pp. 281–288. MIT Press (2006)
25. Cascading, http://www.cascading.org/
26. Lucene, http://lucene.apache.org/core/old_versioned_docs/versions/3_0_1/api/all/org/apache/lucene/analysis/Analyzer.html
27. Willett, P.: The Porter Stemming Algorithm: Then and Now. Program: Electronic Library and Information Systems 40(3), 219–223 (2006)
28. MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: Cam, L.M.L., Neyman, J. (eds.) Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281–297. University of California Press (1967)
29. McCallum, A., Nigam, K., Ungar, L.: Efficient clustering of high-dimensional data sets with application to reference matching. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 169–178. ACM (2000)

SMART: Modeling and Monitoring Support for Business Process Coordination in Dynamic Environments*

Piergiorgio Bertoli[1], Raman Kazhamiakin[1,2], Michele Nori[1], and Marco Pistore[1,2]

[1] SAYservice srl, via alla Cascata 56/C, 38123, Trento, Italy
[2] Fondazione Bruno Kessler, via Sommarive 18, 38123, Trento, Italy
{bertoli,raman,nori,pistore}@sayservice.it,
{raman,pistore}@fbk.eu

Abstract. In several relevant applicative domains, handling a high-level business process requires coordinating a variety of fragmented software systems, and coping with complex dynamics that crucially depend on non-controllable events and actions, also coming from external services. In these cases, the effective management of the process stands on the ability to continuously identify, aggregate, and correlate the relevant events from a variety of independent sources, and to timely enact coordination actions. This is a significant challenge for the existing modeling and monitoring solutions.

In this paper, we discuss *SMART*, a solution designed to support operative business process management under these premises. *SMART* offers a flexible and agile modeling approach, where external services, information sources and proprietary support systems systems can be meaningfully and smoothly integrated according to business logics, to result into "operative consoles" which can be used as effective coordination tools. We demonstrate the key features and advantages of *SMART* through a sample application to support conference organization.

1 Introduction

High-level organizational processes typically involve a variety of actors, and managing their evolution is an inherently complex task which requires tracing and correlating several activities. This is even more true in the large set of applicative fields, including e.g. logistics, production chains, or handling of large-scale events such as e.g. large conventions, massive sport or cultural happenings, or Civil Protection scenarios, where processes are affected by non-controllable choices. Such choices can be taken by single actors, belonging to different categories involved in the process, through independent, 3^{rd}-party services.

In spite of ever-increasing availability of software products to support domain experts in handling specific process phases, such complex business processes are still only supported in a fragmented way, where different uncorrelated software (sub-)systems take

* This work is supported by project "*SMART* 4U: User-Centric Platform for the Internet of Services", funded by the Operational Programme "Fondo Europeo di Sviluppo Regionale (FESR) 2007-2013" of the Province of Trento.

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 243–254, 2012.

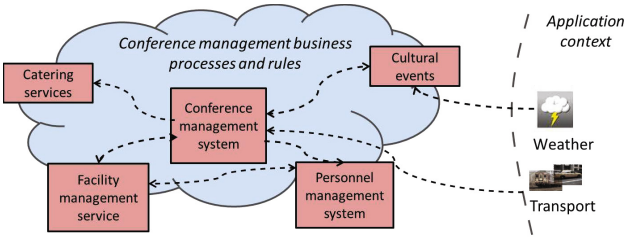


Fig. 1. Conference scenario

care of different aspects or phases. This is natural and unavoidable, since in general such processes are highly customized and integrate a variety of disparate aspects. This is evident even in a smaller-scale conference management scenario, which we will conveniently take as a reference throughout the paper. Organizing a conference involves a variety of activities that include paper submission and evaluation, conference registration, management of the program, management of conference facilities and equipment, assignment and coordination of the supporting staff, catering services, social events, and so on. In turn, as shown in Fig. 1, these different activities are supported by a set of software systems, either general-purpose (e.g., a conference management system such as EasyChair, see http://www.easychair.org) or specific and proprietary (e.g., the venue room-allocation facility, or the personnel management system). None of such systems are thought to interact with any other, since they are not designed having in mind an overarching process of conference management. Attempting to integrate all these systems into a unique and universal application for conference management would indeed fail due to the specific features pertaining to different conferences in different organizational set-ups and for different venues. In addition to this, as also shown in Fig 1, independent 3^{rd}-party services take a significant role in such a scenario, since actors involved in the process may adopt them for executing activities relevant and related to the process. For instance, conference participants typically organize their travel via online providers; then, delays or cancellations about trips of speakers may be brought by services of travel companies.

In this kind of situations, managing the execution of the top level process, which in fact is left implicit as a combination of different and partially codified fragments expressed by different systems, is highly difficult, error-prone and time-consuming, even for a domain expert. Not only it requires substantial domain expertise, but also constant manual aligning and coordination among different activities and software systems, including independent services of which the expert may not be directly aware. At the same time, the fact that activities are carried out through technological gateways opens the possibility of adopting information technologies to analyze and correlate such activities, interacting with the relevant systems and services to coordinate them in the frame of a meaningful and coherent overall process.

The *SMART* proposal aims at exploiting such a possibility, in order to provide operative coordination support for process management under these conditions, taking on a significant conceptual and technological challenge. It does so by offering a coherent approach based on a clear methodology and a corresponding support toolset which:

1. supports domain experts in the agile, cost-effective and non-intrusive integration and orchestration of different independent software systems and services.
2. provides simple yet expressive business domain modeling facilities to capture the key domain elements, their mutual and dynamic relations and impacts, as well as their connection to the underlying systems and services.
3. provides a run-time environment realizing the above model in order to continuously observe relevant phenomena, propagate them to the different elements, notify them as appropriate to human process managers, and implement coordination activities by orchestrating services and systems.
4. provides suitable, easily extensible, front-ends that represent the domain concepts, their evolution, the relevant situations and coordination actions associated to them.

The *SMART* proposal realizes the above by providing (a) a formal tool to model (the dynamics of) activities, and to connect/aggregate/monitor them according to specific orchestration logic; (b) powerful software tools to integrate technologically heterogeneous services and services, and associate them to such activity models, to achieve (c) the ability to monitor and coordinate, by operative consoles, the evolution of complex processes. In particular, the formal core of *SMART* stands in the notion of Domain Object, used to uniformly represent services, user interfaces and domain-specific concepts at different levels of abstraction.

In this paper, we discuss *SMART*, describing its architecture and its core concepts, and showcasing its key features via a sample application meant to support the organization of a conference. The paper is organized as follows. Section 2 describes the conceptual architecture behind *SMART*, while Section 3 discusses its core element, Domain Object model. In Section 4, we describe the concrete application of *SMART* based on the reference scenario; finally, Section 5 presents conclusions and related work.

2 *SMART* Approach: Architecture

SMART supports, both at design and execution time, the integration and coordination of 3^{rd}-party services and proprietary process support applications. The conceptual architecture of the *SMART* platform is represented in Fig. 2, and is structured into an integration, a service model, and a presentation layers.

Specifically, *SMART*'s connection to 3^{rd} party services and proprietary subsystems is realized by the abstraction layer provided by its *Integration* layer. Its task is to compensate the technological and conceptual heterogeneity of the services and systems, providing mechanisms for service invocation as well as for monitoring, contextualizing, filtering and proactively proposing extracted information. This is achieved through the two specific *XS** and *DF* languages, and corresponding compilers and validation tools. The *XS** notation is a proprietary markup language which is used to cope with the wrapping and extraction of information out of loosely-structured information sources, such as e.g. HTML pages. This is fundamental to cope effectively with the vast amount of web-available services. The complementary *Data Flow* (*DF*) language provides a form of enterprise mash-up notation (c.f. [1]), enabling aggregation, transformation, and integration of different structured information sources (including *XS** wrappings) and

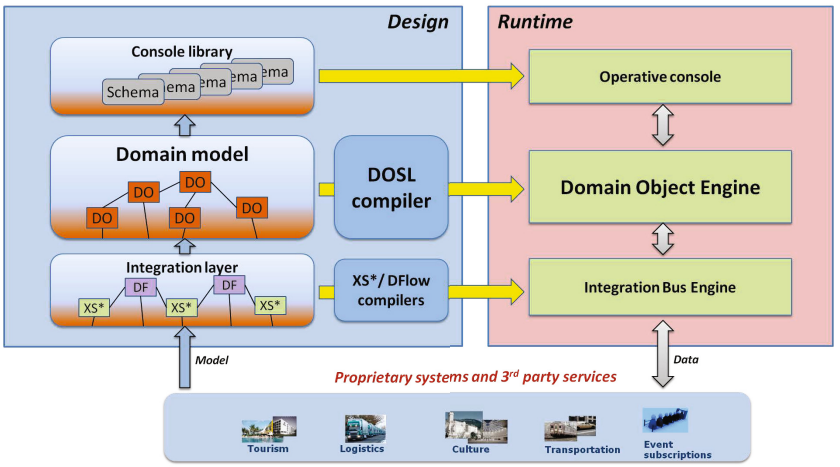


Fig. 2. Conceptual architecture

protocols. *XS** and *DF* can be linked to adapters for standard Web services and components in order to incorporate a wide range of heterogeneous systems and services, via a run-time integration bus.

Central to the platform stands the *Domain model* layer, which provides the model and run-time environment for service integration, monitoring, and processing, and enables different forms of service composition: information integration, events aggregation, orchestration of transactional services. These features are supported by the key conceptual Domain Object element. Domain Objects are the uniform formal tool that *SMART* introduces to model services and relevant domain concepts, in a hierarchy that spans different levels of abstraction and goes up to concepts that are directly connected to the management of the top-level process and can be associated to operative consoles. The Domain Object model is designed to allow representing, in an agile, partitioned and flexible way, the specific domain logic for a business process, through the combination and orchestration of elements, and to faithfully model the key phenomena related to the production of events and actions by the various elements in the play. Further, the model is carefully designed to optimize the usage of distributed computational resources.

Concretely, Domain Objects are specified via a rich Domain Object Specification Language (*DOSL*), which is connected with the *DF* integration language through the sharing of events and actions, and allows a designer to represent a dynamic network of domain objects in terms of their state, relations to other objects, and their state evolution in reaction to events and commands propagated from related objects. For this purpose, the language integrates event-driven and procedural programming styles. A specific *SMART* domain-object runtime engine is used to execute the artifacts resulting from compiling a *DOSL*-specified network of objects.

At the top level, the behavior of the domain, and the progress of the overall process, are traced via operative consoles. *SMART* makes it simple to develop such elements by (a) providing a range of predefined data structures and supporting libraries for

commonly used concepts and representation channels, including geographical data and maps, calendar events, notification messages, and (b) having *DOSL* offer a neat programming interface that represents in a clear way all the phenomena that are relevant to monitor and coordinate the activities of the process. In particular, the interfaces cover event dispatching and notification, access to the state of objects, access to the relation among objects, and management of commands provided by the consoles.

The next section details the core element of this architecture, Domain Objects.

3 Domain Objects

Domain objects are the essential elements of the framework, as they uniformly represent the relevant concepts for a complex business process whose realization stands on top of a combination of software systems of various kinds, including 3^{rd}-party services. This implies significant expressiveness constraints since, at the ground level, they need to map services and systems of various nature and content. Further, the model must be adequately structured, since Domain Objects must represent low-level services and systems as well as high-level concepts apt to be presented to domain analysts and operators, and as well as the intermediate domain-specific concepts links such different abstraction levels.

In particular, domain objects are organized into a directed acyclic type hierarchy whose structure reflects the abstraction levels in the domain at hand. The hierarchy is connected on one side to the infrastructural layer of systems and services, and on the other side must offer suitable interfaces towards operative consoles. Hence, a business domain instance is represented by a dynamically evolving instantiation of domain object types, and of their relationships. The evolution of such a domain is driven on one side by the bottom-up propagation of events triggered by the underlying systems and services, and on the other, by the top-down transferal of actions that can be enforced in response to events, either via business rules or through the operative consoles.

In this respect, the Domain Object model must obey strict requirements that guarantee an appropriate runtime behavior. In particular, the model must be strongly reactive in handling events from systems and 3^{rd}-party services, immediately updating the whole model, notifying relevant changes and updating the operative consoles: since events represent changes in the "environment" of systems and services, no (even temporary) misalignment can be afforded, as it might lead to present wrong information and take incorrect coordination actions. At the same time, the model must avoid over-constraining the run-time execution: while a model of sequential, non-interruptible bottom-up propagation of events interleaved with a lazy top-down propagation of commands is easy to devise and to prove to satisfy the above reactiveness requirement, this would result in a very ineffective usage of computational resources, in a setting where normally sets of independent concurrent events take place over distributed computational nodes.

This is in fact one of the key challenges for the formal model, whose solution calls for a formally secured yet maximally flexible and unconstraining mechanism to ensure the correct and unambiguous handling of concurrency. To solve this hard requirement, the model stands on a notion of *event bundle*, meaning a set of propagating events stemming from a common triggering event, and on allowing concurrent propagation of bundles

by enforcing easy-to-check syntax-based conditions to prevent incorrect interferences between concurrent bundles.

In a nutshell, the Domain Object definition is based on modeling four key phenomena and features:

1. **Relations.** Domain objects form a network according to directed access relations that derive from their type hierarchy. Relations among objects may be many-to-many, and evolve dynamically on the basis of the events and actions, and of the state of objects. This allows for instance quantifying that "an instance of concept X is related to all instances of concept Y that obey a certain property".
2. **Events.** Events are the mean, for an element of the model, to notify changes to hierarchically higher elements, possibly triggering their reaction. Typically, events generated by services cannot be fully controlled or predicted, neither in terms of timing nor of content. Also, services are designed to generate events regardless which entities might decide to use them. Hence, following standard methodologies, the mechanism for propagating events bases on a *publish/subscribe criterion*. Inside an object, an event is handled by a non-interruptible event-handling procedure.
3. **Commands.** Commands allow the transportation of information between an element of the model and a hierarchically inferior element accessed by it, possibly triggering a reaction of the latter. The mechanism of communicating commands differs from that of events in two relevant aspects. First, commands are meant to be point-to-point among object instances; second, the handling of a command in an object takes place via an interruptible procedure.
4. **Object Creation.** The universe of existing objects may evolve due to some object that "spawns" other objects, due to reacting to events, reacting to commands or even simply by enforcing some internal logics. As such, object creation may happen in different phases of the execution of the framework. The formal model imposes certain limitations in the order in which objects can be created, in order not to perturb the event propagation mechanism.

The above elements define the context for the domain object model sits, to which services, systems and consoles connect by means of appropriate interfaces.

More in detail, the formal model of the domain object layer is constituted by a stack of models of decreasing level of abstraction, such that the higher model closely matches the linguistic elements of the *DOSL* counterpart. The stack starts from a set-theoretic model where the key communication elements are perceived as atoms, and domain objects are represented as finite-state machines operating on them; then it steps up to a model where variables and signatures appear as fist-class citizens; and then introduces relations and state visibility as objects of the model. This allows declining semantically defined correctness criteria given in the set-theoretic into easy-to-check (sufficient) syntactic criteria at the higher level. Due to lack of space, we omit the detailed formal discussion of the model, leaving it to the technical report available at [2], and provide a more conceptual discussion of the model. In a nutshell, domain objects can be understood as a form of extended, typed, event-driven state automaton, which at the top level of the formal stack, is based on variables and relations:

Definition 1 (Domain Object Type). *A domain object type is a 8-ple* $\langle SVS, VSVS, SVS_0, E, E_0, C, REL, \mathcal{R}\rangle$ *where:*

- *SVS defines the states of the automaton as a set of pairs representing typed variables. Its subset* $VSVS \subseteq SVS$ *represents the variables visible to other objects, hence defining the visible portion of the state. The initial states are defined by the set of variable-value pairs* SVS_0*, covering all variables in SVS.*
- *E defines the events that can be produced in output, via a set of typed event signatures.* $E_0 \subseteq E$ *defines, in particular, a set of events which are fired upon the creation of an object instance.*
- *C defines which commands the object is capable to handle, through a set of typed command signatures;*
- *REL defines a set of object types that can be accessed by the current type;*
- $\mathcal{R} = \langle \mathcal{R}_E, \mathcal{R}_C, \mathcal{R}_\tau \rangle$ *defines the dynamics of the object type as a partitioned relation split into event-driven dynamics (*$\mathcal{R}_E$*), command-driven dynamics (*$\mathcal{R}_C$*), and internal dynamics (*$\mathcal{R}_\tau$*). All of* $\mathcal{R}_E$*,* $\mathcal{R}_C$*,* $\mathcal{R}_\tau$ *operate by (a) reading the variables in SVS and the visible variables of accessed objects, (b) updating the variables in SVS, (c) generating events and commands, (d) creating new objects, and (e) modifying the set of accessed objects. In a nutshell,* $\mathcal{R}_E$ *is a set of event rules each of which defines a reaction to some set of events, based on the current state and the information brought by the events;* $\mathcal{R}_C$ *is a set of command rules, each defining a reaction to a single command, again based on the current state and the information brought by the command, and* $\mathcal{R}_\tau$ *is a set of internal rules, each defining an internal progression based purely on the current state.*

On this basic formalization, the model enforces a set of additional well-formedness criteria, for instance to inhibit deadlocks and livelocks of sets of objects; the reader can refer again to [2] for details.

Given this setting, a domain object instance (or *object*) is a 3-ple ⟨D,s,I⟩, associating a domain object type D to a valuation s of its state and a set I of currently related objects. The execution snapshot of a domain is then captured by a notion of "universe configuration", which not only amounts to a set of objects, but also associates sets of propagating events stemming from a common triggering event, called *event bundles*, to sets of objects for which they are relevant.

Definition 2 (Universe configuration). *A universe configuration is defined as a 4-ple* $\langle D, H, I, B\rangle$ *where D is a set of Domain Object types,* $H \subseteq D \times D$ *is a relation defining the type hierarchy, I is a set of domain object instances over D, and B is a set of bundles, that is sets of pairs which associate an event* e *to an object* i *meant to handle that (i.e. for which* e *is in the set of its input events* $iEventsOf(i)$*):*

$$B := \{b \mid b = \{\langle e, i\rangle \mid i \in I \wedge e \in iEventsOf(i)\}\}$$

The correct execution of the model to respond to the requirement of immediate handling of events amounts to guaranteeing that each bundle is propagated with no interference from commands or other bundles. At its basic level, the model enforces this in a straightforward way by prioritizing events over commands, and by defining a "sequentialized"

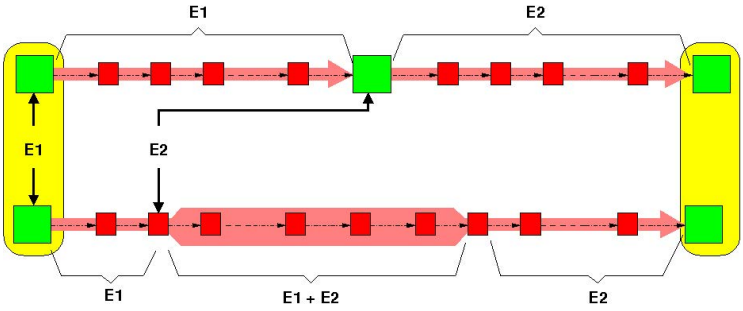


Fig. 3. Convergence of interleaved and sequential execution models

runtime model where (1) "stable" event-free configurations can evolve on the basis of commands, incoming events, or internal logics of objects, while (2) "unstable" configurations are not visible outside the runtime, and evolve only to cope with the pending events until all of them (and the events they create) are consumed.

However, this simple reference model implies that the progression of a bundle "locks" the whole universe for what concerns the progression of any other concurrent bundle; this is an obvious bottleneck with critical performance implications. Hence the model needs to relax this to achieve a situation where concurrent bundle propagation is possible, yet the interleaved propagation of bundles converges to their sequentialized propagation, as graphically depicted in Fig. 3: in the top part, the sequential model fully propagates event E1 prior to handle event E2 (received meanwhile), while in the bottom part, the propagation of E2 takes place concurrently with that of E1. While they traverse different sequences of unstable configurations, they get to the same visible result.

This is achieved, at the semantic level, by a *semantic criterion for interleaving* that allows concurrence insofar any pair of bundles that traverse a common set of objects are executed in the same order on all such objects, and commands are handled in a deferred way (so that a bundle may trigger other bundles only upon its complete propagation). This criterion is proved to satisfy the above convergence requirement w.r.t. the reference sequentialized model. The criterion is then further refined to ease its checking, by standing on two key elements: (a) the representation of the object through variables, which allows syntactically checking whether two bundles interfere on portions of an object, and (b) the choice of a read-commit memory model, where the propagation of each bundle takes place in a private memory space, created at the start of propagation ("read") and copied back to the global memory upon the end of bundle propagation ("commit"). This leads to the following sufficient *syntactic criterion for interleaving*: upon committing bundle B, consider every currently active bundle B', and check whether B is writing on the input variables of B'; if this is the case, discard and restart the execution of B'. This criterion has the substantial advantage of requiring small amounts of computation and at definite points in time; in counterpart, it may require discarding some bundle executions. Again, the key result in this respect is that the same formal guarantees of execution convergence w.r.t. the reference sequential model, shown in Fig. 3, holds.

4 SMART Platform

The *SMART* platform realizes the conceptual architecture represented in Fig. 2, providing suitable compilers for the *DOSL*, *XS** and *DF* languages, and standing on a Java-based run-time execution environment that is organized following Service-Oriented Architecture principles (SOA, [3]).

The key components of the platform, namely service integration bus and domain object container, expose Web service interfaces, enabling Cloud-based deployment. The service orchestration procedures are deployed and executed using the Activiti BPM engine (http://activiti.org/). Further, the use of OSGi technology [4] and of a persistence layer based on non-relational, JSON-based database mongoDB (http://www.mongodb.org/) support dynamic data structures and on-the-fly deployment of the domain object and service models. Using these technologies, *SMART* provides an integrated engine that realizes the execution semantics described in the framework, ensuring an effective yet correct dispatching of events and operator-driven commands, and enabling the monitoring and coordination of the systems and services at hand, in the frame of an overarching business logics.

We now discuss the usage of *SMART* to realize an application that supports operative process management in the reference conference organization scenario and deals with the challenges presented in such scenario.

4.1 A Sample Application

We have used *SMART* to implement an application that aims to support the realization and coordination of various activities for handling a conference, as presented in Section 1, referring specifically to an international conference held in Trento, Italy. The overall reference process includes conference program management (schedule presentations, assign presenters, notify upon modifications), support for the logistics of the presenters (e.g. their trips and hosting), tracking and allocation of resources (appropriately sized rooms, technical equipment and personnel). Hence, in our application, each of these activities must be tracked through the operative console, which must incorporate the three different aspects regarding program, travels, and conference resources. The console must continuously inform the manager about relevant situations and events, highlighting specific problems on specific entities via a "traffic light" metaphor.

To realize these functionalities, the application relies on a set of proprietary systems and services, including in particular a general-purpose conference management system and a venue management facility provided by the University of Trento. The conference management system handles information about the conference program, the presentations and their constraints (the equipment required and the number of participants), as well as some information about presenters and their travels (i.e., from where, number of flight and train if applicable, or whether the car is used). These information are not sufficient to properly support logistics of speakers, and as such, to keep track of transportation means and to find alternatives, we have to incorporate a set of web-available 3^{rd}-party services: a train schedule "ViaggiaTreno" (http://viaggiatreno.it/), a real-time flight status service "FlightStats" (http://www.flightstats.com), and the on-line schedule of Verona airport, which serves Trento (http://www.aeroportoverona.it/).

Fig. 4. Domain object specification excerpt

To exemplify how the application works, consider the critical situation in handling the logistics side, taking place when a presenter faces a transportation problem (e.g., due to delays or strikes) and is not able to arrive on time. Depending on the specific case, this may require finding an alternative transportation or rescheduling the specific presentation. In the latter case, this in turn implies a chain of actions such as releasing and/or reallocating resources (rooms, personnel) and notifying the participants. The application, therefore, has to monitor such situations, relate them to the other elements, represent them in the operative console, and provide the realization of the necessary coordination actions. Our application deals with the above situation by having the domain object model, among the domain concepts, also those related to logistics: speaker travels, and transports. Low-level concepts are linked to the above-mentioned services and systems. An excerpt of the specification is represented in Fig. 4. In particular, it declares a Presentation object related to a Trip object that aggregates information about transport means of the speaker, such as train and flight, and their status. The information extracted from the train schedule service is propagated to individual Train objects, updating the corresponding properties. Realizing a business logics rule, whenever a train arrival time is delayed by more than 15 minutes, the Trip object publishes the "delayed" event. Consequently, if the trip delay becomes critical for the presentation schedule, the status of the presentation becomes "red". Through the console, the manager gets notified about this critical situation, and can handle it can triggering a "changeSlot" operation provided by the Presentation object (and realized as composite procedure).

The operation console is implemented (Fig. 5) on top of the domain specification, providing the three views related to program schedule ("Programma"), presenter logistics ("Viaggi"), and facilities ("Sale"). Each of them reports the information and state of different objects (presentations, transportation, rooms and equipment respectively), as well as the list of relevant events, categorized into "information" and "alert" events (of which the latter are critical and requiring some reaction). Fig. 5 (left), shows some alerts representing the delays of the speakers due to transport problems, and the proposed actions related to changing the time of presentations. Similarly, Fig. 5 (center) shows the state of the speaker travels using the map, where the travel data (flight, train), their status (position with respect to the destination and the "traffic light"), and the service information (e.g., current timetable of the Verona train station and the speakers

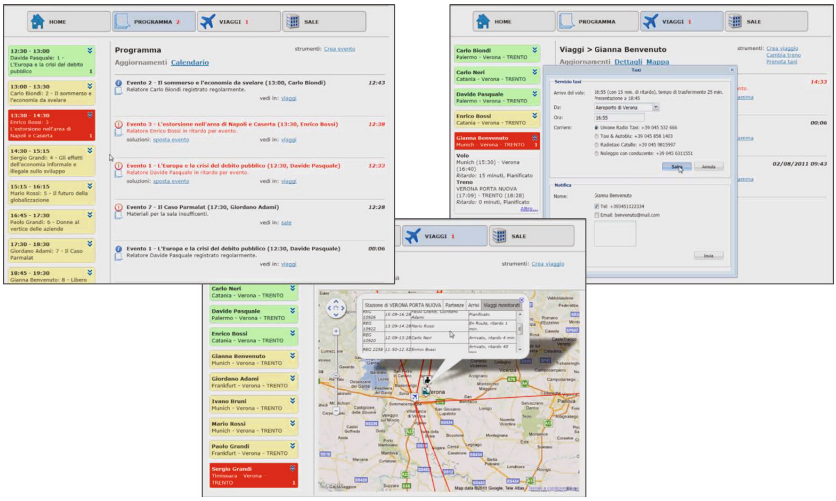


Fig. 5. Management console

expected to travel through the station) are represented. The supporting activities may be invoked directly from the console: Fig. 5 (right) shows the form corresponding to the taxi booking activity for the speaker as an alternative to the train trip, when the flight has been delayed. The modeling of this application has required very limited effort and results in a compact and flexible specification, which provides effective operative support and demonstrates how *SMART* can achieve its goals for a non-trivial scenario.

5 Related Work and Conclusions

In this paper we presented *SMART*, a platform that supports integration and coordination of systems and services within a single business domain, providing necessary means for monitoring, propagating, and representing relevant business events and their impact on the domain artifacts. For this purpose the platform relies on a formal domain objects model, realized through a set of specification notations and a SOA-based runtime environment. We also demonstrated how the platform may be exploited to address process monitoring and management support in a non-trivial conference organization sample scenario.

The problem of monitoring business processes and services is widely addressed in the literature and in industrial realizations. In particular, Business Activity Monitoring (BAM, [5]) allows for near real-time monitoring (identifying trends, calculating Key Performance Indicators, evaluate functional and non-functional properties of the process instances) (c.f., [6,7,8]). Using the Complex Event Processing technologies (CEP, [9]), it is possible to capture complex event patterns regarding a single or multiple executions of a business process. When applied to SOA, monitoring refers to the problem of verifying functional and QoS properties of BPEL orchestrations [10,11] or individual services [12]. While these approaches address the “technical” aspects of executions

(e.g., service quality, process performance) or focus on the activities and events of the overall process, the *SMART* approach targets the problem of coordination and alignment between different elements of the domain. Moreover, the underlying formal model directly supports this vision, ensuring that the information is correctly propagated, which is not the case for the above solutions.

Also close to the vision of the domain object coordination is the approach adopted by the case handling systems [13]. Specifically, the approaches based on artifact-based paradigm [14], allow for associating the activities and rules to a business artifact, assisting the business worker in achieving specific business goals associated to the artifact. It appears then possible and promising to complement such an approach and system with the SMART-provided capability to capture the coordination policies among different artifacts, to ensure their coordinated evolution continuously monitoring and propagating relevant events, and to realize composed coordination activities.

References

1. Crupi, J., Warner, C.: (Enterprise Mashups: The New Face of Your SOA) SOA World Magazine, http://soa.sys-con.com/node/719917 (retrieved March 22, 2012)
2. Bertoli, P., Kazhamiakin, R., Pistore, M.: Formal Model of Domain Objects. Technical report, SayService s.r.l. (2012), http://sayservice.it/docs/TR-D21-2012.pdf
3. OASIS: (Reference Model for Service Oriented Architecture 1.0), http://www.oasis-open.org/committees/download.php/19361/soa-rm-cs.pdf
4. OSGI Alliance: (Osgi service platform core specification) Release 4, Version 4.3 (April 2011), http://www.osgi.org
5. Dresner, H.: Business activity monitoring: Bam architecture. In: Gartner Symposium ITXPO, Cannes, France (2003)
6. Beeri, C., Eyal, A., Milo, T., Pilberg, A.: Monitoring Business Processes with Queries. In: Proceedings of the 33rd International Conference on Very Large Data Bases, pp. 603–614 (2007)
7. Chowdhary, P., Bhaskaran, K., Caswell, N.S., Chang, H., Chao, T., Chen, S.K., Dikun, M., Lei, H., Jeng, J.J., Kapoor, S., Lang, C.A., Mihaila, G., Stanoi, I., Zeng, L.: Model Driven Development for Business Performance Management. IBM Syst. J. 45, 587–605 (2006)
8. Momm, C., Malec, R., Abeck, S.: Towards a Model-driven Development of Monitored Processes. Wirtschaftsinformatik 2 (2007)
9. Verner, L.: The power of events: An introduction to complex event processing in distributed enterprise systems. Addison-Wesley Professional (2002)
10. Ghezzi, C., Guinea, S.: Run-Time Monitoring in Service-Oriented Architectures. In: Baresi, L., Nitto, E.D. (eds.) Test and Analysis of Web Services, pp. 237–264. Springer (2007)
11. Baresi, L., Guinea, S., Pistore, M., Trainotti, M.: Dynamo + astro: An integrated approach for bpel monitoring. In: IEEE International Conference on Web Services, ICWS 2009, Los Angeles, CA, USA, July 6-10, pp. 230–237 (2009)
12. Keller, A., Ludwig, H.: The WSLA Framework: Specifying and Monitoring Service Level Agreements for Web Services. J. Network Syst. Manage. 11 (2003)
13. van der Aalst, W.M., Weske, M., Gruenbauer, D.: Case handling: A new paradigm for business process support. Data and Knowledge Engineering 53, 2005 (2005)
14. Bhattacharya, K., Caswell, N.S., Kumaran, S., Nigam, A., Wu, F.Y.: Artifact-centered operational modeling: lessons from customer engagements. IBM Syst. J. 46, 703–721 (2007)

Towards an Intelligent Framework to Understand and Feed the Web

Anna Fensel[1], Julia Neidhardt[2], Nataliia Pobiedina[2], Dieter Fensel[1], and Hannes Werthner[2]

[1] Semantic Technology Institute (STI) Innsbruck, University of Innsbruck, Technikerstrasse 21a, 6020 Innsbruck, Austria
{anna.fensel,dieter.fensel}@sti2.at
http://www.sti-innsbruck.at

[2] E-Commerce Group, Institute of Software Technology and Interactive Systems, Vienna University of Technology, Favoritenstrasse 9-11, 1040 Vienna, Austria
{julia.neidhardt,pobiedina,hannes.werthner}@ec.tuwien.ac.at
http://www.ec.tuwien.ac.at

Abstract. The Web is becoming a mirror of the "real" physical world. More and more aspects of our life move to the Web, thus also transforming this world. And the diversity of ways to communicate over the Internet has enormously grown. In this context communicating the right thing at the right time in the right way to the right person has become a remarkable challenge. In this conceptual paper we propose a framework to apply semantic technologies in combination with statistical and learning methods on Web and social media data to build a decision support framework. This framework should help professionals as well as normal users to optimize the spread of their information and the potential impact of this information on the Web.

Keywords: Social Media, Semantics, Web Mining, Online Marketing.

1 Introduction

More and more aspects of our life move to the Web; the Web – as the underlying information infrastructure of today – is becoming a mirror of the "real" physical world. And it is obviously transforming this world, where it is hard to distinguish between the physical and the virtual. With developments such as the Semantic Web (moving from the syntax to the concept level) information can be automatically processed, and approaches such as Web analytics and network analysis facilitate statistical / logical inference of new knowledge, moving from data to knowledge. In this context the Web also emerged into a medium where users have become both the most important content consumers as well as the most important content producers – leading to so-called "prosumers".

This development has provided a multitude of interaction possibilities, leading to "the growth of the multichannel monster" [1]. Taking an organizational / business point of view one sees that organizations of all sizes, commercial and

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 255–266, 2012.

not-for-profit, face the challenge of communicating with their clients and partners using a multiplicity of channels, e.g., Web sites, videos, PR activities, events, email, forums, online presentations, social media, mobile application as well as structured data. The social media revolution has made this job much more challenging. The number of channels has grown and communication has changed from a mostly unilateral "push" mode to an increasingly bilateral communication, where individual stakeholders expect one-to-one communication. And the content of communication becomes more granular and increasingly dependent on the identity of the receiver and the context of the communication.

Having such a huge amount of different communication channels as well as this overabundance of information raises crucial questions: Which communication channel should I use? How should I use them? What do I communicate? How does it scale? Current tools only support the simple feedback analysis of social media data. And a straight-forward automation of the information publishing to numerous channels, as currently state of the art, does not cover very complex aspects in the decision making on where to publish the information, particularly, with respect to the optimization of brand and reputation management.

The central issue is to manage and reduce complexity in this Web x.0 world, which we intend to achieve with our framework SCAN (Social Channel Analysis and Networking). The proposed framework is currently under construction and throughout the paper we describe the generic approach and outline the challenges.

SCAN is a smart combination of semantic technologies as well as statistical and learning methods, and achieves the following: It 1) models and analyses the communication in different online channels, 2) forecasts and measures impact of the performed communication and, 3) based on the model and measurements, delivers suggestions which communication channel a user should use in order to increase the reach and spread of information. The evaluation of the framework is based upon the metrics which are the outcome of the impact analysis.

The SCAN framework will decrease the costs of driving different social media marketing campaigns. Furthermore, these activities may lead to higher conversions in online bookings and higher revenues. Since the developed system core is generic and extensible to any type of SME, our approach can be applied to different economic sectors.

The structure of the paper is as follows. In Section 2, we discuss the related work and position our approach within its landscape. In Section 3, we describe the general framework set up and its architecture. In Section 4, we present our approach to "feed" the Web, and in Section 5 our approach to "understand" the Web. Section 6 concludes the paper.

2 State of the Art

Here, we discuss the state of the art approaches from the two major activities of SCAN, namely, understanding and feeding the Web.

Modeling the Flow and Change of Information for Analysis and Forecast. This is a multidisciplinary research, therefore the literature required to follow current state of the art covers Web Science [2], evolutionary modeling, graph transformation systems and text mining.

In [3] the topological structure of the Web has been presented as a directed graph with Web pages as nodes and hypertext links between Web pages as edges between nodes. In [4] and [5] the evolution of the Web graph over time is studied, and some patterns on the emergence of new nodes and edges are presented. Though the "Web graph" model has proven to be efficient to investigate certain topological properties of the Web, e.g., in-degree and out-degree of the nodes follow the power law, it does not take into account the content placed on the Web pages. On the other hand, in [6] and [7] authors focus on the investigation of temporal dynamics of the content. However, the content under analysis is limited to the phrases extracted from quotes, and the Web structure is not taken into account.

In her survey [8], Berendt outlines the major challenges of text mining for online media. She also provides taxonomy of problems in this area. Despite the considerable successes in the areas mentioned above, the major challenge in dynamic modeling of Web documents still remains the problem how to efficiently account both for the content and structure information of Web documents over time. Currently graph transformation systems [9] are successfully applied in the areas of software engineering and model checking. We perceive stochastic graph transformation systems [10], which provide a generic framework to consider both content and structure information of the system, can be applied to the dynamic modeling of Web documents.

Content Dissemination and Impact Analysis. The field of semantics-based or enhanced Content Management Systems (CMSs) has already been quite thoroughly explored. One of the earlier approaches to ontology-based Web site management was the OntoWebber system described in [11]. OntoWebber introduced an integration layer which adapts to different data sources. This is related to our approach, but, in contrast, our approach adapts to different channels rather to different information sources. A year later, in [12], Sheth et al. introduced the SCORE system, which defines four key features: semantic organization and use of metadata, semantic normalization, semantic search, and semantic association. Although being written in the early days of the Semantic Web, the paper covers topics such as metadata extraction from unstructured text and automatic classification that may also become relevant to our approach.

The British national broadcaster BBC started to integrate semantic technologies (i.e., Linked Data) in 2009 in order to integrate various data and content sources distributed throughout the enterprise [13]. As a result of this, reported in [14], BBC's World Cup 2010 site is based on semantic repositories that enable the publishing of metadata about content rather than publishing the content itself. While the data input is fixed, different schemas for the output are defined. However, as only one channel for output is considered, the mapping is performed in a quite straightforward manner. In contrast, our system accounts for different

Table 1. Social Media Management Tools

Toolkits	Number of Supported Channels	Multi Channel Posting	Content Adaption Strategy	Feedback & Statistics	Decision Support
HootSuite http://hootsuite.com	9	Yes	Truncation of long text	Yes	No
HubSpot http://www.hubspot.com	6	Yes	No	Yes	No
Media Funnel http://sti.mediafunnel.com	6	No	Length restriction	Yes	No
Moderation Market http://moderationmarketplace.com	6	No	Length restriction	Yes	No
Ping.FM http://ping.fm	32	Yes	Truncation of long text	No	No
Seesmic https://seesmic.com	24	Yes	Common denominator	Yes	No
Sendible http://sendible.com	20	Yes	No	Yes	No
SproutSocial http://sproutsocial.com	6	Yes	Common denominator	Yes	No

information needs of various and heterogeneous channels and therefore enables the distribution of content throughout different portals.

Marketing plays an important role in the prosperity of a company, and social media is recognized as one of the efficient ways to lower costs for marketing campaigns [15]. However, social media based promotions are less controllable by managers compared to traditional marketing strategies [16]. Therefore, managers are in need of tools and methods to organize promotional activities in social media to reach the performance goals. Initiatives like HootSuite, HubSpot and others (see Table 1) provide toolkits to manage communication via Web 1.0 channels (Email, Blog) and via social Web 2.0 channels, e.g. Facebook or Twitter. Many of them provide capabilities to post in many streams via one clique, using simple mechanisms to adapt the content with regard to channel specifics. Additionally, most of the toolkits allow the user to see activity statistics and retrieve feedback. However, these toolkits neither support the user in selecting the communication channel and the target audience in it, nor provide rules to adapt the content to optimize the efficiency of the campaign.

Moreover, managers require evidence of return on investment in such activities [17]. These questions have been raised in business environment and gained much interest from the side of behavioral scientists, but they still lack thorough quantitative research. In [18] scientists argue that the design and implementation of social media campaigns is specific to national source markets. They provide guidelines for marketers from Russia and FSU countries concerning social media. However, to our best knowledge, there are no up-to-date studies of influence of social media on the Austrian market.

3 Overall Framework and Architecture

Our framework SCAN supports effective social media marketing campaigns by combining analysis of distributed information on the one side and a decision-making process on the other side. In other words, the framework helps to choose "how" (specific phrases/keywords/images/mode), "where" (specific channels) and "when" (specific time) to deliver the information – assuming that the knowledge of "what" (information on the conceptual level) to deliver and "to whom" (target category of users according to certain criteria, for example according to gender, age, location, etc.) is available. To solve this problem, we focus on the following technical subproblems:

1. Semantic modeling of the content and its integration for information on multiple channels as well as the related dissemination mechanisms;
2. Monitoring, analyzing and forecasting information flow in social media, as well as impact analysis;
3. Development of a decision making system ("how", "where" and "when" to publish the specified content to reach the target audience and to increase the effectiveness of social media marketing campaign) and a publishing toolkit.

In subproblem 1), central to the conceptualization is the abstraction of content domain from the communication channels using semantic technologies. Furthermore, interweaving, i.e., content to channel mapping, is combined with an intelligent mechanism for channel selection.

Subproblem 2) could be addressed by a model and a collection of techniques to analyze and forecast distributed information spread over the Web. That would help to estimate the impact of information spread on customer behavior and company performance on the one side, and to forecast the impact over time on the other side, providing means to define return on investment from social media spending.

The analytical part described above is integrated within an information dissemination platform (subproblem 3)). The architecture of the SCAN platform is based on the semantic data model between user and communication channels. It is depicted in Figure 1 and will be discussed in more detail in Section 4. Furthermore, a decision support framework facilitates targeted marketing based on analysis of information spread and the impact of this spread over various Web channels like Facebook, Twitter, etc.

Our approach is based on an iterative cycle, depicted in Figure 2 and to be discussed in Section 5. Austrian eTourism with its huge set of different players and Web sites [19] would represent a concrete case to evaluate our approach.

In the following two sections we discuss the issues of feeding and understanding the Web, following the same structure: "Aim and Outcome", "Methods" and "Challenges".

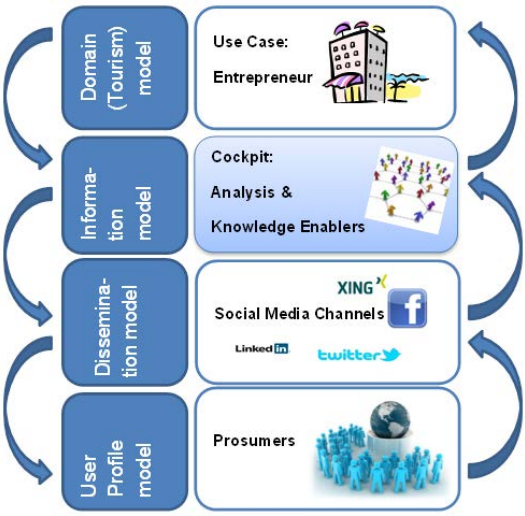


Fig. 1. SCAN Semantic Modeling Framework

4 Feeding the Web

Aim and Outcome. Employing well-developed semantic technology practices on ontology engineering and information integration, we address an under-investigated problem of dissemination and communication in the conditions of the modern multi-channeled social media.

We would use semantics to manually define the information items, the channels, the user groups, and the relationships between them. The overall SCAN semantic modeling framework is depicted in Figure 1 and is comprised of the following types of models (seen at the left hand side of the figure):

- *Domain model (Tourism)* refers to the tourism domain and can be adapted to other use cases; it defines the concepts in the domain and their relationships;
- *Information model* contains information items to be disseminated, such as news items, hotel room prices, pictures, information about services (when applicable, in terms of ontologies such as schema.org, and Dublin Core, GoodRelations) and includes mappings of information items on channels through weavers. The Information model is refined on the basis of the Web analysis, which is described in Section 5.
- *Dissemination model* describes the various channels (e.g., multiple ways to post information on Facebook) and their target groups;
- *User Profile model* contains and represents information about the user – it can be in e.g., FOAF or proprietary social network formats.

Methods. For feeding the Web we take the following steps:

- *Semantic models for content and channels:* Abstracting from the channel of communication and focusing only on the actual content is a prerequisite for scalable online communication. Here, we use semantic technologies to create models for domain content and channels. These semantic models are then used to formalize social media content in order to maximize the effect of the information spread and forecast algorithms. A semantic model is used to capture the content in domain (and not channel specific) terms allowing direct multi-channel communication for non-communication experts. These models and their interweaving with communication channels can also be reused in a vertical domain for various enterprises and other organizations.
- *Web channels identification and collection of relevant data:* For an efficient and high impact communication identifying the right audience, the right time, and especially the right channels on the Web, is crucial. In this task we will specify Web channels in social media which will be used in further analysis.
- *Content and channel interweaving:* Here, our new approach is applied, which is based on distinguishing and explicitly interweaving content and communication as a central means for achieving reusability and thereby scalability. We develop techniques for linking the Domain model with the Dissemination model through a weaver and implement them as a component of the SCAN tool. Formally, a weaver is an ordered list of tuples (see [20] for more technical details):
  1. An information item is defined as an information category that should be disseminated through various channels.
  2. An editor defines the agent that is responsible for providing the content of the information item.
  3. An editor interaction protocol is defined as the interaction protocol defining how an editor collects the content.

To allow the user to abstract from the channel level to the content level, we design the Information model and corresponding methods in our analytical part (see Section 5).

Challenges. Introducing a semantic layer on top of communication channels is required to enable common value management. However, such combination opens a broad variety of new challenges yet to be solved, in particular, as follows:

- *Modeling and interweaving feedback:* Feedback is an important part of all effective communication. Without feedback, the sender - the one who intends to convey information - has no means to validate whether or not the recipient received or understood the message. It is also often preferable to have a fully-fledged two-way conversation instead of simple one-way broadcasting. Therefore, it will be necessary to model feedback and interweave it with content items that we previously published.

- *Modeling target groups:* Companies that pursue common value management usually have a very restricted target group of people they want to address. So far in our channel model we do not distinguish between different target groups in different channels. However, different target groups reside on different communication platforms, even though there is some overlap. For example, you will find more young and hip people on Facebook, and more professional users on Xing or LinkedIn, but there are quite a few users that have a profile on both platforms. Nonetheless, they expect a different way of being engaged in different platforms.
- *Adapting content:* This problem encompasses transformations of the given information item into different formats, such as extracting images, videos or extracting and shortening Web links from piece of content. However, adapting content in a way that requires creativity and human intelligence is still a strenuous problem that reaches the borders of computability. Examples of such adaptations are shortening or translating an essay, or rewriting a text in a way that matches the target group it addresses.
- *Quality management:* An important part of targeted communication is assessing and improving the quality of conveyed content. Whereas trust, reputation and brand management are influenced by how information is perceived, quality assurance is an inbound process. The business processes for quality management and what this actually mean have yet to be defined for common value management. The bigger the campaign is, the more visible the effect of proper quality management.

5 From Data to Knowledge

Aim and Outcome. The goal of the analytical part is to develop appropriate statistical and learning techniques to analyze, forecast and measure the impact of information spread across various Web channels over time. This starts with the specification of Web channels and data to be collected from the Web following our Dissemination and User Profile models. Based on the collected data and Domain model (in our use case Tourism model), we would develop a data model as well as corresponding methods to accomplish the above stated goal.

Methods. The main criteria for the data model is to efficiently account both for structure and content present on the social media. That is why it will be represented as a directed attributed graph, and will encompass the following components: 1) content in social media; 2) users; 3) user attributes, e.g., gender, age, location, etc.; 4) relationships between users. However, the social structure differs for each social media channel, especially the type and structure of relationships between users, e.g., consider Twitter and Facebook. Thus, a separate data model is needed for each social media channel. These models are required for the analysis and prediction mechanisms, and complement the Information model, one of our semantic models.

To retrieve information from social media, we consider open source tools such as Terrier (`http://terrier.org`) and Web mining packages from Python (like BeautifulSoup). Facebook and Twitter APIs are to be used to obtain data from the relevant social networks. Google Refine (`http://code.google.com/p/google-refine`) and Talend (`http://www.talend.com/index.php`) can be applied for data cleaning and transformation. The possible tools to analyze rich semantic text are Open Calais service (`http://www.opencalais.com`), NLTK (natural language toolkit `http://www.nltk.org`), as well as GATE for sentiment analysis and opinion mining (`http://gate.ac.uk`).

Based on the social media channel data model, we would analyze the spread of information in the extracted data by combining text mining and network analysis methods. Afterwards, methods to forecast flow and change of information over the desired time in the given collection of Web channels could be developed by using Markov chains and graph transformation rules together with graph mining.

Ongoing, we specify and measure the impact of information spread on user behavior and company performance. The crucial question in this context is whether there is a correlation between 1) customer based indicators: amount of messages; mentions; comments; polarization of messages (negative or positive), etc.; and 2) company based indicators: amount of bookings; visitors to the company Web site; views of company profile, etc.

Based on the impact analysis and domain knowledge, an objective function would be constructed out of customer and company based indicators to optimize the feeding of social media channels with respect to the company goals and user interests. In the end, the companies will get support in the decision on the design of marketing campaigns. Specifically, by providing "what" (information on the conceptual level, e.g., special offer for Valentine's day) and "to whom" (the targeted audience in terms of age, gender, location, interests, etc.) to our decision support framework, the company will get advice on the aspects: "where" (set of specified social media channels),"how" (set of specified users and keywords) and "when" (specified time period for the marketing campaign) to publish.

The solution approach of SCAN is depicted in Figure 2. Our process is iterative which allows to refine the prototypes of our decision support framework and dissemination tool in each iteration. Semantic models are used throughout the whole process cycle, both contributing to each step and learning from its results.

Challenges. While analyzing collected data, we encounter a set of problems. We classify them and outline possible solutions below.

- *Missing / unavailable data:* Many social networks provide the users possibilities to restrict information visible to individuals outside of their network as well as amount of information coming from the outside. We consider opportunities to upgrade accounts on social networks to gain access to profile information of a bigger amount of users. Additionally, data mining techniques to deal with missing values are applied.
- *Low quality of data:* Data and text pre-processing techniques (e.g., stop words elimination, stemming) are applied to ensure coherent data.

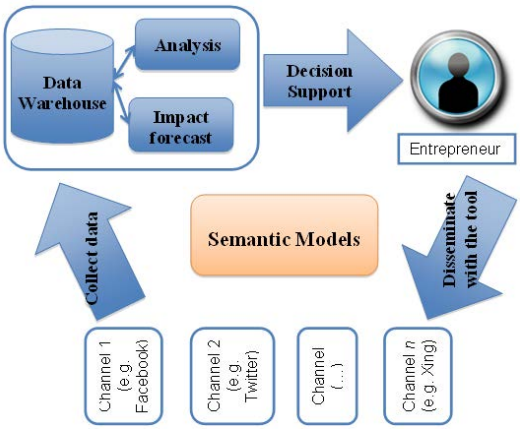


Fig. 2. SCAN Process Cycle

- *Inappropriate choice of time periods:* To evaluate a marketing campaign it is essential to select a proper time period to monitor user behavior and company performance before and after the campaign since campaigns may have either immediate or remote effect. As an option to overcome this challenge, experiments with different time periods are conducted.
- *Scalability in forecast calculation:* Since the number of social media data that is available is rather high and decision support that is provided with a very high latency is useless, the amount of time needed to calculate the predictions needs to be within a feasible time frame.

6 Conclusions

We propose a conceptual framework, SCAN, that combines semantic as well as statistical and learning techniques for efficient and high impact communication, targeting the right audience at the right time through the right channels on the Web. Central to SCAN is the process of interweaving content and communication channels. That would enable channel selection based on the semantics of the information item to be distributed. Moreover, it would provide means for content transformation with regard to the specifics of the communication channels for the distribution of this information item. Thus, non-communication experts would not need to worry about the technical aspects of how to interact with various channels; instead, they would just focus on the content being communicated. Furthermore, by combining methods from Web Science, network analysis and text mining, SCAN provides intelligent analysis of the communication in different social media channels.

The methods and techniques developed in SCAN would be applied to distributed data coming from Austrian eTourism. The two major challenges to exploit the benefits of social media are overabundance of information, out of which

75% is estimated to be redundant, and constant growth of structural complexity. Particularly, Austrian tourism operators are now left with the problem of how to use social media to its full extent. This is exactly the problem which our framework helps them to solve. The effectiveness of SCAN and its social media campaigns would be measured and evaluated through its impact on user behavior, for example, number of comments, and company performance, for example, number of room bookings via the website of a hotel. Such evaluation will need a mixture of tools and approaches from on/off-line interviews with users up to economic analysis of company performance.

References

1. Mulpuru, S., Harteveldt, H.H., Roberge, D.: Five retail ecommerce trends to watch in 2011. Forrester Research Report (2011)
2. Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T., Weitzner, D.: Web science: an interdisciplinary approach to understanding the web. Commun. ACM 51(7), 60–69 (2008)
3. Kleinberg, J.M., Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A.S.: The Web as a Graph: Measurements, Models, and Methods. In: Asano, T., Imai, H., Lee, D.T., Nakano, S.-i., Tokuyama, T. (eds.) COCOON 1999. LNCS, vol. 1627, pp. 1–17. Springer, Heidelberg (1999)
4. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graph evolution: Densification and shrinking diameters. ACM TKDD 1(1) (2007)
5. Bringmann, B., Berlingerio, M., Bonchi, F., Gionis, A.: Learning and predicting the evolution of social networks. IEEE Intelligent Systems 25, 26–35 (2010)
6. Leskovec, J., Backstrom, L., Kleinberg, J.: Meme-tracking and the dynamics of the news cycle. In: Proc. KDD 2009, pp. 497–506 (2009)
7. Yang, J., Leskovec, J.: Patterns of temporal variation in online media. In: Proc. WSDM 2011, pp. 177–186 (2011)
8. Berendt, B.: Text mining for news and blogs analysis. In: Encyclopedia of Machine Learning, pp. 968–972. Springer (2010)
9. Ehrig, H., Ehrig, K., Prange, U., Taentzer, G.: Fundamental theory for typed attributed graphs and graph transformation based on adhesive HLR categories. Fundam. Inf. 74, 31–61 (2006)
10. Heckel, R., Lajios, G., Menge, S.: Stochastic graph transformation systems. Fundam. Inf. 74, 63–84 (2006)
11. Jin, Y., Decker, S., Wiederhold, G.: Ontowebber: Model-driven ontology-based web site management. In: Semantic Web Working Symposium, SWWS (2001)
12. Sheth, A., Bertram, C., Avant, D., Hammond, B., Kochut, K., Warke, Y.: Managing semantic content for the web. IEEE Internet Computing 6(4), 80–87 (2002)
13. Kobilarov, G., Scott, T., Raimond, Y., Oliver, S., Sizemore, C., Smethurst, M., Bizer, C., Lee, R.: Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections. 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. 723–737. Springer, Heidelberg (2009)
14. Bishop, B., Kiryakov, A., Ognyanoff, D., Peikov, I., Tashev, Z., Velkov, R.: Owlim: A family of scalable semantic repositories. Education 2(1), 33–42 (2011)

15. Kirtis, A.K., Karahan, F.: To be or not to be in social media arena as the most cost-efficient marketing strategy after the global recession. Procedia - Social and Behavioral Sciences 24, 260–268 (2011); The Proceedings of 7th International Strategic Management Conference
16. Mangold, W.G., Faulds, D.J.: Social media: The new hybrid element of the promotion mix. Business Horizons 52(4), 357–365 (2009)
17. Weinberg, B.D., Pehlivan, E.: Social spending: Managing the social media mix. Business Horizons 54(3), 275–282 (2011)
18. Fotis, J., Buhalis, D., Rossides, N.: Social media impact on holiday travel planning: The case of the russian and the fsu markets. International Journal of Online Marketing (IJOM) 1, 1–19 (2011)
19. Piazzi, R., Baggio, R., Neidhardt, J., Werthner, H.: Destinations and the web: a network analysis view. Information Technology & Tourism 13(3) (2011)
20. Bauereiss, T., Leiter, B., Fensel, D.: Effective and efficient online communication. Technical report, STI Innsbruck, Innsbruck (2011)

Recommendation of Mobile Services Employing Semantics and Community Generated Data

Alex Oberhauser[1], Corneliu-Valentin Stanciu[1], and Anna Fensel[1,2]

[1] Semantic Technology Institute (STI) Innsbruck
University of Innsbruck, Technikerstrasse 21[a]
6020 Innsbruck - Austria
firstname.lastname@sti2.at
[2] Telecommunications Research Center Vienna (FTW)
Donau-City-Strasse 1 / 3. Stock
1220 Vienna - Austria
fensel@ftw.at

Abstract. The number of online services is growing dramatically. Nowadays they can be semantic or Web 2.0 based, for fixed or mobile device consumption, end-user or provider created, oriented on specific user groups, social networks, etc. Therefore, selection and recommendation of services for the end users on the basis of the service and user data becomes a challenge, and conventional keyword-based information retrieval are no longer sufficient. Here we present an approach for effective selection and recommendation of heterogeneous online services, combining natural language based information retrieval techniques and analysis of semantic annotation, community-generated Web 2.0 type content and location awareness data.

Keywords: Service Recommendation, Semantics, Web 2.0, Context Awareness, Synonyms Identification, Online Communities, Mobile Platform.

1 Introduction

This work introduces an algorithm and an implementation of a recommendation system that uses context information from a searching agent, as well as the tags and ratings generated by the community. The main purpose of the algorithm is to provide an approach that could be used in a rapidly changing environment such as with the mobile devices usage, for generating service recommendations based on consumer opinion as well as similarity between the user who uses the recommender and the consumer who has expressed his/her opinion. For more purposeful results the service query is generated based on context-aware synonyms. The constraints have led to a component based architecture where the logical steps ((1) query creation, (2) query based search and (3) service recommendation) are split into two different, independent components - Query Generation and Service Recommendation. This allows extending the algorithm

W. Abramowicz, J. Domingue, K. Węcel (Eds.): BIS 2012 Workshops, LNBIP 127, pp. 267–278, 2012.

with additional context information from different sources. Throughout the work, the term *service* is defined as a resource that has related ranking information and tags. Further, we assume that practically the services are annotated by users. This definition makes the algorithm applicable for a wide range of service types, such as places, events and products to name a few.

This paper is structured as follows. Section 2 provides the problem statement, motivation and background. In Section 3 the related work is explained. Section 4 presents the approach for service search relying on community generated information. The implementation and example are outlined in Section 5. Finally, Section 6 concludes the paper and offers an outlook for future work.

2 Problem Statement and Motivation

The open-source and developer-supported platforms, for example the Web APIs, have encouraged the exponential growth of mobile applications and services. Because more and more people are joining online communities by creating accounts or profiles, social networks have also turned their attention to mobile platforms.

It is nearly impossible for the end user to choose well suited services on the go without the help of recommendation systems that pre-filter the mass of services and show only a short list of recommendations. The here presented algorithm is also applicable in traditional scenarios, but the mobile use-case simplifies the gathering of context information, such as location or appointments. Such recommender system are also important for e-Tourism stakeholders. It helps to improve the visibility and reach the right target group. This results in more positive reviews, through satisfied customers. One side effect of this growth is that the context information that is produced by users is also growing. This facilitates the development of algorithms that take into account user preferences and consumer opinions in order to recommend services.

One representative example of this problem is searching with the help of abbreviations. If a user searches for *AI*, simple search engines try to match the search term with terms in the underlying data storage. That behaviour could lead to results that include, for example, the word ***Thai*** that has nothing to do with our starting search term. Also, if we assume that the underlying search engine expands the abbreviation, the result is ambiguous (e.g. *AI* expands to *Amnesty International* or *Artificial Intelligence*). This approach is able to select one semantic meaning (with the help of the context, e.g. by user interests) and filter out the rest.

The presented algorithm gives a possible solution of how to overcome this weakness with the help of a context-aware search query in combination with a recommendation system.

The switch of user behaviour from non-mobile web applications to web applications that take into account context information on-the-go, makes it important to include this information into recommendation systems. Another motivation for the development of the algorithm was the public availability of personal information on the web (e.g. social networks, personal websites). With the help of

such user profiles, the search for relevant services could be improved further. The aforementioned algorithm is able to return such services that could use software on mobile devices to be able to react faster to real world changes and improve the user's experience.

3 Related Work

The research work in service recommendation systems field is very wide, but only a few are focused on web applications that take into account context information of searching agents, which are mainly available on mobile devices (see [5], [15]). Another work in the area, *Context-Aware Query Refinement for Mobile Web Search* [8], makes also heavy use of context information, but does not take into account the static context information related to the searching agent, for example the interests.

A close related algorithm is presented in [1] which also use consumer opinions, but only for recommendation of products. The main difference is the flexibility of our approach to be applied in different branches and areas. One drawback of the approach from [1] is the need for an ontology for every service concerning the capability of the recommendation system to compute the ranking.

Another approach is to take into account low level information received from the device used by the searching agent [12]. Such information are for example GPS coordinates, screen resolution, network capabilities (e.g. network speed).

The recommendation algorithm that sovles the one million dollar question advertised by Netflix [4] is suitable for data with a large training set, but not for a fast changing environment. The underlying data set, the movie library, will be extended over the time, but not changed completely.

Another algorithm that works on the base of graphs is described in *Studying Recommendation Algorithms by Graph Analysis* [11]. This approach is demonstrated on the base of movies.

In difference to the mentioned works, our approach makes heavy use of context information that could be extracted from different sources related to the searching agents, e.g. social networks and of user-generated reviews of services, but also frequent changing information such as location or temporary interests.

The presented algorithm leverage also from the knowledge that was made during research in the area of Semantic Web [3]. To mention is the abstraction of context, synonym and review information with the help of well known and used ontologies, such as FOAF, SIOC, SKOT.

4 Approach

The presented approach consists of different steps. The first step is to generate a query to find all relevant services. In this step we take into account context information of the searching agent to improve the relevance of the search result and for a first pre-filtering to minimize the output of unsuitable services. As you can see in Fig. 1 both parts of the algorithm use the context information, such

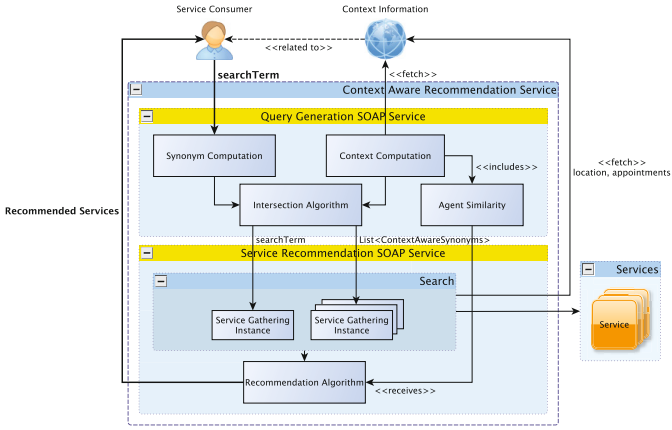


Fig. 1. Architecture View for the Algorithm

as the location, that are related to the searching agent (*Service Consumer*). To not break the recommendation algorithm, the query generation always returns at least the search term given by the searching agent.

Subsection 4.1 describes the algorithm used for the query generation (*Query Generation SOAP Service*) in more detail. After the execution of the query, the output is forwarded to the recommendation algorithm (*Service Recommendation SOAP Service*) that is described in Subsection 4.2.

4.1 Query Extension Based on Context Aware Synonyms

This component addresses the problem of context data classification and the computation of synonyms. After the two information clouds - synonyms and context - are computed, the method is to intersect the two sets. The resulting set then includes context-aware synonyms and is used as input for the recommendation algorithm.

As one can see in Fig. 2 the computation begins with the interaction with another agent (e.g. a human being or a community). There are two types of interaction with the entity. First there is a passive or indirect way of communication. In this phase, the context will be generated from different sources that are related to the agent or the community[1].

The next phase is the active part where a search term is given. This string is used for the input to query different dictionaries that return the synonyms. A dictionary that perfectly suits the computation of context-aware synonyms is WordNet from the Princeton University. The outputs are grouped into blocks, called "synsets". Each synonym in a synset has semantically the same meaning

[1] Through the abstraction of the context information, the gathering process is not bound to specific sources.

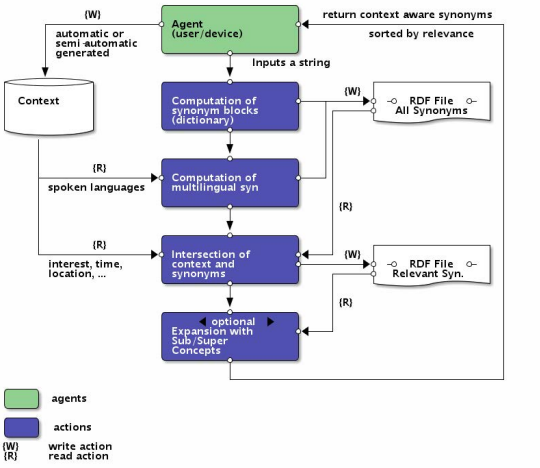


Fig. 2. Context-Aware Synonyms Generation for Search Query

and the block has a natural language description of the meaning with example sentences. To simplify the computation in later steps and to make it possible to extend the approach with other dictionaries, the data is described with help of RDFS[2] ontologies. Additional to the simplification of the computation, it also hides the underlying implementation of the data retrieval part with the advantage of being able to expand the software in a later step with additional dictionaries, without changing the algorithm that is responsible for the generation of context-aware synonyms. After the general computation of synonyms, it is possible to compute the translation of this word for each synonym in the newly gained set. For this purpose the context of the agent is used in order to get the spoken languages. The expanded sets are written to the same RDF file as the synonyms that we have saved in the previous step.

The second major information cloud, alongside the synonyms, is the context. It is useful to split the context into two major categories. One is the Static Context and the other is the Dynamic Context. The Static Context contains information that describes the personal context of the agent and changes rarely, e.g. the interests, in difference to the Dynamic Context, that describes the environmental context and changes more often, e.g. the current location or appointments. The separation is made because the computation on the two sets differs, and additionally the Static Context should not be touched if something in the Dynamic Context changes. Additional to that the Dynamic Context has a higher priority in respective to the Static Context. Other than that the two contexts are used the same way throughout the algorithm.

Along with the capability to add different sources for the context gathering, the algorithm also provides the capability to add context information from

[2] Resource Description Framework Schema [14].

different agents. For the scope of the similarity algorithm we define the bidirectional function *simContext* as a relation between two contexts which outputs a relative value between 0 and 1 that indicates the similarity in percentage. More formally the following Formula 1 describes the *simContext* relation. The function takes as input the two contexts of the two agents and returns the similarity value.

$$simContext : (C_1 \times C_2) \rightarrow R \in [0, 1] \tag{1}$$

It makes sense to use only similar contexts where the percentage is higher than a threshold value S. If the check **simContext** $<$ **S** evaluates to true, then the relation of the foreign context is not used for the algorithm. Otherwise we add the context of the foreign agent to the context of the searching agent with an annotation of the similarity value for each added context tag. It makes sense to choose the threshold value S higher than 0.75 or 75 % and fine tune this value during practical use to receive an optimal result.

To compute the similarity of two contexts we use the Formula 2 and 3. Formula 2 computes the similarity of agent1 to agent2 and Formula 3 makes average of the two computed similarity values. This step is needed, because the presented Formula is bidirectional. The index *a1 / a2* stands for *agent1 / agent2*.

$$sim(context_{a1}, context_{a2}) = \frac{\sum^{\#(matching)} absoluteFrequency}{\sum^{\#(total)} absoluteFrequency} \tag{2}$$

$$simContext = \frac{sim(context_{a1}, context_{a2}) + sim(context_{a2}, context_{a1})}{2} \tag{3}$$

The matching tags in the context of one agent (in respective to the second agent) are divided by the sum of the number of all tags in the two context clouds (Formula 2). The same computation is made for the second agent. The average of both values is the similarity of the two context clouds (Formula 3).

The similarity algorithm does not work directly on the input sources, but on the abstracted context tag cloud.

The standalone implementation uses FOAF (see [3]) files that describe an agent. This description consists of interests of the user (represented by foaf:interests) and of additional information stored in external RDF files, for example links to SIOC files that describes publications (linked with the help of foaf:publications).

A formal description how the intersection algorithm works is given below.

$S_{input} = \cup_{i \in I} S_i$	... Synonyms for the input term *input*. S_i is one synonym block (synset), with synonyms with the same semantic.
$C_{agent} = C_{DYN} \cup C_{STAT}$	... The context information of the *agent* that triggers the search. C_{DYN} for dynamic and C_{STAT} for static context.
R	... The context-aware synonym result set.

The two sets S_{input} and C_{agent} are gathered from different sources. On the base of this two sets the result set R is computed. The following formula shows how the result is defined.

$$\exists i(\exists t(t \in S_i \wedge t \in C_{agent}) \implies \forall r(r \in S_i \wedge r \in R))$$

The formula above claims that a term exists, that is part of one (or maybe more) synonym block(s) and the context set. That implies that all terms in the found synonym block(s) are also part of the result set. If this statement holds, we have found context-aware synonyms.

4.2 Recommendation Based on Community Tags

In order to generate proper service recommendation based on consumer's opinions, it must first assess these opinions, which are primarily written in free-form text. This algorithm is closely related to the approach presented in [1] where the reviews are used to recommend services. The main difference between these two approaches were described in the Section 3. Another possible issue could be the complexity of using parsers for natural language and text-mining techniques to extract information from comments and transform them into recommendations.

The solution is offered by tags, also known as keywords, which are successfully used by a myriad of search engines and applications for various scopes. Therefore, also using tags to recommend services could be the next improvement of recommendation systems.

Ranking. The advantage of using the *opinion quality (OQ)* and *service quality (SQ)* from [1] is also introduced in our approach, with the difference that *OQ* is computed based on the tags attached to reviews and the *SQ* by using the service rating to evaluate the quality value of the service. The *overall service quality (OSQ)*, as also in [1], is a global valuation of the service, which is calculated by the SQ value of each individual review.

Computing the Opinion Quality. In general, people have different opinions, tastes, lifestyles, etc. Therefore it is difficult to calculate the opinion quality for two different categories of services, like recommending devices and places, because for devices, a taxonomy could exist classifying the reviewers skills and expertise, while for places something similar may be applicable. The *opinion quality* is defined as the sum of the relative frequency of tag t_i for each variable i representing the tags provided in review r divided by the numbers of tags appearing in r. The relative frequency is the absolute frequency of a tag t divided by the total number of tags arising in all reviews. Each review r will have a OQ value based on the occurring keywords.

$$OQ_r = \frac{\sum_{i}^{\#t} t_{i_{RelFreq}}}{\#t} \tag{4}$$

Computing the Service Quality. In this step the service will be ranked according to the consumer opinions computed above. For weighing these opinions and to quantify the user valuation of the given tags, the service rating will be used. Most of the platforms offer stars, where consumers can choose between one (very bad) and five (very good) when commenting on a service. To these, ratings are assigned numbers from -1 corresponding to one star, -0.5 for two stars, 0.1 for three stars, 0.5 for four stars and 1 for five stars. This is another improvement regarding [1], because there is not only good and bad, but also something inbetween. The SQ is computed for each review r by multiplying the corresponding value of the rating of r with OQ of r.

$$SQ_r = x * OQ_r,\ x \in \{-1,\ -0.5,\ 0.1,\ 0.5,\ 1\} \tag{5}$$

Computing the Overall Service Quality. The OSQ is defined as the global valuation of the service based on all reviews addressed to this service. This OSQ is computed by the average of SQ value multiplied by the $Scalingfactor$. This $Scalingfactor$ can be any value that adjust the user evaluation.

$$OSQ_s = \frac{\sum(Scalingfactor * SQ_r)}{NumberOfOpinions} \tag{6}$$

In our case the $Scalingfactor$ will be the value computed with the context similarity algorithm from Formula 2 and 3. For this purpose, the similarity will be calculated between the user who does the search and the consumer who publishes the review. The $Scalingfactor$ will then be initialised with a value between 0 and 1, according to the description of Formula 2 and 3. This value is also the score of the service, which will be significant for ordering the results.

Further Improvement. For the services that provide a location, it is helpful to pre-filter them in order to exclude the unreachable ones. Therefore, a radius of reachable location-aware services [10] will be computed based on the time until the next appointment, the GPS coordinates and the transportation way of the user. This may exclude the irrelevant services and the recommendation will then occur only in the services in the radius.

5 Implementation and Example

The implementation consists of two major parts that are implemented as web services with publicly exposed interfaces. The first part (1), as described in Subsection 4.1 and called *Query Generation SOAP Service* in Fig. 1, computes the context-aware synonyms on the base of FOAF files, that describes the context of the searching user. FOAF was chosen to have a higher level representation of context and friend relation information related to one person. The easy transformation from social network information is shown on the example of the Facebook App FOAFGenerator[3]. Additional to the web service architecture, the Query

[3] FOAFGenerator `http://fbtofoaf.networld.to`

Generation Component consists of two independent parts that are responsible for the synonyms, respectively for the context computation. This implementation decision coincides with the flexibility of the algorithm to handle different underlying data sources and the reusability of parts throughout the workflow (see *Agent Similarity* in Fig. 1).

The second part (2) as described in Subsection 4.2 and called *Service Recommendation SOAP Service* in Fig. 1, ranks the services on the basis of consumer opinions, service rating and the similarity between users and consumers. Qype was used as a platform for implementing the service recommendation, platform which also supports the social community. The algorithm could be changed easily to fit other platforms that provide tags and ratings.

To optimize the recommendation process, the implementation only takes places that are near the searching agent. A fictive search for restaurants demonstrates the algorithm as also the additional component that was implemented for visualisation purposes.

Rating			Review 1 4	Review 2 4	Review 3 5
Tag name	**Freq.**	**Relative Freq.**			
London	3	0.3	✓	✓	✓
restaurant	2	0.2	✓	-	✓
center	1	0.1	-	✓	-
panorama	1	0.1	✓	-	-
delicious	1	0.1	-	✓	-
friendly	1	0.1	-	-	✓
expensive	1	0.1	-	✓	-
Similarity of agents			0.3	0.5	0.7

The OQ is calculated using Formula 4 and the information from table above.

$$OQ_1 = \frac{0.3+0.2+0.1}{3} = 0.2$$
$$OQ_2 = \frac{0.3+0.1+0.1+0.1}{4} = 0.15$$
$$OQ_3 = \frac{0.3+0.2+0.1}{3} = 0.2$$

Then the SQ is computed based on the OQ result values and the corresponding rating. Therefore, if a review has a rating of five, the OQ value will be multiplied by 1 and for a rating of four by 0.5.

$$SQ_1 = 0.5 * 0.2 = 0.1$$
$$SQ_2 = 0.5 * 0.15 = 0.075$$
$$SQ_3 = 1 * 0.2 = 0.2$$

At the end the OSQ is computed as the average of all SQ values multiplied by the $Scalingfactor$. For this example we gave some fictive numbers of the similarity between the user and the consumers. During the execution of the complete implementation, this value is received from the *Query Generation* component that is able to compute the user similarity.

$$OSQ = \frac{(0.3*0.1)+(0.5*0.075)+(0.7*0.2)}{3} = 0.116$$

6 Contribution

In contrast to previous work that is optimized for search improvements on the web (see [5], [15]) the context consists of context information related to the searching agent and not of previous search queries. This approach has the advantage that it perfectly suites mobile environments where the query changes with location related information and is independent in relation to the previous search queries. Another difference in the query creation is the use of synonyms as extension, whereas the algorithm in [7] uses only context information. The ongoing work [2] plans to use the same technology (e.g. RDFS/OWL ontologies) and concepts in the same scenarios. The base of the scenarios is that an agent searches on-the-go for resources that are relevant for its context. A difference is that our work presents an algorithm for the computation of context-aware synonyms. This part is mentioned shortly in [2]. For the recommendation of services we also take into account user-generated quality measurements, such as rating values. This additional layer improves the quality of the recommended systems and allows an improvement over time. Another improvement is the use of the friend relation to find suggestions from users that are similar to the searching agent. This part sets the agent in a social context and helps to choose the right community recommendations for the searching agent.

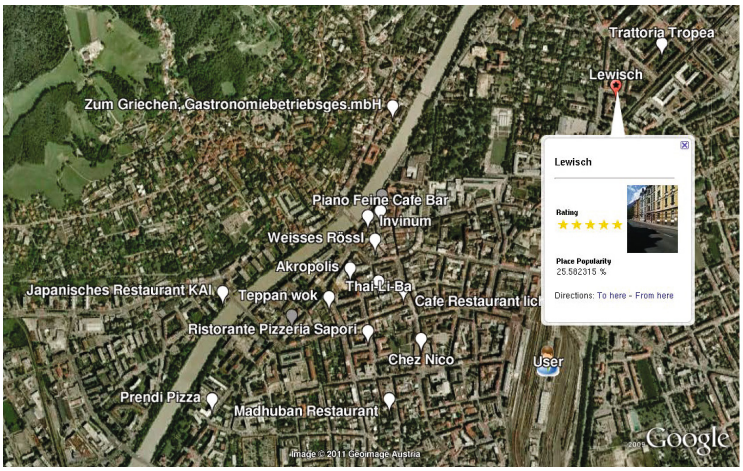


Fig. 3. Results after Searching for *restaurants* in *Innsbruck*

For evaluation purposes we have also developed a **visualisation component** which exports the results entries into KML files[4] and show them on Google Earth[5]. Figure 3 is a screenshot of the visualisation component, showing the

[4] KML stands for Keyhole Markup Language and is used for location storage in XML.
[5] Google Earth: `http://www.google.com/earth/index.html`

results after applying the service recommandation engine on the input received from the user extended with the list of context-aware synonyms from the query generation.

We have integrated and evaluated the service search techniques in a novel design and platform for user-generated mobile microservices from m:Ciudad project [9]. The Web 2.0 content search techniques extends the model for the definition of a microservice, as well as an advanced implementation for supporting search and authoring of such services [6]. This sets a step towards the vision of user-generated mobile microservices and their deployment in an open service platform. m:Ciudad platform is, to our knowledge, the first designed and implemented infrastructure enabling end-user mobile service creation, with its approach described in [13].

7 Summary and Conclusions

We have presented an algorithm providing a very flexible, component based heterogeneous service recommendation approach able to react within rapidly changing mobile environments. In contrast to previously developed recommendation systems, this approach uses a wider definition of context information for users that are searching for resources. The algorithm is spcifically designed to recommend all types of services and takes into account the current context information (e.g. location, appointments) and the context information that are related to the searching agent (e.g. interests, working topics). This approach makes the algorithm suitable for mobile devices, where this information is available and could be used locally. With the consistent use of user-generated information and context information related to the searching agent and its social connections to other agents, the algorithm guarantees a highly context related result. Further practical evaluation can be performed deploying the system in the real life and asking the users to find the services (e.g. restaurants) with and without the techniques, and provide feedback on which outcomes and methods they prefer and why.

References

1. Aciar, S., Zhang, D., Simoff, S., Debenham, J.: Informed Recommender: Basing Recommendations on Consumer Product Reviews. In: IEEE Intelligent Systems, pp. 39–47. IEEE Computer Society, Washington, DC (2007)
2. Arabshian, K.: A framework for personalized context-aware search of ontology-based tagged data. In: Proceedings of the 2010 IEEE International Conference on Services Computing, SCC 2010, pp. 649–650. IEEE Computer Society, Washington, DC (2010), http://dx.doi.org/10.1109/SCC.2010.73
3. Auer, S., Bizer, C., Müller, C., Zhdanova, A.: The social semantic web. Lecture Notes in Informatics (LNI), vol. 113. Bonner Köllen Verlag (September 2007)
4. Bell, R.M., Koren, Y., Volinsky, C.: The bellkor 2008 solution to the netflix prize. Seminars in Laparoscopic Surgery 9(4), 197–197 (2008), http://www2.research.att.com/~volinsky/netflix/

5. Cao, H., Jiang, D., Pei, J., Chen, E., Li, H.: Towards context-aware search by learning a very large variable length hidden markov model from search logs. In: Proceedings of the 18th International Conference on World Wide Web, WWW 2009, pp. 191–200. ACM, New York (2009), http://doi.acm.org/10.1145/1526709.1526736
6. Danado, J., Davies, M., Ricca, P., Fensel, A.: An Authoring Tool for User Generated Mobile Services. In: Berre, A.J., Gómez-Pérez, A., Tutschku, K., Fensel, D. (eds.) FIS 2010. LNCS, vol. 6369, pp. 118–127. Springer, Heidelberg (2010)
7. Häkkilä, J., Korpipää, P., Ronkainen, S., Tuomela, U.: Interaction and End-User Programming with a Context-Aware Mobile Application. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 927–937. Springer, Heidelberg (2005)
8. Hattori, S., Tezuka, T., Tanaka, K.: Context-aware query refinement for mobile web search. In: IEEE/IPSJ International Symposium on Applications and the Internet Workshops, p. 15 (2007)
9. (Information, I., Technlogies), C.: m:Ciudad FP7. http://www.mciudad-fp7.org/ (Online; accessed April 28, 2011)
10. Kaasinen, E.: User needs for location-aware mobile services. Personal Ubiquitous Comput. 7, 70–79 (2003), http://dx.doi.org/10.1007/s00779-002-0214-7
11. Mirza, B.J., Keller, B.J., Ramakrishnan, N.: Studying recommendation algorithms by graph analysis. Journal of Intelligent Information Systems 20(2), 131–160 (2003), http://www.springerlink.com/index/v74q03776485vml4.pdf
12. Polyviou, S., Evripidou, P., Samaras, G.: Contextaware queries using query by browsing and chiromancer. In: Second International Conference on Pervasive Computing (2004)
13. Shin, Y., Yu, C., Chung, S., Kim, S.: End-user driven service creation for converged service of telecom and internet. In: Fourth Advanced International Conference on Telecommunications, AICT 2008, pp. 71–76 (June 2008)
14. W3C: RDF Schema, http://www.w3.org/TR/rdf-schema/ (Online; accessed August 04, 2010)
15. Xiang, B., Jiang, D., Pei, J., Sun, X., Chen, E., Li, H.: Context-aware ranking in web search. In: Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2010, pp. 451–458. ACM, New York (2010), http://doi.acm.org/10.1145/1835449.1835525

Author Index