LNBIP 160

Witold Abramowicz (Ed.)

Business Information Systems Workshops

BIS 2013 International Workshops
Poznań, Poland, June 2013
Revised Papers

BUSINESS INFORMATION SYSTEMS 2013

Springer

Lecture Notes in Business Information Processing 160

Series Editors

Witold Abramowicz (Ed.)

Business Information Systems Workshops

BIS 2013 International Workshops
Poznań, Poland, June 19-20, 2013
Revised Papers

Springer

Volume Editor

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

ISSN 1865-1348 e-ISSN 1865-1356
ISBN 978-3-642-41686-6 e-ISBN 978-3-642-41687-3
DOI 10.1007/978-3-642-41687-3
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013951363

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

There are many topics in business information systems that deserve attention. BIS Workshops give researchers the possibility to share preliminary ideas and first experimental results and to discuss research hypotheses. Past experience has shown that workshops are also a perfect instrument to create a community around very specific research topics. BIS Workshops are an ideal place to confront authors' research ideas with a well-focused audience. Finally, the discussions held during presentations strengthen the paper and prepare it for the publication.

Four workshops took place during the 16th edition of the BIS conference. They covered such topics as applications and economics of knowledge-based technologies (AKTB), business and IT alignment (BITA), formal semantics for future enterprise (FSFE), and legal informatics and legal information technology (LIT). Additionally, as a response to a need from young researchers for exchanging their PhD research results with more experienced colleagues, the first edition of the PhD Symposium was organized.

The volume opens with two papers prepared by BIS keynote speakers. The 25 articles contained in this volume are an extended version of papers accepted for BIS workshops as well as for the PhD Symposium. There were 47 submissions for all mentioned events. Based on reviews, the respective workshop chairs accepted 25 in total, yielding an acceptance rate of 53%.

We would like to express our thanks to everyone who made the BIS 2013 workshops successful. First of all, our 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.

June 2013 Witold Abramowicz

Conference Organization

BIS 2013 was organized by Poznań University of Economics, Department of Information Systems.

Conference General Chair

Witold Abramowicz	Poznań Univeristy of Economics, Poland

BIS Organizing Committee

Elżbieta Bukowska (Chair)	Poznań University of Economics, Poland
Barbara Gołębiewska	Poznań University of Economics, Poland
Włodzimierz Lewoniewski	Poznań University of Economics, Poland
Szymon Łazaruk	Poznań University of Economics, Poland
Bartosz Perkowski	Poznań University of Economics, Poland
Wioletta Sokołowska	Poznań University of Economics, Poland
Milena Stróżyna	Poznań University of Economics, Poland
Krzysztof Węcel	Poznań University of Economics, Poland

AKTB 2013 Workshop Chairs' Message

The 5th Workshop on Applications of Knowledge-Based Technologies in Business (AKTB 2013) was organized in conjunction with 16th International Conference on Business Information Systems (BIS 2013) held in Poznań, Poland. It continued the successful series of AKTB workshops in Poznań (2011, 2009), Berlin (2010), and Vilnius (2012).

The workshop aims to engage researchers and practitioners, specialists, and market analysts in sharing their research experiences and domain knowledge in the application of contemporary computational intelligence methods for modeling and implementation of business information systems.

The AKTB 2013 kept the tradition of presenting innovative and robust computational solutions validated by experimental research and based on the in-depth domain knowledge of various business areas.

The main attention of the workshop was focused on the most urgent scientific triplet: business, computational intelligence, and big data. The approved papers discuss advanced services for information systems users, and propose innovative solutions for modeling systems and processes. The conference workshop themes are related to methods of solving complex business tasks, analytical insights, and application of intelligent and knowledge-based technologies for high-quality experimental research in various fields of business problems.

The main topics of the workshop concentrate 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 big 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, company rating systems.

Ten articles were submitted to the AKTB2013 workshop. Each paper was evaluated by at least three independent reviewers of the Program Committee. The four highest ranked articles were accepted for presentation during the conference and the second stage of reviewing before including them in the proceedings. One paper was selected as invited paper.

Each reviewer evaluated the quality of the article by taking into account the criteria of relevance of the article to the workshop topics, adequacy of the article title and the content, its originality and novelty, coherence of methodological background, substantiation and validity of the conclusions, and quality of presentation of the paper.

The Program Committee comprised 23 outstanding researchers who represented prestigious scientific institutions from 13 countries.

We appreciate the expertise and quality of work of the Program Committee members, whose reviews provided a deep analysis of the submitted research works and highlighted valuable insights for the authors. The high standards followed by the reviewers enabled us to ensure the high quality of the workshop event, excellent presentations, intensive scientific discussions, and added value to the workshop proceedings. We would like to express our gratitude for the success of AKTB 2013 to all authors of submitted papers, members of the Program Committee, Vilnius University, Masaryk University, and Department of Information Systems of the Poznań University of Economics and to acknowledge the outstanding efforts of Organizing Committee of the 16th International Conference BIS 2013.

June 2013

Virgilijus Sakalauskas
Dalia Kriksciuniene
Tomas Pitner

Organization

Chairs

Dalia Kriksciuniene	Vilnius University, Lithuania
Tomas Pitner	Masaryk University, Czech Republic
Virgilijus Sakalauskas	Vilnius University, Lithuania

Program Committee

Rajendra Akerkar	Western Norway Research Institute, Norway
Lia Bassa	Foundation for Information Society, Hungary
Tânia Bueno	Instituto de Governo Eletrônico, Brazil
Saulius Gudas	Vilnius University, Lithuania
Mario Hernandez	University of Las Palmas de Gran Canaria, Spain
Daning Hu	University of Zurich, Switzerland
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
Laima Papreckiene	Kaunas Technological University, Lithuania
Razvan Petrusel	Babes-Bolyai University, Romania
Tomas Pitner	Masaryk University, Czech Republic
Vytautas Rudzionis	Vilnius University, Lithuania
Virgilijus Sakalauskas	Vilnius University, Lithuania
Eric Schoop	TU Dresden, Germany
Mikhail Simonov	ISMB, Italy
Darijus Strasunskas	NTNU, Norway
Rimvydas Simutis	Kaunas University of Technology, Lithuania
Margaret Tan	Nanyang Technological University, Singapore
Sebastián Ventura	Cordoba University, Spain
Danuta Zakrzewska	Technical University, Lodz, Poland

BITA 2013 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 an obvious demand today 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 modelling 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 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 the 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 process, and (3) formal communications.

BITA 2013 was the 4th workshop on this subject following an event in 2012, which was located at the 16th International Conference on Business Information Systems in Poznań. The workshop aimed to bring together people who have 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. A specific focus was on practices of business and IT alignment, i.e., on case study and experiences papers.

The workshop received 11 submissions, of which six were selected by the Program Committee.

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

June 2013

Ulf Seigerroth
Kurt Sandkuhl

Organization

Chairs

Ulf Seigerroth	Jönköping University, Sweden
Kurt Sandkuhl	University of Rostock, Germany

Program Committee

Ulf Seigerroth	Jönköping University, Sweden
Kurt Sandkuhl	University of Rostock, Germany
Ulrike Borchardt	University of Rostock, Germany
Rimantas Butleris	Kaunas University of Technology, Lithuania
Stefan Cronholm	Linköping University, Sweden
Steven De Haes	University of Antwerp, Belgium
Khaled Gaaloul	CRP Henri Tudor, Luxembourg
Anders Hjalmarsson	Viktoria Institute, Sweden
Stijn Hoppenbrouwers	Radboud University Nijmegen, The Netherlands
Tomasz Kaczmarek	Poznań University of Economics, Poland
Marite Kirikova	Riga Technical University, Latvia
John Krogstie	Norwegian University of Science and Technology, Norway
Birger Lantow	University of Rostock, Germany
Mikael Lind	Viktoria Institute, Sweden
Andreas Opdahl	University of Bergen, Norway
Anne Persson	Skövde University, Sweden
Erik Proper	Radboud University Nijmegen, The Netherlands
Nikolay Shilov	SPIIRAS, Russia
Dirk Stamer	University of Rostock, Germany
Janis Stirna	KTH, Sweden
Vladimir Tarasov	Jönköping University, Sweden
Wim Van Grembergen	University of Antwerpen, Belgium
Vivian Vimarlund	Jönköping International Business School, Sweden
Matthias Wißotzki	University of Rostock, Germany

FSFE 2013 Workshop Chairs' Message

The papers published in this volume were presented at the Second Workshop on Formal Semantics for the Future Enterprise (FSFE 2013) held in Poznań, Poland, in conjunction with 16th International Conference on Business Information Systems (BIS 2013).

The main goal of this workshop is to bring together researchers to share their knowledge and experience in the formal aspects of the future enterprise. The workshop welcomed research on any topic related to enterprise transformation and adaptation to the challenges of the future. The workshop emphasized the formal approach over research areas. This year, two papers were presented. One dealt with mobile applications and proposed a new way of distributing mobile tasks in the cloud. The second paper approached security in network communications and proposed a method for modeling and predicting modifications of security-related factors. The papers in this volume are extended and revised versions of the initial submissions, which incorporate reviewers' comments and outcomes of the discussion during the presentation at the workshop. Each paper went through a second review to make sure the suggestions were implemented. Overall, there were four submissions of which two were accepted. Thus, the acceptance rate was 50%.

June 2013 Razvan Petrusel

Organization

Chairs

Razvan Petrusel	Babes-Bolyai University of Cluj-Napoca, Romania
Gheorghe Cosmin Silaghi	Babes-Bolyai University of Cluj-Napoca, Romania

Program Committee

Razvan Petrusel	Babes-Bolyai University of Cluj-Napoca, Romania
Gheorghe Cosmin Silaghi	Babes-Bolyai University of Cluj-Napoca, Romania
Alvaro Arenas	IE Business School Madrid, Spain
Benjamin Aziz	University of Portsmouth, UK
Robert Andrei Buchmann	University of Vienna, Austria
Tomasz Kaczmarek	Poznań University of Economics, Poland
Jan Mendling	Wirtschaftsuniversität Wien, Austria
Mihaela Muntean	West University of Timisoara, Romania
Syed Naqvi	CETIC Charleroi, Belgium
Ioan Petri	Cardiff University, UK
Paul Pocatilu	Bucharest University of Economic Studies, Romania
Ioan Toma	University of Innsbruck, Austria
Nicolae Tomai	Babes-Bolyai University of Cluj-Napoca, Romania
Boudewijn Van Dongen	Eindhoven University of Technology, The Netherlands
Irene Vanderfeesten	Technical University of Eindhoven, The Netherlands
Krzysztof Węcel	Poznań University of Economics, Poland

LIT 2013 Workshop Chairs' Message

This volume contains the papers presented at LIT 2013: the 5th Legal Informatics and Legal Information Technology Workshop held on June 20, 2013, in Poznań, Poland. The event was co-organized with the 16th International Conference on Business Information Systems. It was a successful continuation of the LIT workshop series after a one-year break.

Judging by the numbers, LIT 2013 was comparable to its direct predecessor. As in 2011, six papers were submitted and reviewed. Out of this set, three works were accepted for presentation and publication after additional editorial review. This resulted in the low acceptance rate of 50%.

This level of improvement – in comparison with earlier LIT workshops acceptance rate – was the consequence of applying strict procedures for accepting papers [1]. We are certain that this factor inevitably lead to an interesting discussion during the workshop and specifically to a high quality of the works contained in this volume.

The work of Nečaský et al. presents an interesting approach of unifying the idea of Linked Open Data (LOD) with legal information. Such a combination allows one to merge legislative data from different sources and publishers. Another potential use is to enrich the already available legal information with public third-party datasets for further processing. The paper introduces the motivation for the creation of legislative LOD datasets as well as proposing the necessary infrastructure for Czech legislative dataset construction.

The work of Araszkiewicz et al. discusses the concept of a knowledge-based support system for divorcing parents. The system is developed to help in making decisions on the issues of the exercise of parental custody and contacts with children. The authors try to incorporate advanced knowledge so that the system will be able to reflect the problematic, in their opinion, legal concept of "well-being of the child." The paper concludes that the presented multi-layered knowledge base with case-based reasoning knowledge representation structures is sufficient to produce a system that operates on the basis of the above-mentioned concept.

The last paper by Żurek elaborates on instrumental inference. This kind of reasoning is very often employed in legal practice. It is also the one that seems to be problematic in terms of representation in a formal manner. Therefore, the author builds a sophisticated model of relations of necessity and sufficiency in order to propose a generic way to formalize instrumental inference. Additionally, the author discusses the issues related to the application of an instrumental

[1] Review notes were independently given by at least two reviewers assigned randomly from the Program Committee. Cases of possible conflicts of interest were avoided in advance. In some cases of largely diversified reviews, we took advice from additional reviewers.

reasoning model in argumentation with special attention paid to the conflict resolution area.

This year's event covered a wide area of topics with many challenges appearing on the horizon. The availability of such a forum is vital for supporting the development of knowledge in the area of IT and the law.

Finally, we would like to give our cordial thanks to all the authors as well as participants whose work and attendance enabled us to organize the workshop. We are aware that there would be no workshop in this shape without the participating community. We also would like to express our deepest thankfulness to the Program Committee members who supported us with their knowledge and judgment. Without the silent work of these people, the selection of the works would not be feasible. Finally, as always we would like to express our profound gratefulness to the organizers of the hosting conference (BIS 2013). Their support allowed for the organization of the 5th Legal Informatics and Legal Information Technology Workshop.

June 2013

Erich Schweighofer
Piotr Stolarski
John Zeleznikow

Organization

Chairs

Erich Schweighofer	University of Vienna, Austria
Piotr Stolarski	Poznań University of Economics, Poland
John Zeleznikow	Victoria University, Australia

Program Committee

Erich Schweighofer	University of Vienna, Austria
Piotr Stolarski	Poznań University of Economics, Poland
John Zeleznikow	Victoria University, Australia
Emilia Bellucci	Deakin University, Australia
Tania C.D. Bueno	IJURIS, Brazil
Marta Poblet	Autonomous University of Barcelona, Spain and RMIT University, Australia
Andrew Stranieri	University of Ballarat, Australia
Tomasz Żurek	Maria Curie-Skłodowska University, Poland

PhD Symposium Chair's Message

The 16th Conference on Business Information Systems (BIS 2013) that took place in June 2013 in Poznań, Poland, besides a number of scientific sessions, also hosted the PhD Symposium. The PhD Symposium was a response to the needs noticed during the previous editions of the conference, where we observed that it is very important for young researchers to present and discuss their PhD research results with more experienced colleagues.

The idea behind the symposium was therefore not only to present the research results but also to create links in the community and foster collaboration between young scientists from diverse research teams. We hope that contacts established during the symposium will result in future successful co-operation between young researchers (this concerns students as well as potential PhD advisors/reviewers that were also present).

Similarly to the main BIS conference, the papers submitted to the symposium covered a broad range of topics; however, a significant fraction of research devoted to semantic technologies may be observed. In total, 16 papers were submitted, out of which 14 were invited for presentation at the symposium and 10 were accepted for publication in the symposium proceedings. The Program Committee (PC) provided students with remarks that enabled them not only to improve the papers but also to shape further research in the topic. I would like to thank the PC members for this guidance!

Finally, the symposium was opened by the invited talk of Prof. Martin Bichler from the Technical University in Munich, which was an enlightening introduction to the talks and presentations that followed.

The papers included in this book are the result of the initial proposal submitted by the PhD students, discussions that happened during the symposium, and the reviews received from the PC. I hope that you will enjoy them and they will be an inspiration for your future research. The quality of the papers, as noticed by the reviewers, is as high as for the workshop papers.

June 2013 — Agata Filipowska

Organization

Chair

Agata Filipowska	Poznań University of Economics, Poland

Program Committee

Witold Abramowicz	Poznań Univeristy of Economics, Poland
Agata Filipowska	Poznań University of Economics, Poland
Rafael Accorsi	University of Freiburg, Germany
Matthias Book	University of Duisburg-Essen, Germany
Efthimios Bothos	National Technical University of Athens, Greece
Sabine Buckl	TU Munich, Germany
Flavius Frasincar	Erasmus University Rotterdam, The Netherlands
Anastasios Gounaris	Aristotle University of Thessaloniki, Greece
Wiesława Gryncewicz	Wroclaw University of Economics, Poland
Monika Kaczmarek	Poznań University of Economics, Poland
Marcello La Rosa	Queensland University of Technology, Australia
Andre Ludwig	University of Leipzig, Germany
Babis Magoutas	National Technical University of Athens, Greece
Baljeet Malhotra	SAP Research Singapore, Singapore
Viorel Milea	Erasmus University Rotterdam, The Netherlands
Marina Mongiello	Technical University of Bari, Italy
Joanna Palonka	University of Economics in Katowice, Poland
Adrian Paschke	Freie Universität Berlin, Germany
Florin Pop	University Politehnica of Bucharest, Romania
Esteban Robles Luna	LIFIA-F. Informatica, UNLP, Argentina
Artur Rot	Wroclaw University of Economics, Poland
Yiannis Verginadis	National Technical University of Athens, Greece
Krzysztof Węcel	Poznań University of Economics, Poland
Wiesław Wolny	University of Economics in Katowice, Poland
Hai Yun Lu	SAP Research Singapore, Singapore

Additional Reviewer

Victor Asavei	University Politehnica of Bucharest, Romania

Table of Contents

FSFE Workshop

LIT Workshop

PhD Symposium

Market Design - Foundations and Applications

Martin Bichler

Department of Informatics, TU München, Munich, Germany

Market Design is a discipline focusing on the theory, design, and development of auctions and markets. Examples of innovative market design include combinatorial auctions for selling spectrum as they have been used in the recent years worldwide, as well as multi-lot auctions in procurement and in transportation. Discrete optimization, game theory, lab experiments, and behavioral studies all play a role in the design of such markets. The talk focuses on spectrum auction design as a pivotal example for multi-object auction design. This area allows illustrating some of the fundamental problems in market design.

The simultaneous multi-round auction (SMRA) has been used to sell spectrum licenses worldwide since 1994. Complementarities in bidder valuations typically cause strategic problems for bidders in such auctions. In particular, a bidder with complementary valuations might be exposed to losses if he cannot bid on his packages of interest. This exposure problem caused much research in the recent years on the design of multi-item, more specifically combinatorial auctions. Much of this research has focused on the computational complexity of the winner determination problem and game-theoretical analyses of different auction formats. Initial experimental research showed significant efficiency increases when using combinatorial auctions in cases where bidders have complementary valuations [1].

This has led to a new generation of auction designs, which are currently being used for selling spectrum licenses world-wide. Since 2008, the combinatorial clock auction (CCA) has been introduced by regulators to sell spectrum licenses in countries around the world. The CCA combines an ascending clock auction for price discovery with a second sealed-bid phase, and a bidder-optimal core-selecting payment rule [2]. The auction format should (1) solve the exposure problem for bidders, and (2) provide incentives for truthful bidding.

The exponential growth of the number of packages bidders can bid on in a combinatorial auction does not allow results of small experiments with a few items only to carry over to larger auctions with many items. Only recent experimental work has focused on larger auctions beyond 10 lots. Scheffel et al. [3] found restricted package selection to be the main cause of inefficiency in a variety of combinatorial auction formats. Bichler et al. [4] analyzed the CCA and compared the results to those of SMRA based on value models similar to those in spectrum auctions in Europe. They found the efficiency of the CCA to be significantly worse than that of SMRA in a multi-band environment with four bands and six licenses each.

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 1–2, 2013.

This raises fundamental questions on the efficiency of combinatorial auctions in larger markets with many objects for sale. Spectrum auctions with up to 100 items have been conducted or are being planned in a number of countries such as Austria, Australia, Canada, Switzerland, the Netherlands, Ireland, or the UK. For example, in the Dutch auction in 2012 41 separate spectrum licenses in the 800 MHz, 900 MHz and 1800 MHz bands were up for auction. Switzerland auctioned 61 licenses in 11 bands in 2012. Bidders in such auctions might just not be able to submit bids on all possible combinations which have a positive valuation in such auctions. Note that with 30 licenses only, a bidder could already submit package bids on more than a billion packages. This can lead to inefficiencies as the winner determination allocates the spectrum as if all of these missing package bids had no value to bidders.

Given the exponential blowup of packages that a bidder is facing, there is little hope for high efficiency in larger combinatorial auctions with fully expressive bid languages. The design of compact bid languages which require a lesser number of bids could provide a workaround. In recent experiments, we analyze compact bid languages for multi-band spectrum auctions. The bid language allows expressing the main complementarities that typically exist within bands, but it does not allow expressing valuations for all possible packages. These experiments suggest that compact bid languages could address the exposure problem, but at the same time avoids the missing bids problem in fully combinatorial auctions. While there is no one-size-fits-all bid language for all domains, the experiments show that much efficiency can be gained by well-designed bid languages. This design parameter has often been ignored in previous discussions on spectrum auctions and market design overall.

References

1. Banks, J., Olson, M., Porter, D., Rassenti, S., Smith, V.: Theory, experiment and the federal communications commission spectrum auctions. Journal of Economic Behavior & Organization 51, 303–350 (2003)
2. Day, R.W., Cramton, P.: Quadratic core-selecting payment rules for combinatorial auctions. Operations Research 60, 588–603 (2012)
3. Scheffel, T., Pikovsky, A., Bichler, M., Guler, K.: An experimental comparison of linear and nonlinear price combinatorial auctions. Information Systems Research 22, 346–368 (2011)
4. Bichler, M., Shabalin, P., Wolf, J.: Do core-selecting Combinatorial Clock Auctions always lead to high efficiency? An experimental analysis of spectrum auction designs. Exp. Econ. 1–35 (2013)

Resilience - A New Research Field in Business Information Systems?

Günter Müller, Thomas G. Koslowski, and Rafael Accorsi

University of Freiburg, Germany
{mueller,koslowski,accorsi}@iig.uni-freiburg.de

Abstract. Being responsive in cases of unplanned disruptions has been difficult for management in the past, but for IT it is even more challenging: IT Systems are developed to fulfill predefined properties, and offer a hard-wired set of exception handling functionalities. Resilience encompasses reaction on disturbances beyond the scope of known properties. An organization is resilient if its capabilities can be adapted to new requirements which have not been explictly incorporated into the existing IT design. This paper introduces the concept of resilience and its implications in the fields of business information systems.

Keywords: IS Resilience, IS design, IS management, Resilient BPM.

1 Introduction

Modern societies operate in an increasingly complex and turbulent world marked by interconnection and interdependence across global networks [9, 23]. Information Technologies (IT) have an ambivalent effect on the performability of critical infrastructures: On the one hand, IT has become a platform of innovation and economic growth by means of efficient coordination and distribution of information. On the other hand, the increasing IT-enabled interconnectedness and interdependence is leading to the emergence of unintended, unpredictable safety and reliability problems [17, 31]. As a consequence, societal infrastructures are facing growing propensity for system breakdowns and sense of vulnerability to new and future threats such as terrorism, pandemic potential, energy volatility, and climate [27]. While such failures and breakdowns have proven relatively rare, the consequences of failures within an interconnected world can cause serious problems beyond geographical and functional borders [9].

Today, most decision makers, either public administrators or private organizations, have come to understand that protection of information systems (IS) is of high priority. But the expanding landscape of emerging risks illustrates the borderless and unpredictable nature of risk and uncovers the limits of traditional risk management practices and theories in the face of highly interconnected systems: *new emerging risks* or *new surprises* lack a priori indication of occurrence, they exhibit the potential to "cascade" systems at different speeds and their relation between origin, evolution and final consequence are frequently ill-understood

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 3–14, 2013.

[23]. But just because some systems are complex does not mean they are unmanageable or impossible to govern: Though, managing them requires different methods and rests on other assumptions than predominated in classical risk and security management. Where we had come to expect predictability and consistency, we now must accept the necessity of dealing with the consequences of uncertainty [23].

Against this background, the notion of resilience is getting attention as a denominator to move beyond survival and even prosper in the face of challenging conditions [17, 22]. Resilience is an emergent property associated with an organization's capacity to continue its mission despite disruption through mindfulness [34], resourceful agility, elastic infrastructures and recoverability, e.g. [12, 17]. Therefore, resilience is a combination of technical design features, such as fault-tolerance and dependability [6], with organizational features such as mindfulness, training and decentralized decision making [5, 34]. In contrast to traditional risk and security management approaches attempting to assess operational risks based on threat probabilities (focus on the *cause* of events), resilience shifts attention on systems interdependence and the potential of incidents to cascade (focus on the *consequences*). In line with [12], we define operational resilience management as "the processes by which an organization designs, develops, implements, manages, and improves strategies for protecting and sustaining high-value services and associated assets such as people, information, technology, and facilities" [12, p. 19].

The primary objective of this paper reflecting a keynote at the Business Information Systems Conference is to capture and to establish a relationship between resilience research and the business information systems field. The remainder of the paper introduces the various facets of resilience (Section 2), reports on the status quo wrt. research (Section 3) and provides a scientific-programmatic view of the upcoming research questions in this area (Section 4).

2 Conceptionalizing Resilience

2.1 Definitional Thicket

The concept of resilience has been a prominent and emerging topic in various scientific fields, but also on the agenda of public and private institutions, recognizing the uncertain and inter-connected nature of (socio-economic) systems. However, as resilience research encompasses a wide range of disciplines such as ecology, psychology or engineering, and different research contexts and topics, it is not surprising that the concept lacks an accepted common definition across disciplines ranging from descriptive to normative aspects, from strategies to heuristics [22] or even as a new paradigm for managing variability and uncertainty [17, 32].

Rooted from the Latin *resilire*, to spring back or to rebound, the term resilience refers to "the ability/power to recover form and position elastically" following a disturbance. The conceptual development (see for an overview [17, 20, 22, 24]) shows that the straightforward concept in physical science evolves to a complex multi-interpretable concept with contested definitions and even relevance. After looking at the definitions of resilience from a wide variety of disciplines, we can see that they almost always contain the basic idea of *bouncing back* from unanticipated challenges by means of absorbtion and recovery or *bouncing forward* by means of adaption and transformation [22]. Moreover, the relation between the different resilience concepts and its relation to *change* further illustrates the importance of temporal aspects [24]: systems are in a continuous state of change and require the development of multiple capabilities and strategies to cope *before, during* and *after* disturbance [20, 22]. We resume a more detailed discussion on temporal aspects in the subsequent sections when we highlight the different orientations between IS Resilience and IS Risk and Security Management and a concluding introduction of a Resilience Management Cycle.

2.2 Foundations of IS Resilience

The rationale for putting resilience into an organizational or IS research context has its roots in the study of safety-critical socio-technological systems characterized by high uncertainty [17, 23, 24]. Recent works and theoretical developments such as the "Normal Accident Theory" introduced by [25] and "High Reliability Organizations" [34] indicate that some failures are not only hard or impossible to predict, but also inevitable products in complex and tightly-coupled systems. Resilience is an emergent property associated with an organization's capacity to continue its mission despite disruption through mindfulness [34], resourceful agility and recoverability, e.g. [12, 17]. Therefore, resilience is a combination of technical design features, such as fault-tolerance and dependability [6], with organizational features such as mindfulness, training and decentralized decision making [5, 34]. Hence, this socio-technological conception of resilience has recently attracted IS scholars' attention (e.g. [5, 12, 15, 26]).

Interestingly, within the areas of IT/IS risk and security management, the tensions between IT-enabled productivity gains versus emerging vulnerabilities and risks are well recognized for decades. However, while IS architectures such as ERP and BPMS are often associated with significant performance improvements by means of standardization, high formalization, automatization and service decomposition [7], their potential for enhancing security and risk management is gaining momentum but still is not exploited thoroughly [15, 18]. A key challenge arises with the fact that existing approaches of information security and risk management (ISRM) mainly assume stable, predictable and isolated process types. This is sometimes in contrast to the business reality, as large organizations have often hundreds or more of interdependent processes in place. As a consequence, today's complex and fragile IS are prone to unforeseeable disruptions [12]. This is supported by Butler and Gray [11] who identified the paradox of relying on complex systems composed of unreliable components for reliable outcomes as a mostly neglected field in IS research so far [11]. The next section provides an overview of existing work on IS resilience and its shortcomings.

3 Status Quo and Shortcomings

Although resilience is widely recognized in related disciplines such as Computer Science [35], Contingencies and Crisis Management [9], or Safety Engineering [17], there is an apparent incongruity between the level of interest paid by business managers and the attention that organizational and IS scholars have given to resilience [32]. Today, only a limited number of resilience research exists. This research gap is surprising, as resilience is often said to be a combination of social or organizational and technical qualities and, therefore, a research topic well-suited for IS research. Hence, we provide a brief overview of existing work on resilience in IS research in order to identify current research gaps and challenges as a foundation for a IS resilience research agenda.

The majority of recent work on IS resilience and related research remains on a pure conceptual level. For example, a recent literature review on the related concepts IS reliability and mindfulness has been carried out by Butler and Gray [11], examining how organizations achieve reliability when operating in complex, fragile, and often unreliable IS environments. Although the authors intended to contribute to IS reliability, they introduced mindufulness, an organizational resilience concept into IS research [34]. Mindfulness describes an organization's capability to perceive cues, interpret them, and respond appropriately [11, 20]. Accordingly, they provide a foundational framework of IS reliability achieved by balancing routine-based strategies (focus on reducing variability and deviation) and mindfulness-based strategies (focus on cognitive and organizational capabilities for contextual sense-making). Based on a comprehensive literature review, they conclude that IS research provides little guidance for organizational reliability and highlight the need for conceptual tools and artifacts that help mindful management to support surviving and thriving in complex, dynamic environments. Similarly, Riolli and Savicki [26] emphasize individual and organizational characteristics against pressuring information system work environments. The authors conclude that much more empirical work has to be done to analyze the interrelations between stressors and resilience outcomes on an individual and organizational level. Based on a broader literature review on resilience across multiple disciplines, Erol et al. [14] propose a framework to discuss the moderating role of IS on assisting connectivity and collaboration in order to support resilience.

Another research stream addresses the issue of resilient IS architectural design. Inspired by biological systems, Zhang and Li [36] introduced a set of resilience axioms and derived five principles of engineered artifact systems. These principles encompass technical and managerial recommendations to increase system resilience such as inherent redundancy, flexible coordinative responsibilities, components for monitoring and continual training. Liu et al. [21] discuss the relationship between resilience and other similar architectural properties and present seven constraints to consider in the architecture for enhancing resilience. A set of fundamental requirements for supporting resilient business process management (BPM) is given by Antunes and Mourão [5]. While these works capture basic requirements for resilient IS design, they lack empirical validation, concrete implementation guidelines, as well as artifacts to support the implementation of resilience in IS, e.g. [5][12].

A further research stream focuses on measurement issues of resilience in the IS context. Wang et al. [33] present a measure in the context of enterprise information systems focusing on recoverability. They develop a formula that calculates the weighted relation between request time and completion time. A more advanced contribution with regard to measuring and visualizing resilience is given by Zobel et al. in the context of disaster events and cyber attacks [37]. Their approach captures multiple dimensions of resilience as a function of the predicted amount of initial loss and associated recovery time. Derived resilience curves provide further decision support for appropriate selection of countermeasures.

4 Research Questions and Objectives

Despite the wide spread of resilience across multiple disciplines, a number of open research issues remain. These encompass conceptual and definitional vagueness of resilience, a lack of empirical research and a lack of applicable (organizational) solutions and IT-artifacts to bring resilience into action. Hence, we articulate a research agenda on resilience and resilience management comprising four research questions spanning conceptual perspectives, research methods and prototypical implementations of a resilience supporting information system.

4.1 Divergent Understanding and Construct Clarity

The first research gap refers to the conceptual vagueness of resilience and organizational/IS resilience in particular. Researchers in varied and distinct disciplines have struggled with the concept of resilience in their respective fields for decades. Against the backdrop of manifold conceptual usage across multiple fields, it is not surprising that extant resilience research is surrounded by diversity and ambiguity of definitions, scope conditions, antecedents and outcomes. Is resilience a metaphor, a capacity, a capability, a strategy, a goal, a guiding principle, a philosophy, a measure or a behavior? Although an elastic notion of resilience may facilitate communication across disciplines (or even divergent lines of research within a discipline), a tradeoff may exist due to clarity confusion that may hinder operationalization and lead to unclear or even contradicting evaluations of results. A definition that is too broad could hinder empirical research results and even cause some to question the relevance of the concept [29]. Thus, as stated by Suddaby, a clear construct might not only facilitate communication between scholars, it also “enhances researchers’ ability to empirically explore the phenomena” and further enhance research creativity by “allowing managers to redefine problems in ways that are more amenable to resolution”. As a consequence, a deeper investigation and development of a wide-accepted definition and specification of (IS) resilience is crucial for both, theory development as well as further empirical analysis and artefact developments. This raises the following research questions (RQ):

RQ1: *How does resilience manifest itself across multiple disciplines? And how could resilience be defined and specified in order to fertilize IS research and practice?*

4.2 Empirical Exploration

In line with [29], the lack of construct clarity also impedes empirical exploration of organizational resilience. In their review of organizational resilience, [8] highlight that "there appears to be a strong focus around building theories and definitions of resilience. However, the literature is lacking in empirically proving the theories." Vogus and Sutcliffe [32] further posit that "given the dearth of empirical work exploring resilience in organization theory, many (if not all) avenues are open for future research in resilience". Dependent on the underlying theoretical assumptions, the nature of resilience will change. As for example Colbert [13] highlights changing implications for strategic human resource management due to a complex theory perspective of resource-based-view. Also, [10] exemplify the fundamental re-evaluation of organizational effectiveness based on the network-structure of the organizational system. Applying either a "Gaussian perspective" emphasizing linear-additivity and predictability or a "Paretian perspective" emphasizing non-linearity and emergence has strong implications for the analysis of system behaviors and structures and therefore substantially modifies the required variety to adapt and survive (p. 126, both widely described as related concepts of resilience). Thus, IS scholars should aim to untangle the underlying puzzle of organizational resilience and its related concepts, for instance vulnerability and adaptability by recognizing underlying assumptions about stability and normativity [31]. Furthermore, this requires to acknowledge different levels of abstraction (ranging from vague principles to less abstract policies, practices and outcomes) of resilience as well as the contextual scope of different levels of analysis [13, 20]. This research challenge concludes the following research questions:

RQ2: *How does resilience relate to other organizational and technical factors? More precisely, what are determinants and antecedents of organizational resilience and IS resilience in particular?*

4.3 Operationalization and Methodological Quest

The aforementioned trade-off between the potential of complex perspectives to enrich and question simpler assumptions at the expenses of academic rigor and a wide repertoire of quantitative statistics is already acknowledged in organization science, e.g. in [10], as well as IS research [31]. We expect a similar problem when operationalizing and measuring organizational resilience. For example, a simple version of resilience (as bounce back) may be well served in the more stable "Gaussian worlds" but may bring limits to a more complex or "Paretian world". While in particular technical indicators for the earlier resilience types are already established (e.g. in [37]), the development of more complex indicators and modeling techniques [24] still remain at a formative stage. Along with measurement issues, new method-sets from other disciplines such as mathematics, quantum mechanics and complexity science may enable new streams of resilience research in an organizational context. Future efforts may increasingly include empirical analysis and simulation modeling of resilience and its interaction with related constructs. Fruitful insights are to be expected for instance by a deeper investigation into the role of unique contextual settings, such as the comparative studies of resilience frameworks between different organization types (business vs. non-profit, service vs. industrial, etc.).

Independent of the particular resilience type, we expect that the incorporation of resilience as an important system feature will change the organizational object function and therefore leads to a re-evaluation and extension of organizational effectiveness. Evaluating effectiveness involves transforming managerial decisions into action and measuring the performance of that action. Performance measurement requires a systematic and deep analysis of business objects, which includes not only a re-structuring of processes but also the development of innovations in the light of resilience [4, 19]. Hence, both researchers and practioneers need to derive a set of meaningful indicators of organizational resilience on both, operational and strategic levels. As resilience is contextualized given a specific challenge, it is crucial to identify factors that are believed to enhance IS resilience, such as margin tolerance, buffering capacity and flexibility. Depending on the specific purpose, the meaningful indicators might be either quantitative or qualitative and claim to determine the gap between the expected and current status of the relevant business unit [4, 28]. This leads to the following questions:

RQ3: *How can resilience be translated to the principles and measurements of organizations and IS? What are appropriate scientific methodologies to recognize the complex and emergent nature of resilience?*

4.4 Resilient IS Design

The foundations of IS resilience introduced in Section 2.2 also have a variety of implications for the design of IS. Recent studies on resilience emphasize the integration of organizational and technological views, as well the integration of related, but usually disjointed activities of IS security, business continuity and IT operations [4, 12] as depicted in Fig. 1.

As for other business information systems, the elicitation of requirements and their system-wide enforcement are of utmost importance. At the technical level resilience requirements are intricate to capture, as they merge the capability of absorbing failures and unexpected situations (e.g. cushioning the ripple effect on the advent of change) with detecting the continuous deterioration of systems throughput (e.g. when running out of resources). While the former can be somehow estimated at design-time, the latter requires the a posteriori analysis and intervention. Interestingly, current literature reviews on risk-aware BPM by Jakoubi et al. [18] or Suriadi et al. [30] show that the vast majority of contributions

Fig. 1. Operational Resilience Management System

focus on design-time risk-management in BPM systems, while approaches at run-time and the exploitation of process-related log files a posteriori are largely neglected. Process-oriented resilience management might have the potential to fill these gaps.

According to the CMU-CERT Resilience Management Model [12], an operational resilience requirement is defined as a constraint that an organization places on the productive capability of assets to ensure viability when charged into business processes. These requirements provide the foundation for how to enhance the resilience of assets and related processes. They embody organizational objectives, risk appetite and tolerance, critical success factors, and operational constraints [12]. Moreover, Antunes [5] propose fundamental requirements for resilient BPM: They support (i) various levels of severity, ranging from simple failures of key resources to catastrophic accidents; (ii) the coexistence of stable processes with unstable changes in the operating environment; (iii) the dynamic construction and update of situation awareness; (iv) assistance for knowledge representation and management, a fundamental drive to decision-making [15]; (v) flexible operations and unplanned tasks whenever necessary; (vi) the opportunity to experiment with and learn from the novel, innovative and challenging situations that emerge from hazards; and finally (vii) the transition from emergency to normal operations.

Up to now, there are techniques and formal foundations that can, when assembled, provide for resilience mechanisms at the level of BPM. However, the current state of the art do not offer corresponding mechanisms. Similarly, vendors of BPM systems and workflow management systems have not yet focused their solutions on resilience.

Based on a literature review, we developed a resilience management cycle (depicted in Fig. 2) for automated support for resilient BPM according to the well-established BPM lifecycle. The cycle contains four phases adapted primarily from [5], beginning with (i) *Detection* in order to identify failures, potential

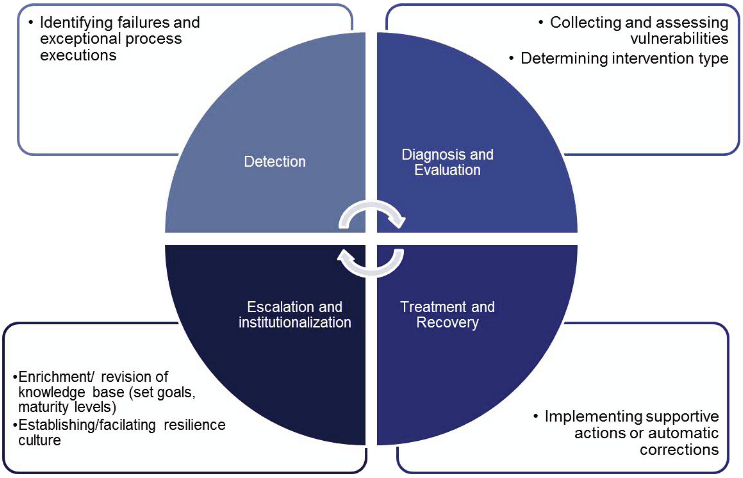


Fig. 2. The Resilience Management Cycle

weaknesses and exceptional process executions. (ii) The purpose of *Diagnosis and Evaluation* is to collect and assess vulnerabilities, and consequently to determine a set of intervention types. (iii) The next stage covers *Treatment and Recovery*, including the actual selection and implementation of supportive actions and automatic corrections. (iv) Finally, the phase of *Escalation and Institutionalization* guarantees enrichment or revision of the current knowledge base, and aims to establish and facilitate an organization-wide resilience culture.

In accordance with the resilient management cycle, it is natural to focus on the detection stage first. The purpose of this phase is to collect, record, and distribute information about the operational resilience of BPM on a timely basis. Effective resilience detection provides essential information about changes/deviations [17, 24], such as hazard occurrences and exceptional process executions, but also potential weaknesses such as high utilization at the margin of resources' or processes' capacity. Data collection, logging, and measurement are at the heart of resilience detection: they addresses the organization's competencies for identifying, collecting, logging, and disseminating information needed to ensure that operational resilience management processes are performed consistently and within acceptable tolerances [12]. This requires an effective measurement and analysis process that transforms operational resilience objectives and requirements into visible measures. Measures need to express the gap between intended process-goals and actual process-goals.

Works on BPM re-engineering [16], risk-aware [7], and resilient BPM in particular [12], provide a solid basis for measures for the attempted resilience detection service. However, deriving meaningful measures for resilience detection requires the alignment with organizational goals and missions [12]. As these objectives need to be interpreted and aligned for a specific organization, the well-established objective-driven approach suggest by [4] seem promising. The rationale behind it is to assure that resilient measures for extraction and detection have a direct link with operational goals and therefore impact the resilience of diverse organizational missions.

Another implication for resilient IS design arises with the concept of mindfulness, an organization's capability to perceive cues, interpret them, and respond appropriately [11, 20]. Ongoing research aims at elaborating the conception and implementation of intuitive user interfaces. For instance, process resilience detectors as an a posteriori checking module could complement and support established risk-aware BPM architectures. In contrast to those approaches with emphasis on design-time analysis to calculate operational risks based on (either subjective or historical) threat probabilities (focus on the cause of events) [30], the a posteriori resilience approach will focus on the business processes' interdependencies and potential to cascade, so-called *ripple effect* [20][17]. Such detection service addresses further questions, such as: (i) Do the actual process models correspond with the intended concepts?; (ii) Does the observed system behavior meet requirements of the respective compliance or security standard?; and (iii) Can we derive further information about the dynamic system behavior (e.g. recovery time, rate of degradation)? In order to extract the interdependencies and dynamics, new tools such as exploratory mining techniques for conformance checking [1] as well as process discovery [2, 3] might be suitable approaches.

RQ4: *What are fundamental requirements for resilient IS design? How can they be integrated into the architecture of IS? What tools and approaches are applicable to support and enhance IS resilience?*

5 Conclusion

Resilience is getting attention as a denominator to survive and even thrive in face of turbulent and complex operational environments. This paper sets out to argue that resilience can be featured as a new and valuable research field in Business Information Systems. Despite the wide spread of resilience across multiple disciplines, IS are a crucial but still unadequately explored enabler for organizational resilience. As we highlighted in this paper, a number of open research issues remain. These encompass conceptual and definitional vagueness of resilience, a lack of empirical research, and a lack of applicable (organizational) solutions, as well as IT-artifacts. Based on the plethora of research gaps, we proposed a research agenda on resilience and resilience management comprising four major research directions. The research agenda and associated research questions provide a basis for a new consolidated discourse on IS resilience.

References

1. Accorsi, R., Stocker, T.: On the exploitation of process mining for security audits: the conformance checking case. In: ACM Symposium on Applied Computing, pp. 1709–1716. ACM (2012)

2. Accorsi, R., Stocker, T., Müller, G.: On the exploitation of process mining for security audits: the process discovery case. In: ACM Symposium on Applied Computing, pp. 1462–1468. ACM (2013)
3. Accorsi, R., Ullrich, M., van der Aalst, W.M.P.: Process mining. Informatik Spektrum 35(5), 354–359 (2012)
4. Allen, J.H., Curtis, P.D., Gates, L.P.: Using defined processes as a context for resilience measures (2011)
5. Antunes, P., Mourão, H.: Resilient business process management: Framework and services. Expert Syst. Appl. 38(2), 1241–1254 (2011)
6. Avizienis, A., Laprie, J.-C., Randell, B., Landwehr, C.: Basic concepts and taxonomy of dependable and secure computing. IEEE Trans. Dependable Secur. Comput. 1(1), 11–33 (2004)
7. Balasubramanian, S., Gupta, M.: Structural metrics for goal based business process design and evaluation. Business Process Mgmt. J. 11(6), 680–694 (2005)
8. Bhamra, R., Dani, S., Burnard, K.: The concept, a literature review and future directions: International Journal of Production Research. Int. J. of Production Research 49(18), 5375–5393 (2011)
9. Boin, A., McConnell, A.: Preparing for critical infrastructure breakdowns: The limits of crisis management and the need for resilience. J. Contingencies & Crisis Management 15(1), 50–59 (2007)
10. Boisot, M., McKelvey, B.: Connectivity, extremes, and adaptation: A power-law perspective of organizational effectiveness. J. Management Inquiry 20(2), 119–133 (2011)
11. Butler, B.S., Gray, P.H.: Reliability, mindfulness, and information systems. MIS Quarterly 30(2), 211–224 (2006)
12. Caralli, R.A., Allen, J.H., Curtis, P.D., Young, L.R.: CERT Resilience Management Model, version 1.0 (2010)
13. Colbert, B.A.: The complex resource-based view: Implications for theory and practice in strategic human resource management. Academy of Management Review 29(3), 341–358 (2004)
14. Erol, O., Sauser, B.J., Mansouri, M.: A framework for investigation into extended enterprise resilience. Enterprise Information Systems 4(2), 111–136 (2010)
15. Fenz, S., Neubauer, T., Accorsi, R., Koslowski, T.: FORISK: Formalizing information security risk and compliance management. In: Annual IEEE/IFIP Conference on Dependable Systems and Networks (2013)
16. Harrington, H.J.: Business process improvement: The breakthrough strategy for total quality, productivity, and competitiveness. McGraw-Hill (1991)
17. Hollnagel, E., Woods, D.D., Leveson, N. (eds.): Resilience engineering: Concepts and precepts, Ashgate, Aldershot, England, Burlington, VT (2006)
18. Jakoubi, S., Tjoa, S., Goluch, G., Quirchmayr, G.: A survey of scientific approaches considering the integration of security and risk aspects into business process management. In: Workshop on Database and Expert Systems Application, pp. 127–132.
19. Koslowski, T., Strüker, J.: ERP on demand platform - Complementary effects using the example of a sustainability benchmarking service. Business & Information Systems Engineering 3(6), 359–367 (2011)
20. Koslowski, T.G., Geoghegan, W., Longstaff, P.H.: Organizational resilience: A review and reconceptualization. In: Conf. Strategic Management Society (2013)
21. Liu, D., Deters, R., Zhang, W.J.: Architectural design for resilience. Enterprise Information Systems 4(2), 137–152 (2010)

22. Longstaff, P.H., Koslowski, T.G., Geoghegan, W.: Translating resilience: A framework to enhance communication and implementation. In: Symposium on Resilience Engineering (2013)
23. McCann, J.E., Selsky, J.W.: Mastering turbulence: The essential capabilities of agile and resilient individuals, teams, and organizations. Jossey-Bass (2012)
24. Meyer, J.F.: Model-based evaluation of system resilience. In: Annual IEEE/IFIP Conference on Dependable Systems and Networks (2013)
25. Perrow, C.: Normal accidents: Living with high risk technologies. Princeton University Press (2011)
26. Riolli, L., Savicki, V.: Information system organizational resilience. Omega 31(3), 227 (2003)
27. Risk Response Network. Global risks 2012. World Economic Forum (2012)
28. Somers, S.: Measuring resilience potential: An adaptive strategy for organizational crisis planning. J. Contingencies & Crisis Management 17(1), 12–23 (2009)
29. Suddaby, R.: Editor's comments: Construct clarity in theories of management and organization. Academy of Management Review 35(3), 346–357 (2010)
30. Suriadi, S., Weiss, B., Winkelmann, A., ter Hofstede, A., Wynn, M., Ouyang, C., Adams, M., Conforti, R., Fidge, C., La Rosa, M., et al.: Current research in risk-aware business process management-overview, comparison, and gap analysis. QUT ePrints, 50606 (2012)
31. Tanriverdi, H., Rai, A., Venkatraman, N.: Research commentary - Reframing the dominant quests of information systems strategy research for complex adaptive business systems. Information Systems Research 21(4), 822–834 (2010)
32. Vogus, T., Sutcliffe, K.: Organizational resilience: Towards a theory and research agenda. In: IEEE Int. Conf. Systems, Man and Cybernetics, pp. 3418–3422 (2007)
33. Wang, J., Gao, F., Ip, W.: Measurement of resilience and its application to enterprise information systems. Enterp. Inform. Sys. 4(2), 215–223 (2010)
34. Weick, K.E., Sutcliffe, K.M.: Managing the unexpected: Resilient performance in an age of uncertainty. Jossey-Bass (2007)
35. Wolter, K.: Resilience assessment and evaluation of computing systems. Springer (2012)
36. Zhang, W., Lin, Y.: On the principle of design of resilient systems – Application to enterprise information systems. Enterp. Inform. Sys. 4(2), 99–110 (2010)
37. Zobel, C.W., Khansa, L.: Quantifying cyberinfrastructure resilience against multi-event attacks. Decision Sciences 43(4), 687–710 (2012)

Analytic Hierarchy Process to Improve the Evaluation Process in R&D Management System

Tania C.D. Bueno, Fabricio T. Donatti, Thiago P.S. Oliveira, Angela I. Zotti, and Hugo C. Hoeschl

Institute of Electronic Governance, Intelligence and Systems, Rua Lauro Linhares, 728 sala 105, Florianópolis, Santa Catarina, Brazil
{tania.bueno,fabricio.donatti,thiago.paulo, iara.zotti,hugo.hoeschl,i3g}@i3g.org.br

Abstract. The present work was developed to assist in the selection of projects in R & D for Electric Power companies in Brazil. On every evaluation system it is fundamental the correct selection of criteria, relevant and adequate for each situation. The normalization of such criteria will guarantee quality on the promotion of technological innovation in electric sector, which plays a key role for the integration of economic, social and environmental factors inside Electric Power Companies. Therefore, a project evaluation model was developed based on knowledge representation methodologies and multicriteria decision problem. This model used aspects from Albertyn, Saaty and Bueno to support the process of quality measurement of R&D Projects.

Keywords: Analytic Hierarchy Process (AHP), knowledge representation, multicriteria decision, projects evaluation, decision making.

1 Introduction

The Brazilian Electric Energy National Agency – ANEEL defines that the proposed Research and Development (R&D) projects should be evaluated solely based on predefined criteria. The analysis in these criteria must be done considering the main product of the project, the state of the art, challenges and proposed advances in scientific and/or technology terms. The problem to be solved and the absence or the high cost solution available in the market must be considered when is relevant. However, the Brazilian Electric Power companies have difficulty in sustaining a significant portfolio of investments in innovation and the main problem is the delay in the evaluation of projects [1], [2]. One reason for this problem is the complexity of interpretation and contextualization of what is innovation for these companies.

The selection problem of R&D projects has being seen as a multicriteria decision problem. In fact, due to a number of criteria ranging from financial investments and materials to possible competitive advantages, this work intend to prioritize projects according to their importance to be implemented at first moment [3]. On every evaluation system it is fundamental the correct selection of criteria, relevant and adequate for each situation. The normalization of such criteria – for innovation project

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 15–22, 2013.

selection – will guarantee quality on the promotion of technological innovation in electric sector, which plays a key role for the integration of economic, social and environmental factors inside Electric Power Companies. Therefore, a project evaluation model was developed. This model used aspects from Albertyn [4],[5], Saaty [6],[7],[8] and Bueno at al [9],[10],[11] to support the process of quality measurement of R&D Projects. In order to turn this process evident, this work was divided on the following sections: on Introduction is presented the subject of the present work, followed by the presentation of the R&D projects evaluation process. The third item explains the methodology. The fourth and fifth items present the Results and Conclusion, respectively. Finally, the sixth item presents the References consulted for this study.

2 Brasilian Electric Power Innovation Analysis

The evaluation process for R&D projects is the main bottleneck for the Electric Energy institutions in Brazil because of its subjectivity and lack of relevant and adequate criteria for each situation. To help on this demand, the Electric Energy National Agency (ANEEL) [1],[2] has defined some criteria, so that the companies in electric sector could evaluate R&D projects. These criteria, however, are too broad and subjective, becoming difficult the creation of patterns for the evaluation process for the electric energy companies. The absence of solid criteria comes out with delays, and in many cases, might cause the selection of inappropriate projects.

2.1 Criteria by ANEEL

The criteria established and adopted by ANEEL are described in sequence, which currently parameterize the selection and evaluation process of the companies. The evaluation of a R&D project is performed based on the four criteria: 1.Originality, 2.Applicability, 3.Relevance and 4.Reasonableness of Costs. The projects are evaluated at the end of the period of implementation. But, there is also the possibility of an initial evaluation, if requested, not obligatory though. Nevertheless, the initial evaluation is highly recommended in order to evaluate the framework of the project as R&D activity, its relevance to the technological challenges facing the sector and the reasonableness of the planned investments in the face of expected results and benefits.

Currently, in this research, the projects are submitted to a committee of evaluators supported by a management system. These evaluators are chosen from a list of employees with great knowledge of the area evaluated. The analysis of the project is contextualized exclusively within the themes and subthemes allowed by ANEEL and, in some cases, defined by the power company. Strategic themes or subthemes are those whose development is of national interest and of great importance for the electricity sector, involving highly complex in terms of scientific and/or technological and low attractiveness for investment (see table 1 and 2).

Table 1. Priority themes for investment in R&D from ANEEL

Priority Themes
FA- Alternative energy sources of eletric power generational
GT - Thermal Generation
GB - Watershed Management and Reservoirs
MA - Environment
SE - Safety
EE - Energy Efficiency
PL - Planning of Power Systems
OP - Operation of Power Systems
SC - Supervision, Control and Protection of Power Systems
QC - Quality and Reliability of Electric Energy Services
MF - Metering, Billing and combat commercial losses
OR - Other

Table 2. Examples of ANEEL priority subthemes

Themes	SubThemes
FA	FA01 - Alternative energy sustainable service to small isolated systems
	FA02 - Generating energy from urban solid waste
	FA03 - New materials and equipment for power generation from alternative sources
	FA04 - Technologies for use of new fuels in generating plants
	FA0X - Other
GT	GT01 - Evaluation of risks and uncertainties continuing to supply natural gas for thermal generation
	GT02 - New techniques for increasing efficiency and decreasing emissions of pollutants from power plants to fuel derived from petroleum
	GT03 - Optimization of power generation in industrial plants: increased efficiency in cogeneration
	GT04 - Micro-cogeneration systems residential
	GT05 - Techniques for capture and sequestration in plants
	GT0X - Other

2.2 Electric Power Company Evaluation Process

The project proposal is reviewed by the Company area where the knowledge to be gained in the reserch, and/or the potential outcomes may be used. If there is a favorable analysis by the area, the R&D Committee of evaluators reviews the project proposal based on parameters and evaluation criteria provided in the Research and Development Guide Program. If the Commission considers relevant, the proposals coordinators are invited for a presentation [10].

After this step and after the Regulatory Agency (ANNEL) initial evaluation, the R&D Commitee establishes a conclusive record containing a list of the projects considered "able" to join the R&D Program in the Company. Projects can only be considered "able" if they answer the following premises:

a) To obtain at least the concept “3” in Originality criterion given by ANEEL;

b) To achieve grade equal or higher than 2.8, the average applied to the concepts that guide the project (Originality, Applicability, Relevance and Reasonableness of Costs);

c) The favorable assent of the R&D Commitee. In final evaluation in case of inadequacy as an R&D activity the project is disapproved. The Originality criterion has eliminatory character, so the project must have grade equal to or higher than three (3) and still be characterized as R&D project.

Otherwise the value of the project should be reversed for the account of R&D [11].

2.3 CELESC R&D Management System

The Celesc R&D Management System consists of two separate structures: one for management, and the other for searching information on R&D documents – a Web page accessible without password by the employees. These both environments serve as management instruments for consultation about the R&D information and provide to keep track – daily – on the job produced by all the people involved on production, storage and recovering of strategic information, inside the parameters defined by ANEEL.

The Celesc R&D Management System includes Artificial Intelligence artifacts, which gives best quality and answering speed through the organization of information in form of cases, and allows recovering information based on similarity measures. The system is based on ontologies [9] for indexing and storing the documents inserted on the knowledge base. This form of information representation allows a more organized storage of cases (the R&D Projects), becoming their recovery easier and presenting clear answers. In this sense, considering the complexity of activities involved here, the Celesc R&D Management System is an integrated suite for organizational knowledge management, which supports all the knowledge construction process, and becomes decision making process faster. The management environment is composed by a set of submodules. The submodule *Project Evaluation* is one of them and the object of the present research. In this work, evaluation forms were built to integrated metadata which was already available in the organization with an ontology automatically extracted from the corpus in order to improve ANEEL´s criteria.

3 Methodology

The structure of the Electric Power Company´s R&D Management System allowed the inclusion of an evaluation process through a framework that enables the normalization of the weights of questions and answers. The reason for joining was to realize the occurrence of some classes of the metamodel that present connections with classes pointing to quality aspects adapted to the Electric Power Company R&D Project. Some purpose and expected outputs from the process are defined before starting the analysis.

The knowledge engineering team determined that the objective pursued by the application of the model was to identify a set of questions and expressions which could fit the criteria set by ANEEL through knowledge engineering methodology where the definition of the relevance of the expressions is related to frequency of use of these expressions in context.

Set the goal of the analysis - it was the choice of the projects that were using the process - we chose to work on projects that were running. This process was consolidated through a list of questions and answers. After that, the MCDA worked by Albertyn [5] is a method to support evaluators who are faced with many different and conflicting solutions to a problem.

Deriving this method, we chose to focus the evaluation on the application of the AHP, because the method is based on the innate human ability to make judgments about different problems [7][8].The method is characterized by decomposing the problem into descendant hierarchical levels, starting with the overall goal, criteria, subcriteria and alternatives in successive levels, until it reaches a prioritization of its indicators, approaching a better response. The AHP has the advantage of allowing comparison of quantitative and qualitative criteria. At each stage (criteria and subcriteria) pairwise comparisons are made to determine the relative importance of each criterion to reach the goal. After this hierarchical structure phase, AHP includes other two important steps: the judgment of value and priority, where the evaluator establishes a peer comparison of elements of the various hierarchical levels, prioritizing them and in sequence, the analysis of consistency of these trials. This pairwise comparison is done according to Saaty Scale [6],[7] that allows the conversion of the analysis on a scale from 1 to 9, ending in an array for each level of criteria and sub-criteria, showing the result of the comparisons made in pairs.

Through pairwise comparisons, the priorities calculated by AHP capture subjective and objective measures and demonstrate the strength domain of a criterion over another or one alternative over another. Through synchronization meetings, the teams began the process of knowledge engineering, with the task of standardizing the language that would apply in the development of Knowledge-Based System - KMS. We identified the main concepts used in the domain of evaluation and selection of R&D projects and their understanding and determination of national and regional context of the procedures adopted so far the Electric Power Company. Prospecting criteria, in agreement with the criteria established by ANEEL, and definition of the Energy Power Company strategic criteria were issues discussed.

Inventories of contents, processes and people were conducted following the premises of knowledge sharing, visualization and definition of relevance. It is the synchronization of these factors that enables the knowledge understanding or expertise in a particular field. There is a need to change the formal model of innovation management in the Electric Power Company to support and become innovation more effectively within the parameters of ANEEL and thus adapting it to the national context.

4 Results - AHP Application for R&D Evaluation

While inventories allowed the selection of the questions available to the R&D Projects Evaluation System and therefore the adoption of criteria to be evaluated and prioritized, the application of AHP method allowed calculating the weights for each alternative of evaluation project in relation to the proposed criteria. At this stage, based on the questionnaire already prepared, we performed pairwise comparisons using the intensities of importance by Saaty Scale. The pairwise comparison followed the next hierarchy: pairwise evaluation among criteria, pairwise evaluation among questions within the criteria, pairwise subcriteria evaluation. Thus, the intensities of importance were converted into numbers, according to that scale. The test calculations were performed in Excel software. The tables present the comparison among all the criteria (Originality, Applicability, Relevance and Reasonableness of Costs). The values resulting from the evaluation of researchers and the matrix obtained from calculations that make up the method. According to the data presented the Originality criterion is the one that has a greater impact over the other criteria defined by the company. This result quantifies the guidance established by the energy national agency and the extreme importance of the originality criterion for the evaluation and selection of R&D projects.

The comparison among the questions that form the Originality criterion shows the “Unpublished” as the most important as “Unprecedented, Intellectual Property, Innovativeness and Scope”.

The alternatives inside each question of each criterion were considered as subcriterion. The result of pairwise comparison made among the subcriteria of Originality criterion that evaluates the innovative character of the proposed project (Innovativeness). The most relevant subcriterion was “It is well defined and innovative, generating new product, process and methodology”. All criteria have issues and these issues have subcriteria. A comparison is made between the issues of each Criterion and between the subcriteria of each issue to calculate the most relevant.

4.1 Evaluation

In previous items, AHP and the pairwise comparison method have been described in detail. Also the necessary factors and constraints for the case study have been determined. The importance of these factors over each other has been analyzed and the comparison ratios have been assigned to form the pairwise comparison matrix. We performed a comparison of the results in a more conceptual manner in order to achieve greater user acceptance. The test samples were summarized in selecting the projects, applying the solution implemented and checking if the result of this application is corresponding to the result expected from the model.

Due to the complexity of the execution of a real project evaluation, we conducted a simulation based on randon choise made by user from the data in excel spreadsheet. We compared the data with the previous case that was a simple sum of the evaluation questions.

Buckley [12] has raised questions about certainty of the comparison ratios used in the AHP. He had considered a situation in which the decision-maker can express feelings of uncertainty while he/she is ranking or comparing different alternatives or criteria. The method he has used to take uncertainties into account is using fuzzy numbers instead of crisp numbers in order to compare the importance between the alternatives or criteria. Saaty [7] is against fuzzifying the AHP. He concludes that fuzzification of the process does not give much different results. He also believes AHP is already a fuzzy process because most ratios for ranking are not absolute or crisp numbers. At final stage it was also made final validation of the functionality of the system with the CELESC team.

5 Conclusions

This methodology allowed the identification of the project that best match with the criteria defined by the Electric Power Company, considering the different weights for each alternative. As main result it produced a tool that allowed simulating scenarios depending on the questions to evaluate R&D projects.

Every new scenario allows using the same methodology to support decision making in other situations, specific to each edict or company's proposal. It means that depending on the proposal, another criterion as Relevance can be considered the most important. Therefore, each scenario created can permit the use of the methodology applied in this study as support for decision making in other circumstances.

Thus, for each new scenario will be necessary to compare the questions according to each edict adopted, calculating the weights again and subsequently, ranking criteria, and producing a new result.

Since ANEEL´s R&D Program is recent, results from this research have contributed effectively for the Electric Power Company intention, on the challenge for an evaluation and selection of submitted projects with good quality, about a task previously assumed by ANEEL.

Acknowledgments. Our thanks to Centrais Elétricas de Santa Catarina – CELESC, especially to Celesc Team: Mr. Amaro Koneski Filho, Mr. Luiz Afonso Pereira Athayde Filho e Mrs. Débora Simoni Ramlow.

References

1. ANEEL - Agência Nacional de Energia Elétrica (2002). Atlas de Energia Elétrica do Brasil.ANEEL, Brasília: ANEEL (2002), `http://www.aneel.gov.br/arquivos/PDF/livro_atlas.pdf` (retrieved)

2. ANEEL - Agência Nacional de Energia Elétrica, Guia do avaliador de projetos de P&D:Programa de Pesquisa e Desenvolvimento Tecnológico do Setor de Energia Elétrica. Brasília:ANEEL, p. 30 (2010)
3. Gusmão, A.P.H.: Modelo Multicritério para Priorização de Portfólio de Projetos de Pesquisa e Desenvolvimento. Recife. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Pernambuco, 101p. (2005)
4. Albertyn, F.E.: Process Selection using Case-Based Reasoning Techniques. In: IADIS International Conference E-Society 2007, Part of the IADIS MultiConference on Computer Science and Information Systems (MCCSIS 2007), Lisbon, Portugal, July 3-8 (2007), `http://www.esocietyconf.org/`
5. Albertyn, F.: A Decision Making Approach: a meta-model for e-process selection. In: IADIS International Conference e-Commerce, Freiburg, Germany, July 26-30 (2010) ISBN: 978-972-8939-24-3
6. Saaty, T.L.: Theory and Applications of the Analytic Network Process: Decision Making with Benefits, Opportunities, Costs, and Risks. RWS Publications, Pittsburgh (2005)
7. Saaty, T.L., Tran, L.T.: On the invalidity of fuzzifying numerical judgments in the Analytic Hierarchy Process. Mathematical and Computer Modelling 46, 962–975 (2007)
8. Saaty, T.L.: Extending the Measurement of Tangibles to Intangibles. International Journal of Information Technology & Decision Making 8(1), 7–27 (2009), `http://ssrn.com/abstract=1483438`
9. Bueno, T.C.D., Stradiotto, C.R.K., Bedin, S.P.M., Mirapalheta, V.O., Hoeschl, H.C., de Oliveira, T.P.S., Athayde Filho, L.A.P., de Carvalho Jr., P.A.S.: Case-based System for Innovation Management in Electric Power Enterprises. In: Abramowicz, W., Tolksdorf, R., Węcel, K. (eds.) BIS 2010. LNBIP, vol. 57, pp. 3–6. Springer, Heidelberg (2010)
10. Bueno, T.C.D., Bueno, C.O., Zotti, A.I., Mirapalheta, V., Oliveira, T.P.S., Hoeschl, H.C.: Using Multi-Criteria Decision and Knowledge Representation Methodologies for Evaluation of Innovations Projetcs. Lecture Notes in Information Technology, vol. 14, pp. 1–621 (2012)
11. Bueno, C.O., Bueno, T.C.D., Hoeschl, H.C.: Ontologies to Automating Categorization of Innovation Projects in Knowledge-Based Systems. In: XXXVIII CLEI Conferencia Latinoamericana em Informática (Simposio Latinoamericano en Informática y Sociedad), Medellin, Colombia. Electronic Notes in Theoretical Computer Science, pp. 482–490 (2012)
12. Buckley, J.J.: Fuzzy Hierarchical Analysis. Fuzzy Sets and Systems 17(3), 233–247 (1985)

Using Software Agents to Enhance the Functionality of Social Knowledge Portal

Mariusz Żytniewski and Radosław Kowal

University of Economics in Katowice
ul.1 Maja 50
40-287 Katowice Poland
{mariusz.zytniewski,radoslaw.kowal}@ue.katowice.pl

Abstract. Modern knowledge-based organizations require IT solutions that support execution of their business processes through enabling access to codified knowledge regarding these processes. Such knowledge can be gathered and shared using social software such as wikis and blogs. But user generated content often leads to granularity of knowledge. It is often necessary to provide the bigger picture and, especially when organization is divided into many teams, it's necessary to provide a competence map. One of the concepts that can be applied to address these problems is software agents society, operating on the top of the social software platform. In this paper we present a concept of the architecture of such solution.

Keywords: software agent societies, knowledge based organization, knowledge portal, social software, wikis, blogs, tags.

1 Introduction

The recent advancements in the theory of business process orientation in knowledge based organizations indicates the possibility to include in this research area the concept of software agent societies, which use mechanisms of semantic knowledge representation. As an area of application for such solutions we propose an improvement to the processing of knowledge regarding business processes. Processing of knowledge carried out under the condition of pervasive communication, semantic Internet and Web 4.0 is an important tendency in contemporary researches.

The knowledge based organizations not only utilize intellectual capital to create products, services but also consciously manage this kind of capital and try create conditions that enables organizational learning. In other words knowledge based organization are those organizations which adjust their offer, strategy and procedures to the knowledge that is an outcome of their reflections on the experience and which manage their knowledge in a conscious way.

Knowledge portals help such organizations in simplifying the process of gathering and storing of codified (explicit) organizational knowledge. Knowledge portal offers the following advantages to organizations:

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 23–34, 2013.

- Simplified management of organizational knowledge [1], through the creation of social networks [2], [3] and communities of practice [8] in an organization,
- it facilitates the work of the knowledge workers [4], [5],
- it supports the changes in the corporate culture and knowledge exchange among organization's members [6],
- it enables access to knowledge through the mobile devices and from any place any time,
- it assures access to new communication channels, that helps with archive portions of created knowledge [7],
- it supports business process modeling based on defined knowledge about these processes in dynamically changing environment and assures mechanism of knowledge storing and sharing.

But the main problem regarding the usage of such portal is efficient browsing and searching through the existing knowledge.

In this paper we present a concept of applying the distributed, autonomic and mobile IT solutions to enhance the ability of social software platform in the area of knowledge management. Such solutions are usually referred to as software agents. We describe a case of a consulting company as a perfect example of knowledge based organization. Our aim is to present how social software can be used for knowledge management at the consulting company and how using software agents can enhance its usability.

The goal of applying software agents at the top of social software platform is to help its users in the process of knowledge browsing and in searching for knowledge. It is often relatively hard for users to find knowledge that is useful for their specific needs. For example wikis can comprise of hundreds of pages but without proper classification user can be overwhelmed with the abundance of information. Information overload is a common problem which can be solved by distillation of information based on user's characteristics – users' profiles. Given the specific nature of the organization being a subject of our research we propose two areas of improvements that are to be achieved due to application of software agents: a) creation and taking advantage of users' profiles, b) creation of competence maps which helps people with finding a right person for solving specific problem.

First part of this paper describes the state-of-art in the research area of application of software agents for supporting knowledge based organization. In the second part we presented process-oriented social software portal and how we believe it should be designed in order to effectively address the problem of knowledge management. The last part is devoted to present the concept of how software agents society can be utilized to automate searching for persons holding key competences based on the analysis of the links between knowledge parts in the social software platform.

2 Software Agents as a Support for Knowledge Based Organizations

The issue of the use of software agents to support knowledge based organizations is one of the areas for the development of concept agent technology in contemporary organizations.

Software agents, thanks to their autonomy, pro-activeness, personalization of activities and mobility can become an element that supports and perfects the functioning of IT systems in the knowledge based organizations.

There are a number of studies on applying software agents in this area. The context of their applications and functionality in the area of knowledge management has been presented in [18], [19], [20], [21],[22],[23].COMMA [19] is one the classic solutions in which software agents have been applied to support knowledge management. This solution is supposed to govern storing of the organizational knowledge and can be treated as an example of corporate memory concept. It utilizes web resources and improves sharing of knowledge in terms of searching for conceptual dependency which leads to improvements in communication between different groups of organization members.

Recent researches refer to application of agents in the area of supporting knowledge management and how they affect the processes in the organizations.

In [20], an IT solution that supports software management processes has been described. The authors presented the context of the application of software agent society and the mechanisms of its semantic representation. The study [21] presented elements of the concept of agent-mediated knowledge management system, which may find its use as a support for call center operations. Its use is associated with the construction of rule-based system for the selection of employees for specific customers based on their characteristics. But the work has not identified sources of information to determine the competence of the employees.

In the literature, we also find research projects regarding the use of software agents as a part of a searching mechanism. There are also other works showing how agents are used to integrate unstructured sources of information with codified knowledge resources of the organization on the semantic level. Indicated in [22] solution ACQUIRE shows the use of agents in the process of information retrieval by use of SQL. These studies, however, do not take into account the issue of knowledge representation in the organization and integration of this solution with knowledge management systems. Another solution [23] indicates the necessity of considering the software agents society as part of the mechanism that supports the integration of knowledge in the organization, showing the possible use of agent's system in the integration of information systems in the context of the use of semantic knowledge representation methods. In this case, the solution is directed to automating the processing of semantic knowledge based on the data in a relational database.

These solutions show that software agents can be used to provide semantic-savvy connection between information systems of the organization with the resources of its codified knowledge. In particular, in the context of a knowledge-based organizations, such arrangements may support functioning of knowledge portals, which are used in such organizations.

In the process of implementation and maintenance of integrated information systems knowledge about business processes plays a major role. Knowledge is a key resource for mapping these processes to the functionality offered by the implemented software package as well.

For each company, which daily operation is conditioned by the effective use of the knowledge and skills of its employees it is important to properly manage the organizational knowledge. There is no single generally acceptable concept of knowledge management. There are many definitions of the term, pointing to various aspects, such knowledge itself (for example, the approach of resource allocation, process, etc.) and to the process of knowledge management in organizations.

Research conducted by the authors showed that the main areas of application for IT solutions in knowledge-based organizations include:

- The need to develop an integrated knowledge management system to support selected business processes and management of knowledge about them.
- Allowing the sharing of knowledge, not only within the organization studied, but also among business partners.
- Enhancing the communication within project teams and between them, along with consulting their business partners.
- Automation of the process of knowledge management by providing innovative solutions in the form of software agents society, enabling in particular [12]:
    - the tasks related to the search, acquisition, structuring of, archiving and integration of knowledge in organizational knowledge repositories,
    - notify users about updates regarding organizational knowledge,
    - the negotiation and acquisition, also in the form of a purchase, the knowledge for users,
    - clarifying the issues related to the appropriateness and quality of knowledge.

The outcome of diagnosis of the requirements is the developed concept of the knowledge portal.

3 Concept of the Knowledge Portal

Designed social system uses a combination of wiki engine capabilities of social knowledge processing [13] and the concept of structuring knowledge by describing it with metadata, created with the use of ontology description languages. Metadata is used to improve the process of searching the knowledge base, and taking into account the relationships between different information, making up the knowledge of the organization.

Each wiki page, comprising a knowledge base, is an instance of codified knowledge. Specification of the concept of knowledge instances has been taken from [14]. To describe such instance we use the so called knowledge registration form. In case of this system, this form contains information about the source of knowledge, including a person who is the author of this portion of knowledge and about initiator – a person who is responsible for creation of this specific part of knowledge. In addition, the form contains data related to the location of expertise in the form of precise description of the knowledge whereabouts, ie. a system and a place within this system where the knowledge is applied. Thanks to this the knowledge, stored in the portal does not exist in isolation from the site of application, but is closely linked to it, making it easy to access the specific example of its usage and simplifies the process of assessing the suitability of the instance of knowledge for the persons concerned.

The typology of knowledge objects existing in the system is derived from the types of used information technology solutions and the types of objects present in a clearly identifiable form in the process of implementation and maintenance of an integrated management information system. For example there is a common issue with defining numbering series within the system – the knowledge about how to do it according to developed best practices becomes a knowledge object which type is best practice. Such type is connected to the metadata, specific for this type. Another example can be an additional solution, developed to satisfy a specific need of the organization where integrated system is implemented. Although this solution is very specific and potentially unrepeatable it shares some common characteristics that can be applied to all IT solution of this type. The aim is to facilitate the search through the knowledge base and provide for retention of knowledge in fragments as small as possible which in turn should make it easier for users and encourage them to generate content. Granulation results from the application of different perspectives to the accumulated knowledge [15]. This knowledge granularity allows the mapping of the actual connections between the various elements of the knowledge accumulated by the consulting company. In other words, according to this approach it is possible to obtain semantic, contextual links between elements / clusters of knowledge.

For the purposes of the presented problems the following areas of social software platform has been chosen, which are a representation of the areas / clusters of knowledge used by an implementation company:

- Projects – contains information about projects in progress and completed. Knowledge which is collected here is related to the organization of projects. For example, pre-analysis, the documents related to the project, meeting notes, etc.
- Solutions – atomized knowledge of IT solutions implemented by various consultants. For example, technical reports and the details of their implementation, additional solutions, validations, print layouts, etc.
- Customers – collected in the form of customer cards, gathers information on the companies, where projects have been conducted. Also, information on the current methods of support - the data describing connection details to customers' systems.
- Best practices – gathers knowledge on how to solve common problems encountered during implementation projects.

The platform can be used to provide easier start for junior consultants, who do not have adequate knowledge and experience. It becomes a kind of map that allows the new employees to get accustomed with ties existing between the various elements of organizational knowledge and to learn about what are the company's worked out ways of dealing with problems – best practices. Also it is very helpful when someone needs to find persons with specific skills.

For each area there are ontologies defined. As a result they are better suited to the knowledge that is contained in them. An example of a cluster of the knowledge is presented in Fig. 1.

```
(...)
<owl:ObjectProperty rdf:about="#hasProject">
<rdfs:range rdf:resource="#Project"/>
<owl:inverseOf rdf:resource="#hasClient"/>
<rdfs:domain rdf:resource="#Client"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="hasCreator">
<rdfs:range rdf:resource="#Employee"/>
<rdfs:domain rdf:resource="#Solution"/>
</owl:ObjectProperty>
<owl:FunctionalProperty rdf:ID="project_type">
<rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#st
ring"/>
<rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#Datat
ypeProperty"/>
<rdfs:domain rdf:resource="#Project"/>
</owl:FunctionalProperty>
(...)
```

Fig. 1. Part of the knowledge ontology on the portal

Platform makes it possible to create templates for editing and displaying a particular topic. The use of ontologies and templates is designed to increase the usability of the platform for an average user. One of the main problems with the use of knowledge management systems supported by social software is that its users do not want to spend time on generating content. In order to overcome these problems it was decided to facilitate the content creation by suggesting how it should be done.

4 Software Agents Society in Enhancing Knowledge Portal

Portions of knowledge in the portal are de facto interrelated through hyperlinks. Such a linkage can be considered appropriate in a situation where people who are looking for a solution to their problem do not need to know more about the context of the knowledge stored in a specific topic. This occurs most often when the system is used by an experienced consultant, whose purpose is merely to check the availability of ready-based IT solution. But the knowledge accumulated in the system can also be used in other ways. The system can provide a knowledge base for the less experienced users or for a user what just wants to learn more on a particular subject. For such person, it will be important to know how the portion of knowledge fits to a broader context, most often expressed in the form of a business process.

To take into account the need to satisfy the requirements of less experienced users and increase usability of the portal in terms of training it was decided to add a semantic layer to the existing social platform. This layer is supposed to mediate the interaction between the user, and the knowledge accumulated on the platform. In particular, it is intended to facilitate the mapping of the fragments of knowledge to the business processes. In addition this semantic layer is intended to facilitate browsing the knowledge stored on the platform by providing browsing personalization.

To accomplish these tasks we use software agents society. In the case of knowledge management software agents society can be considered as a component of a knowledge management system or as a part of the decision support system [16]. Software agents society can be considered as a decision-making support based on knowledge resources codified by the ontology description languages. At the current stage of research, knowledge portal architecture supported by software agents society is divided into five layers:

The layer of personalization and communication – is responsible for contacting with the user. Its goals are: finding an instance of knowledge, knowledge filtering according to the context of the user, generating tips, assisting the user in his or hers activities, the analysis of the work context of the user.
The layer of management of knowledge portal – prepared for the purpose of allowing users to codify and share of knowledge.
The layer of management of software agents society - its job is to manage the activity of software agents society. It is an intermediate layer between the layer of personalization and the layer of knowledge sharing. One of the components of this layer – knowledge brokers agents and personalization agents – is presented below.
The layer of knowledge sharing - its task is to provide codified knowledge stored in the knowledge portal for the software agents society activities.
The layer of knowledge storing – it is responsible for storing codified knowledge regarding the context of use of the software agents society and knowledge portal itself.

Personalization agents are created when a user logs on the platform. They create or update data describing the user by a series of characteristics:

- frequency of visits,
- the total amount of time spent in a single work session,
- amount of time, divided by category – webs (in terms of knowledge portal – for instance: projects or solutions),
- number of topics created and/or edited,
- place in the organizational structure - function performed in the organization.

Brokers agents are responsible for developing recommendations for the user based on information obtained from personal agents. They facilitate the user finding linkages between knowledge portions, making up the business process according to the user profile. For example if user profile contains information about his or hers engagement in creation of IT solutions broker agent can present a link between currently watched solution and solution dependent on this one. They also help in finding people with a specific expertise. One of the concepts they use is the analysis of the quality of the knowledge stored in the portal.

5 The Idea of Application of Software Agents Society for Determining the Quality of Knowledge in the Portal

As [17] indicates one of the steps in the lifecycle of knowledge management system is knowledge audit, which „… implies analyzing its [audit] impact on decisional processes, the contribution of each piece of knowledge in optimizing business processes, therefore in raising the quality of the products and services offered. At the end of this process, it is possible to have to regenerate some knowledge pieces".

The evaluation of the quality of knowledge can contribute to the improvement of business processes. For instance application of a social software, which enables social ranking, can help in evaluating an employees' knowledge and can provide the basis for determining their competences.

For the purpose of the designed system we identified three parameters used in the knowledge evaluation (as indicated by Hassenzahl and Tracinski [24]): its usefulness, completeness and readability. The complexity of the definition of usability [25] and its possible multi-faceted examining was deliberately limited in this study. For the evaluation of user-generated content one can use various methods. According to a study prepared by Chai et al. methods used for the evaluation for wikis and blogs often concentrate on gathering direct response from the readers of knowledge portion being evaluated[26].

Earlier shown part of the knowledge defined in the semantic layer of knowledge portal refers to business processes executed by employees. Based on information provided by the user agent, a broker agents analyzes codified knowledge shared by the users. The process of agent communication designed in JADE was presented in Fig. 2.

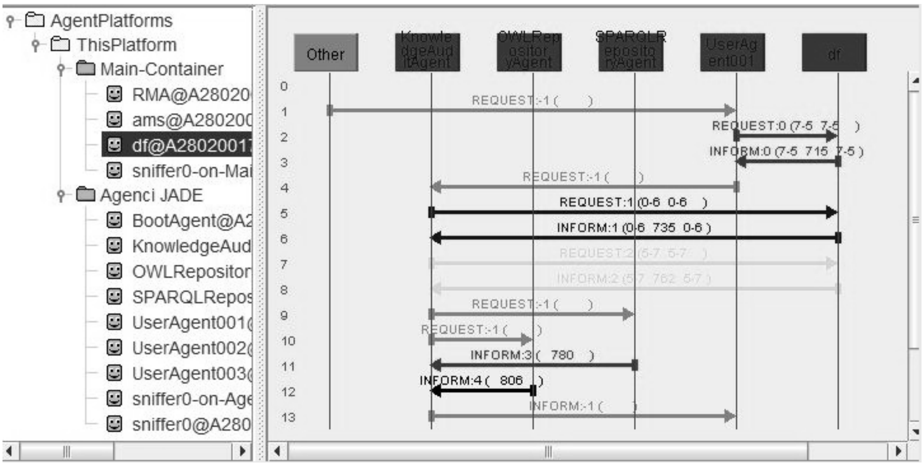


Fig. 2. Agent communication process in JADE

This piece of knowledge refers to problems encountered in the implementation of business processes and methods for their solutions. On this basis, for each employee a ranking is calculated for a specific area of expertise. In presented case, 43 codified solutions created by 6 employees, were used for agent evaluation. Sample of portal knowledge saved in OWL ontology is shown in Fig. 3.

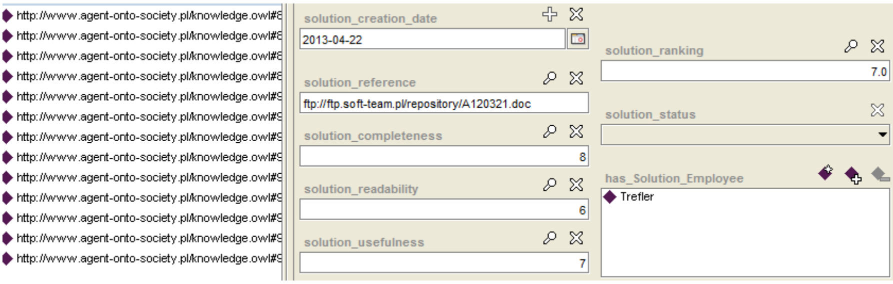


Fig. 3. A sample of agent knowledge about possible problem solutions

The ranking algorithm used in this solution for assessment of employees on the basis of their competence is presented below:

1. Find knowledge objects that refer to the solutions for problems that arise during the implementation of business processes.
2. Find knowledge objects that refer to employees who are the authors of the solution.
3. Find knowledge objects concerning the projects for which the solution has been developed.
4. Specify the number of the solutions generated by each employee.
5. Find ratings for the knowledge object regarding the developed solutions.
6. Generate ranking for each employee on the basis of rankings of the defined elements of knowledge.

An example of code presenting this algorithm is shown in Fig. 4. In order to limit the number of results the summary presented below contains only the average ranking of selected parameters for the five employees involved in the project with the highest ranking for the selected month.

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX know: <http://www.agent-onto-society.pl/knowledge.owl#>
PREFIX xsd:  <http://www.w3.org/2001/XMLSchema#>

SELECT ?project ?lname ?fname (Count(*) AS ?ilosc) (AVG(?rank) AS      ?sredni_ranking)
WHERE {
  ?solution rdf:type ?type.
  ?type rdfs:subClassOf* know:Solution.
  ?solution know:solution_ranking ?rank.
  ?solution know:has_Solution_Employee ?emp.
  ?emp know:employee_lname ?lname.
  ?emp know:employee_fname ?fname.
  ?solution know:inverse_of_hasSolution ?project.
  ?project know:project_name "Obieg Dokumentow".
  ?solution know:solution_creation_date ?date.
  FILTER(xsd:dateTime(?date) >= "2013-04-01T00:00:00Z"^^xsd:dateTime &&
       xsd:dateTime(?date) <= "2013-04-30T23:59:59Z"^^xsd:dateTime)

} GROUP BY ?project ?lname ?fname
ORDER BY DESC(?sredni_ranking)
LIMIT 5
```

Fig. 4. An example of code that generates the ranking of user-generated content

```
---------------------------------------------------------------------------------------
| lname                    | fname                    | ilosc | sredni_ranking          |
=======================================================================================
| "Trefler"^^xsd:string    | "Mateusz"^^xsd:string    | 5     | "8.0"^^xsd:float        |
| "Kowal"^^xsd:string      | "Radoslaw"^^xsd:string   | 14    | "7.214286"^^xsd:float   |
| "Nowak"^^xsd:string      | "Piotr"^^xsd:string      | 8     | "7.0"^^xsd:float        |
| "Sawicz"^^xsd:string     | "Agnieszka"^^xsd:string  | 6     | "6.8333335"^^xsd:float  |
| "Kolodziej"^^xsd:string  | "Daniel"^^xsd:string     | 6     | "6.1666665"^^xsd:float  |
---------------------------------------------------------------------------------------
```

Fig. 5. An example of the ranking

Fig. 5 shows an example of the ranking that is a result of the analysis of codified knowledge.

Thanks to the usage of software agents in the analysis of the codified knowledge it becomes possible to better assign employees to tasks within business processes running in the organization. The designed system is able to generate suggestions regarding the selection of employees, indicate which employees have sufficient competence in the area, thus increasing the level of satisfaction the beneficiaries of the business process. At the same time integrating this solution with a portal of organizational knowledge contributes to the promotion of knowledge about the business processes and supports the exchange of information between employees, increasing their knowledge of these processes and their stages of their execution.

Summary

The proposed concept of using software agents society to support the knowledge based organization refers to the observed need to find a solution which can address the problem of knowledge representation in the manner that makes it possible to interpret that knowledge by users and software agents, and, at the same time, makes it possible to support the users in execution of business processes. The studies show that agent-based solutions should be a part of the organizational knowledge management systems. The result of their work is the support for the knowledge processing in the knowledge portals. Software agents also expand the functionality of such portals. One of the main aspects here is the analysis of explicit knowledge stored in the portal and determination of its usefulness in the processes of the organization.

What distinguishes the idea presented in this paper is integration of software agents society as a part of the organization's knowledge portal and application of the mechanisms of semantic knowledge interpretation and processing. The basic problem of the construction of such solutions is the heterogeneity of the sources of knowledge that can be processed by the organization. Use of knowledge portal allows preparing one common ontology for the organizational knowledge base, which can be used by software agents in the process of knowledge evaluation.

Current research on the development of portal's functions are related to the concept of knowledge representation for building the maps of business processes in organizations. We also expect that using software agents society can support a simulation of business processes and improvement of their implementation. These elements will be the subject of further studies by the authors.

Acknowledgements. The issues presented constitute a preliminary stage of the authors' research into the aspect of modeling software agent societies in knowledge-based organizations. The project was financed from the funds of National Science Centre 2011/03/D/HS4/00782.

References

1. Unold, J.: Web 2.0 w kreowaniu wiedzy zbiorowej. In: Filipczyk, B. (ed.) Wiedza i komunikacja w innowacyjnych organizacjach. Systemy ekspertowe wczoraj, dziś i jutro, pp. 104–111. Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice (2010), ST – Web 2.0 w kreowaniu wiedzy zbiorowej
2. Jackson, S.E., Erhardt, N.L.: Building Social Connections to Gain the Knowledge Advantage. In: Goldsmith, M., Morgan, H., Ogg, A.J. (eds.) Leading Organizational Learning, pp. 255–266. Jossey-Boss, San Francisco
3. Swan, J., Newell, S., Scarbrough, H., Hislop, D.: Knowledge management and innovation: networks and networking. Journal of Knowledge Management 3, 262–275 (1999); ST – Knowledge management and innovation
4. Schneckenberg, D.: Web 2.0 and the empowerment of the knowledge worker. Journal of Knowledge Management 13(6), 509–520 (2009)
5. Pfaff, C.C., Hasan, H.: Overcoming Organizational Resistance to Using Wiki Technology. In: Proceedings of the 10th Pacific Asian Conference on Information Systems
6. Bajenaru, A.: Using wiki as a corporate knowledge sharing system. Annals of University of Craiova - Economic Sciences Series 2 (2010)
7. Jackson, P.: Capturing, structuring and maintaining knowledge: a social software approach. Industrial Management & Data Systems 110, 908–929 (2010), ST – Capturing, structuring and maintaini
8. Gunawardena, C.N., Hermans, M.B., Sanchez, D., Richmond, C., Bohley, M., Tuttle, R.: A theoretical framework for building online communities of practice with social networking tools. Educational Media International 46(1), 3–6 (2009)
9. van Elst, L., Dignum, V., Abecker, A.: Towards agent-mediated knowledge management. In: van Elst, L., Dignum, V., Abecker, A. (eds.) AMKM 2003. LNCS (LNAI), vol. 2926, pp. 1–30. Springer, Heidelberg (2004)
10. Schonstrom, M.: Creating knowledge networks: lessons from practice. Journal of Knowledge Management 9(6), 17–29 (2005)
11. Vaast, E.: The Use of Intranets: The Missing Link Between Communities of Practice and Networks of Practice. In: Hildreth, C., Kimble, P.M. (eds.) Knowledge Networks: Innovation Through Communities of Practice2, pp. 202–216 (2004)
12. Dignum, V.: Agent Mediated Knowledge Management. In: IAT 2007 Tutorials (2007)
13. Wagner, C., Bolloju, N.: Supporting Knowledge Management with Conversational Technologies: Discussion Forum, Weblogs and Wikis. Journal of Database Management 16(2), i–vii (2005)
14. Calmet, J., Maret, P., Hammond, M.: Agent Societies for Corporate Knowledge Issues. In: The Fifth International Workshop on Engineering Societies in the Agents World (2004)
15. Kay, J., Barg, M., Fekete, A., Greening, T., Hollands, O., Kingston, J.H., Crawford, K.: Problem-based learning for foundation computer science courses. Computer Science Education 10(2), 109–128 (2000)
16. Calmet, J., Maret, P., Endsuleit, R.: Agent oriented abstraction. Royal Academy of Sciences Journal (2004)

17. Muntean, M.: IT Infrastructure for Collaborative Environments. Journal of Applied Collaborative Systems, 5–10 (2009)
18. Di Nitto, E., Pianciamore, M., Selvini, P.: The role of agents in knowledge management. In: WOA, pp. 29–34 (2002)
19. Gandon, F., Berthelot, L., Dieng-Kuntz, R.: A multi-agent platform for a corporate semantic web. In: Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 3, pp. 1025–1032 (2002)
20. Nor, Z.M., Abdullah, M., Selamat, R., Azrifah Azmi Murad, M., Hasan, M.: Managing Knowledge in Collaborative Software Maintenance Environment. In: Virtanen, P., Helander, N. (eds.) Knowledge Management, pp. 73–93. In-Teh (2010)
21. Popirlan, C.I.: A Multi-Agent Approach for Distributed Knowledge Processing in Contact Centers BT. In: Proceedings of The 14th WSEAS International Conference on COMPUTERS (Part of the 14th WSEAS CSCC Multiconference), July 23-25. Latest Trends on Computers, pp. 214–219 (2010)
22. Das, S., Shuster, K., Wu, C.: ACQUIRE: agent-based complex query and information retrieval engine. In: Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 2, pp. 631–638 (2002)
23. Monticolo, B., Lahoud, I., Bonjour, E., Demoly, F.: SemKnow: A Multi-Agent Platform to Manage Distributed Knowledge by using Ontologies. In: International Conference on Artificial Intelligence and Applications, Hong Kong (2012)
24. D. 9241-210:2008, Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems (2008)
25. Hassenzahl, M., Tractinsky, N.: User Experience - a research agenda [Editorial]. Behavior & Information Technology 25(2), 91–97 (2006)
26. Chai, K., Potdar, V., Dillon, T.: Content Quality Assessment Related Frameworks for Social Media. In: Gervasi, O., Taniar, D., Murgante, B., Laganà, A., Mun, Y., Gavrilova, M.L. (eds.) ICCSA 2009, Part II. LNCS, vol. 5593, pp. 791–805. Springer, Heidelberg (2009)

Research of Conventional Data Mining Tools for Big Data Handling in Finance Institutions

Darius Tamasauskas, Marius Liutvinavicius, Virgilijus Sakalauskas, and Dalia Kriksciuniene

Vilnius University, Lithuania
{darius.tamasauskas.85,marius.liutvinavicius}@gmail.com,
{virgilijus.sakalauskas,dalia.kriksciuniene}@khf.vu.lt

Abstract. The article investigates the usability of conventional data mining tools for handling data sets generated in financial institutions. It discloses the characteristics of "big data" which reveal limitations and new requirements for analytical software to deal with huge data flows related to financial transactions. The performance characteristics of four different conventional data mining tools, their visualization and clustering models are tested for experimental set of big data extracted from bank local data warehouse. The ranking of critical characteristics is provided for each stage of analysis of big data set.

Keywords: big data, data mining tools, banking data, risk assessment.

1 Introduction

The problems in financial sector that emanated after 2008 financial crisis showed the need to improve risk management. As the financial sector is distinguished by its large amounts of data generated by banking, insurance and other financial institutions, it is difficult to evaluate all the information adequately and discover potential threats. Data mining techniques and improved models of risk management could help to solve this problem.

Nowadays banks and other financial institutions use various business intelligence systems for analysis of their data repositories. Business intelligence is a data-driven decision support system that combines data gathering, data storage, and knowledge management with analysis to provide input to the business decision process (Burstein and Holsapple, [2]). It is quite common for banks to combine several tools for data analysis and to make considerable investments for data handling. However, they use conventional multi-purpose software such as Excel, Access, Statistica, SPSS or open source tools as well, for carrying out various managerial analytical tasks. Conventional software tools used in business have the required flexibility for processing and interpretation of analytical tasks. They are widely applied not only in large financial institutions such as banks- this type of tools is mostly widespread in small financial institutions, providing financial services, consulting, insurance.

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 35–46, 2013.

McAfee [9] revealed the important relationship: the more companies characterized themselves as data-driven, the better they performed according to measures of financial and operational results. One of the key findings of Lavalle's [8] assured that top-performing organizations use analytics five times more than lower performers.

According to IBM [5] 2.5 quintillion bytes of data is created every day — as many as 90% of data in the world today has been created in the last two years alone. The problem related to various aspects of handling contemporary data sources is often called "big data" problem.

Gantz and Reinse in [3] define the term "digital universe" as a measure of all the digital data created, replicated, and consumed in a single year. They predict that during 2005 to 2020 the digital universe will grow from 130 exabytes to 40,000 exabytes, or 40 trillion gigabytes.

Big data usually is described as high volume, big velocity and big variety information assets that demand cost effective, innovative forms of information processing for enhanced insight and decision-making (Gartner IT Glossary, 2013, [4]). TechAmerica Foundation's Federal Big Data Commission [12] also determines big data as a phenomenon defined by the rapid acceleration in the expanding volume of high velocity, complex, and diverse types of data.

The data in financial institutions meet the criteria for big data: massive (e.g., terabytes in volume), temporally ordered, fast changing, potentially infinite. Kovalerchuk and Vitaev [7] state that financial institutions produce huge datasets that build a foundation for approaching these enormously complex and dynamic problems with data mining tools in order to extract knowledge from large amounts of data.

It is notified that successfully harnessing big data can help banks achieve three critical objectives for banking transformation: create a customer-focused enterprise, optimize enterprise risk management and increase flexibility and streamline operations IBM (2013). Understanding and managing financial risk, credit rating, loan management, bank customer profiling and money laundering analyses are core financial tasks for data mining.

Data mining can contribute for solving business problems in banking and finance by finding patterns, causalities, and correlations in business information and market prices that are not immediately apparent to managers because the data volume is too large or is generated too quickly to screen by experts (Rajanish Dass, [10]).

In this article we use conventional software packages in order to determine their suitability and limits for various aspects of big data analysis. Although we are aware of special large scale systems which can deal with tremendously huge sets of data, but we focus on situations where there are no possibilities to use powerful servers, specific databases, cloud systems or expensive data mining tools. We perform analysis with our big data example by using conventional hardware and tools that are open source or common in business. The experimental data consists of monthly customer transaction data from medium size bank, extracted from several sources included to local data warehouse.

The article is structured in the following way: data mining software is overviewed in the section 2, research framework and its performance with experimental data is presented in section 3. The last section compares the results of analysis made with selected software.

2 Data Mining Software Solutions

According to Baru et al [1] big data analytics has become a major force of innovation across enterprises of all sizes, but the vast selection of tools brings problems of their comparative evaluation and fit for enterprise purposes. In general, data mining methods can be classified into the following categories: association rule mining, classification, clustering analysis, sequential pattern finding by using four main data mining techniques: neural networks(NN), genetic algorithms(GA), statistical inference(SI), data visualization(DV) (Shaikh and Chhajed, [11]).

We selected the tools that have main features of conventional analytical software, as they are either free or don't require large resources, potentially can handle big data and have basic set of functions (Table1). This should allow the companies that don't have resources for costly solutions to work with big data sets.

Table 1. Comparison of data mining tools

Tool	Free	Open Source	NN	GA	SI	DV	Features of data mining tools
R	Yes	Yes	+	+	+	+	Decision trees, clustering, association rules, text mining, tools for creating chart and graphs.
Datameer	No	No	-	-	-	+	Data visualization.
SAS	No	No	+	+	+	+	Clustering, time series analysis, text mining, forecasting trends.
Google Chart	Yes	No	-	-	-	+	Data visualization, web-based tool.
Orange	Yes	Yes	+	-	-	+	Data visualization, fast and versatile visual programming.
Rapid Miner	Yes	Yes	+	+	+	+	Text mining, predictive analytics, clustering, for research and real-world data mining tasks.
Weka	Yes	Yes	+	+	+	+	Regression, clustering, association rules, visualization, classification.
JHep Work	Yes	Yes	-	-	-	+	Data visualization, designed for scientists, engineers and students.
Knime	Yes	Yes	+	+	+	+	Data transformation, predictive analytics, visualization.
Rapid Analytics	Yes	Yes	+	+	+	+	Text mining, predictive analytics, clustering.
PSPP	Yes	Yes	-	-	+	+	Descriptive statistics, linear regression, T-tests, non-parametric tests, graphical user interface.
Rattle	Yes	Yes	-	-	-	+	Data transformation, data visualization.
SPSS	No	No	+	+	+	+	Predictive analytics.
Pentaho	No	No	-	-	-	+	Interactive dashboards, data visualization.
XLminer	No	No	+	-	+	+	Neural nets, classification and regression trees, logistic regression, linear regression, association rules, clustering, principal components.
Viscovery Somine	Yes	Yes	-	-	+	+	Visual cluster analysis, self-organizing maps.

For further research we selected the four with the declared capacity of big data handling, namely Rapid Miner, Viscovery Somine, XLMiner and SPSS. Rapid Miner was selected because it's open source and has a wide range of intelligent analysis features. Viscovery Somine is also easily accessible tool designed for a particular data mining task. XLMiner was selected as one of the simplest ways to become ready for data mining. Finally SPSS was chosen in order to test how popular statistics tool can handle with big data.

RapidMiner, formerly called YALE (Yet Another Learning Environment), is an software environment for machine learning and data mining experiments oriented for both research and real-world. It enables experiments consisting of a huge number of arbitrarily unstable operators, which are made with the graphical user interface of RapidMiner[1] and detailed in XML files. RapidMiner is freely available open-source data mining and analysis system that runs on every major platform and operating system. It has GUI mode, server mode (command line), or access via Java API. RapidMiner contains more than 500 operators for data integration and transformation, data mining, evaluation, and visualization and also has extension mechanism. Machine learning library WEKA is fully integrated in this software. RapidMiner uses graphical process design for standard tasks and scripting language for arbitrary operations. It has standardized XML interchange format for processes. Different data sources can be accessed: Excel, Access, Oracle, IBM DB2, Microsoft SQL, Sybase, Ingres, MySQL, Postgres, SPSS, dBase, Text files[2].

RapidMiner has got references as the world-leading open-source system for data mining Koutonin (2013). It is claimed that thousands of applications of RapidMiner in more than 40 countries give their users a competitive edge[3].

Viscovery SOMine is a system which combines methods of classical statistics and self-organizing maps (SOM) for explorative data mining and predictive modeling. The SOM technology is used for representing and visualizing data distributions that may contain large number of variables data sets[4]. The visual interface allows direct, context-sensitive access to the original data records underlying the map formations and their complete statistics at any point in the analytical workflow. The system enables exploratory data mining, statistical analyses, profiling and segmentation, visual cluster analyses, various approaches for classifications[5]. The system provides a number of classical statistical functions including descriptive statistics, group profiles, correlation analysis, principal component analysis, histograms, or scatterplots.

XLMiner is comprehensive data mining add-in for Excel, including numerous models, such as neural nets, classification and regression trees, logistic regression, linear regression, Bayes classifier, K-nearest neighbors, discriminant analysis, association rules, clustering, or principal components[6].The authors of XLMiner claim

---

[1] http://www.junauza.com/2010/11/free-data-mining-software.html

[2] http://rapid-i.com/content/view/181/190/

[3] http://www.siliconafrica.com/the-best-data-minning-tools-you-can-use-for-free-in-your-company/

[4] http://www.viscovery.net/somine/

[5] http://www.the-data-mine.com/Software/ViscoverySOMine

[6] http://www.solver.com/xlminer-data-mining

that is possible to sample data from virtually any database and clean it with a comprehensive set of data handling utilities including categorizing data and handling missing values. The visualization include wide selection of tools starting from simple bar, line and histogram charts to advanced multiple linked charts, one-click changes to axes, colors and panels, zooming and brushing.

IBM SPSS Statistics is the specialized statistical software for performing the data analysis process. It includes tools for loading data from various sources and preparing it for work by enabling necessary transformations, such as handling missing values, sorting variables). Different statistical methods are used for full data analysis and obtaining meaningful results. The results can be visualized in graphs and tables and the output data can be exported to various formats. SPSS is widely used environment, adopted by more than 300,000 companies over the world. Among the customers there are 95 percent organizations from the Fortune 1000 list of the most profitable businesses[7]. The encapsulated mathematical and statistical expertise enables to extract predictive knowledge that when deployed into existing processes makes them adaptive to improve outcomes[8].

3 Research Framework and Experimental Data Analysis

For exploring performance of conventional software for handling big data, we analyzed customer transaction information of one of the middle size banks in Lithuania. Data contains income and balance information of bank customers collected on monthly basis from 2011 January till 2013 March.

We further overview characteristics of experimental data if it can be classified as big data for exploring its processing by the selected analytical tools.

Volume: the experimental data represents 27 periods, denoting months, data set consists of about 270 000 cases. It can be considered as big data as it is massive, temporally ordered, fast-changing and potentially infinite.

Velocity is not involved for our experiment as the analysis aims to provide long term insights of customer behavior. However this characteristic was very important for evaluation of the performance of selected data mining tools.

Variety factor is very important, as data is extracted from several sources using SQL scripts. Data extraction is very fast, but there is no tool for getting it. User has to be familiar with SQL and write queries himself.

Data is collected from several sources. The main source is bank's local data warehouse. Data in local data warehouse is not enough for calculating proper bank result. There are various adjustments, which are required for management decisions. Adjustments include extracting data for Liquidity premium, Economic Capital, Income on Economic Capital, Expected losses, data from subsidiaries and various others. For this reason the other data sources, such as data collected in SAP and financial data warehouse were used for data extraction. In SAP we can see legal result and management result with all monthly adjustments. The financial data warehouse is

[7] http://www.insol.lt/statistics/

[8] http://www.spss.com.hk

a new emerging data source which is currently being developing at explored bank. It is implemented in order to collect all financial data for all departments of various countries in the same place. The financial data warehouse is used to generate automated reports, which can provide data in different aggregation levels. Reports are done to calculate various results, profit-loss information, balance sheet information, loan losses, interest income, lending and deposit portfolios and their margins, funding gap, number of various clients, number of housing loans, not performing loans, income statements by customer segments, expenses and others.

Variability characteristics has important role, as the data extracted from bank's local data warehouse and other sources was arranged in 13 variables. Though the number of variables is not big, there were cases of missing data, some of the categorical values changed very often and brought inconsistency in analysis, and therefore it makes data analysis even more complex. The definitions of the attributes of data are presented in Table 2.

Table 2. Description of variables

DataDate	**Monthly date**
CustomerID	Customer's identification code
Line	Type of income
Sector	Field, characterizing customer's segmentation on the type of work area
Target	Field, characterizing customer's segmentation on his income and assets
PLcategory	Detailed breakdown of the line / group of different accounts
ProdCateg	Product code
DAO	Department account officer
ShortName	Type of customer
Desk	Detailed breakdown of *Shortname*
Income	Amount of income
AvgBal	Average monthly balance of customer's portfolio
CloseBalance	End of month balance of customer's portfolio

The variables in Table 2 serve for defining categories of customer characteristics. ‘Line’ defines the type of income. Basically there are 4 types of income: interest income, commission income, trading income and other income. Field ‘Line’ describes these types of income with more detailed split.

‘PLcategory’ is even more detailed split of ‘Line’ and basically it is group of different accounts.

‘ProdCateg’ defines the product, for example different kinds of accounts, loans, over drafts and so on. ‘Sector’ and ‘Target’ describes customer's segmentation. ‘Sector’ is more of customer's working area, for example: Large corporate customer, Small corporate customer, Credit institutions, Banks and so on. While ‘Target’ is more about customer's segmentation on his income and assets, for example: Gold customers, Silver customers, VIP customers, Potential customer, International customers and so on. ‘DAO’ describes internal account officer codes and departments.

‘Shortname’ is for describing customer's type. The largest groups are Corporate customer and Household customers. ‘Desk’ has information about Corporate smaller units, Household physical branches and other desks.

Field ‘Income’ is the main amount of income. If the ‘Line’ stands for interest income type, ‘Income’ will be lending or deposit income, get from customer. If the

'Line' has the value income type of commission, 'Income' will be commission paid by customer. 'AvgBal' and 'CloseBalance' are customer's balance information: cash in his account, remaining of loan, overdraft balance and so on. 'CloseBalance' has information about end of period balance, while 'AvgBal' is average month balance.

Veracity characteristic was involved partly. As the banking transactions are aggregated from trustful sources, the trust aspect was not present. This characteristic is interconnected to the variability due to missing data. As the tools deal with missing data cases in different ways, offering various replacements, the outcome of analysis contains has stochastic approach.

Volume characteristic was crucial during experimental analysis. Due to data volume the analysis in most cases was reaching or overcoming performance limits of the selected tools. Therefore any refinements in analysis done by adding more variables or selecting more complex functions led to reduction of processing speed, instability and even automatic skipping of part of the data cases.

4 The Results

In this section the results of performed analysis with selected software is presented.

Baru et al [1] suggested four step procedure for software benchmark: system setup, data generation, data load, execution of application workload. We have several core tasks: to verify software's capabilities to load big data, as related to required support of available data formats and the volume), to perform cluster analysis and data visualization, assess the speed of those processes, evaluate the ease of use of the software and the completeness of the results (Fig. 1).

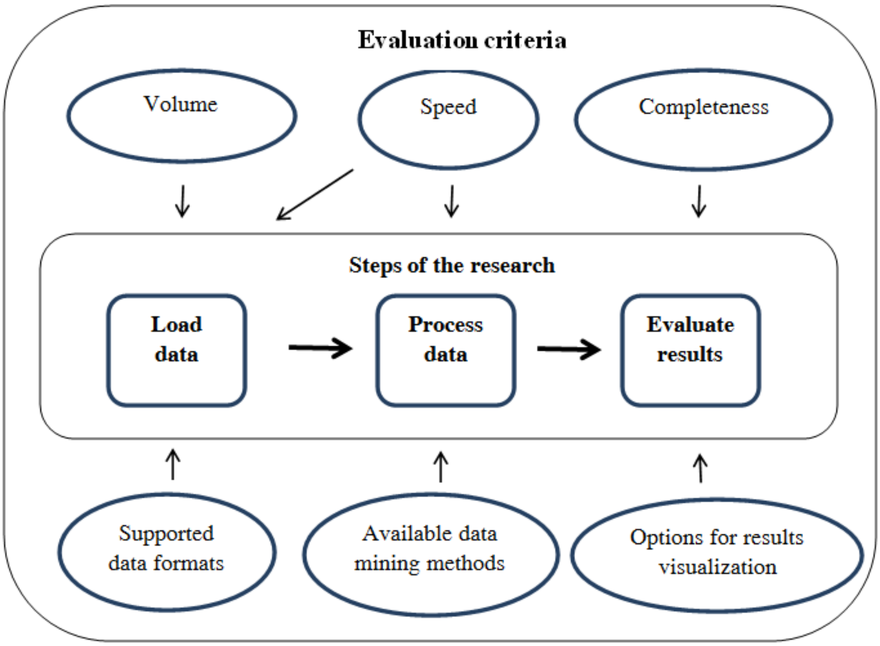


Fig. 1. The workflow of the research

For these tasks we used conventional computer with Intel Core i3 processor, 4 GB of RAM and HDD with 7200 rpm.

The first program used for data analyses was **Rapid Miner**. Despite the fact that it can work with Excel spreadsheets, in practice it was impossible to import such big data into data library from Excel file. Rapid Miner works only with .xls (Excel 2003) format spreadsheets, that cannot exceed 65 000 rows. But even with smaller amount of data the data import process took more than 48 hours and was canceled.

In order to solve this problem, all data was converted to .mdb (Access 2003) format. Indexed data was loaded to data library within 10 minutes.

The first task was to explore the data by using visualizing options. Rapid Miner presented all customers according to their Income and Line in several seconds. Visualization results in Rapid Miner are presented in Fig.2.

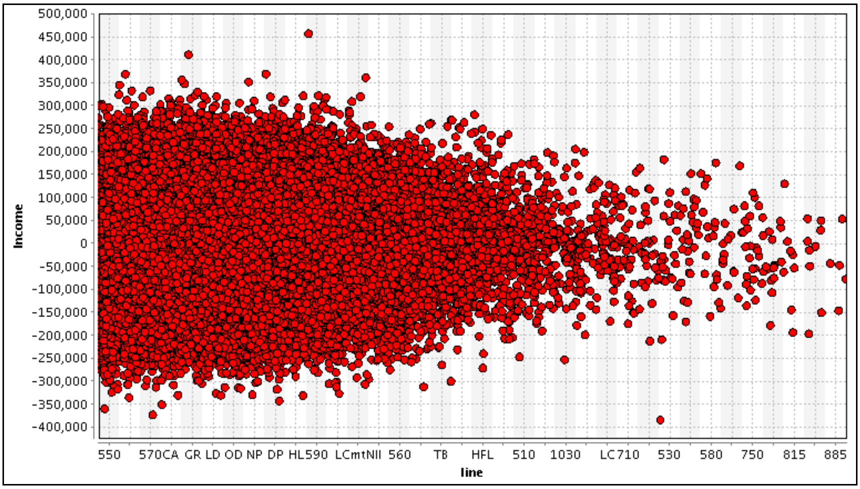


Fig. 2. Line (Type of Income) scatter plot by Rapid Miner

The Fig.2 shows that majority of income is generated in the first half of Lines. The second half of Lines generate much less income. This information enables to divide all data into smaller groups and perform deeper analysis of differencies between groups of customers.

If simple vizualization task were made without problems, the same can‘t be said about more difficult tasks. For example, Pareto ranking and SOM vizualization processes continued for more than 10 hours and quit without any results.

The Fig.3 shows the results of the main task – clusterization. It took less than 1 minute for Rapid Miner to put 270 000 cases into 8 clusters. But after this we could only find out the members of each cluster. In order to determine the characteristics of each cluster the model in Rapid Miner‘s environment should be improved with more operators.

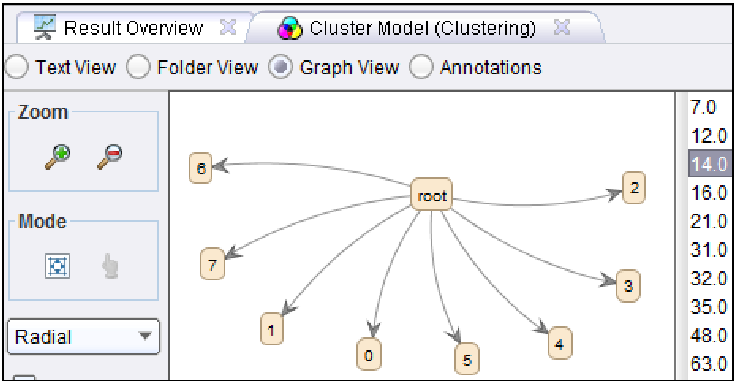


Fig. 3. Clusters in Rapid Miner

The second tested program was **Viscovery Somine.** This program encountered the same problem while trying to load all the 270 000 cases. Only .xls format was supported, therefore only 65000 cases could be analyzed.

The Fig.4 presents all steps required to generate SOM map. Firstly, the data is imported and data mart is created. After these steps, the clusterization model is created. The import of data took 11 seconds, and data mart recording took less than 3 seconds. During "Create model" stage the map was created in 2 minutes and 5 seconds. So with smaller amount of data the load process is quite fast, but with huge datasets it become unsatisfactorily delayed.

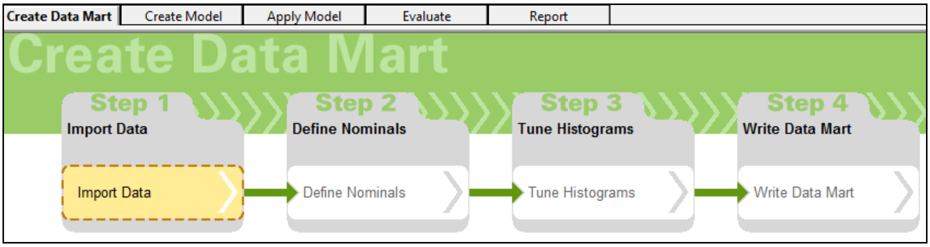


Fig. 4. Clustering in Viscovery Somine

All the data was classified into 8 clusters. Versatile and detailed data visualization is the biggest advantage of Viscovery Somine. The Rapid Miner only could classify data into clusters, but Viscovery Somine immediately presented the influence of different input variables for each cluster. For example, one cluster contains customers that have products only belonging to the same category. It also highlighted small group of customers who generated far greater income. With this software is was possible to discover characteristics of each cluster or automaticcaly determine the cluster where specific customer belonged.

The third tested solution was **XLMiner**. This is Excel add-in, that means this program can be easily integrated in the most popular spreadsheet tool.

The first task made with XLMiner was K-Means clustering. Despite the fact that the spreadsheet with 270 000 cases was loaded, XLMiner could handle only 20000 records. All the uploaded records were classified into 8 clusters in 75 seconds (Fig.5).

Data summary

Cluster	#Obs	Average distance in cluster
Cluster-1	9511	66,094
Cluster-2	3481	143,133
Cluster-3	1846	91,882
Cluster-4	7	25988,035
Cluster-5	916	547,251
Cluster-6	3463	232,069
Cluster-7	771	1347,46
Cluster-8	5	14699,088
Overall	20000	194,79

Elapsed Time

Overall (secs)	75,00

Fig. 5. Clustering results in XLMiner

The second task was to perform the hierarchical clustering. For this case only 6000 records were clustered. In addition to statistical information XLMiner presented the dendrogram (Fig.6). It allowed to determine the optimum number of clusters, as it shows the attribute distances between different merged classes.

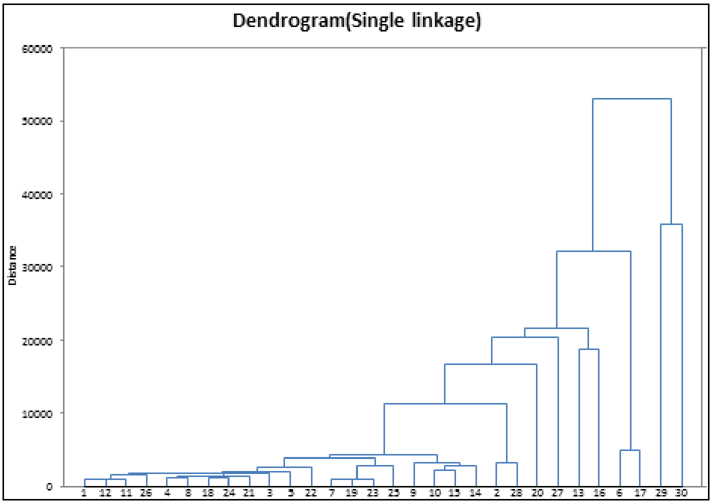


Fig. 6. Dendogram in XLMiner

The last tested software was popular statistical tool **SPSS**. According to the specification it should be possible to load .xlsx spreadsheets in SPSS v17, but the majority of 270 000 experimental records uploaded from Excel spreadsheet were corrupted and it was impossible to perform analysis. Only the import from Access database was succesful and it took less than 3 seconds.

Three continuous variables (Income, Average balance, Close balance) and four categorical variables (Sector, Target, Plcategory, Prodcateg, Shortcame) were selected for clustering.

Cluster Distribution

		N	% of Combined	% of Total
Cluster	1	5270	18,8%	2,0%
	2	5627	20,1%	2,1%
	3	931	3,3%	,3%
	4	2987	10,7%	1,1%
	5	5689	20,3%	2,1%
	6	3414	12,2%	1,3%
	7	1349	4,8%	,5%
	8	2698	9,6%	1,0%
	Combined	27965	100,0%	10,5%
Excluded Cases		239108		89,5%
Total		267073		100,0%

Fig. 7. Clustering results in SPSS

All records were grouped into 8 clusters in 1 minute and 50 seconds (Fig.8). But clustering results weren't satisfactory. The majority of data samples (89,5 %) were not included into analyses (Fig.7).

Table 3. The summary of experimental results

	RapidMiner	ViscoverySOM	XLMiner	SPSS
Loaded records	270000	65000	270000	270000
Data loaded without format changing	No	Yes	No	Yes
Time to load data	~10 min.	~3 s	~10 s	~3 s
Analyzed records	270000	65000	20000	27965
Time to analyze data	~1 min.	~2 min.	~1,5 min.	~2 min.
Statistics	No	Yes	Yes	Yes
Visualization	Partial	Detailed	Partial	Detailed

In table 3 the performance of the software tools is summarized. The differences among the four data mining environments are quite evident and show high risk for their application in real data mining problem solving. This is the reason why we need to look for more advanced software and intelligent data mining mothods to deal with such kind of big data for extracting meaningful information for decision making processes.

5 Conclusions

There is a great need for data mining tools in financial institutions that could perform deep analysis with data sets, which have critical characteristics of "big data" volume, velocity, veracity, variety, variability and value. The software should be able to load big data sets extracted and combined from different types of data sources and perform fast analysis despite cases of missing, changing and incomplete data.

Conventional tools used in business have required flexibility for processing and interpretation of analytical tasks; they are widely applied not only in large financial institutions such as banks, but in small businesses providing financial services, consulting, insurance. The amount of financial data analyzed in both small and big financial institutions is huge; and it is demanding same high quality of analysis as well. However, data mining tools of conventional use can't handle huge datasets due to different aspects of big data.

The experimental research with the data set of medium size bank revealed major problems and discrepancies while applying widely implemented and highly ranked software tools: Rapid Miner, Viscovery Somine, XLMiner and SPSS, selected due to their ability of handling big data sets. There were crucial problems with data preparation for analysis related to the big volume. Only DataMiner was able to analyze all 270 000 records correctly. The functionality of Viscovery Somine or SPSS were more appropriate for data mining tasks, however they couldn't deal with bigger databases. The speed factor is critical for data analysis. The research showed that personal computer with technical parameters commonly used by majority of managers for their analytical purposes were too slow for more advanced tasks, especially related with visualization. The further research is foreseen for performing experiment with more advanced software, such as cloud based systems, in order to compare their advantages against conventional tools.

References

1. Baru, C., Bhandarkar, M., et al.: Benchmarking big data systems and the big data top100 list. Big Data 1(1), 60–64 (2013), doi:10.1089/big.2013.1509
2. Burstein, F., Holsapple, C.W.: Handbook on Decision Support Systems 2: Variations, p. 798. Springer, New York (2008)
3. Gantz, J., Reinse, D.: The digital universe IN 2020: Big Data, Bigger Digi tal Shadows, and Biggest Growth in the Far East (2012)
4. Gartner IT Glossary (2013), `http://www.gartner.com/it-glossary/big-data/`
5. IBM: Deriving Business Insight from Big Data in Banking (2013), `http://www-01.ibm.com/software/data/bigdata/industry-banking.html`
6. Koutonin, M.R.: The Best Data Mining Tools You Can Use for Free in Your Company (2013), `http://www.siliconafrica.com/the-best-data-minning-tools-you-can-use-for-free-in-your-company/`
7. Kovalerchuk, B., Vityaev, E.: Data mining for financial applications. In: Data Mining and Knowledge Discovery Handbook, pp. 1203–1224 (2005)
8. Lavalle, S., Lesser, E., et al.: Big data, analytics, and the path from insight to value. MitSloan Management Review 52(2), 21–31 (2011)
9. McAfee, A., Brynjolfsson, E.: Big data: the management revolution. Harvard Business Review, 1–9 (2012)
10. Dass, R.: Data Minin. In: Banking And Finance: A Note For Bankers- Technical note, Note No.: CISG88 (2006)
11. Shaikh, M., Chhajed, G.: Review on Financial Forecasting using Neural Network and Data Mining Technique 5(2), 263–267 (2012)
12. TechAmerica Foundation's Federal Big Data Commission.: Demystifying Big Data: A Practical Guide to Transforming the Business of Government (2012)

Lean Agile Approach in Development of Modern Mobile and Web Applications

Jaroslav Skrabalek and Tomáš Pitner

Faculty of Informatics, Masaryk University
Botanická 68a, 602 00 Brno, Czech Republic
{skrabalek,tomp}@fi.muni.cz

Abstract. There are significant differences between the development of classical software project and a web/mobile-based project. I try to define new project management approach derived from methodology of large software projects. This approach called LAWA – Lean Agile Web Approach – shall be easily applicable on small projects, to IT product developed from scratch, especially startups, but it should be also applicable in an initiative within a large corporation. The design of LAWA is inspired and benefits from Agile Unified Process, SCRUM and LEAN start-up.

Keywords: Web 2.0, Mobile, Application, Development, Methodology, shortage, Lean, Agile, LAWA.

1 Introduction

Besides the meaning of Web 2.0, strong trend focusing on mobile applications, and what it stands for the Internet, ICT and society, there are significant questions regarding the software development. We can date first basis of systematic project management in the beginning of 20th century. Since then many approaches and software development processes have been established and took into the practice [1]. From simple models like Waterfall we went through Critical Chain Project Management, Extreme Project Management including Scrum and Agile Software Development to the sophisticated methodologies like PRINCE2 or Rational Unified Process by IBM [2].

But those approaches result from an earlier degree of software development appropriate to desktop programs. Due to the changes that Web 2.0 and Mobile introduce to the software development, management methodologies need to adapt to the dynamic and challenging environment. Requirement specification can be changed quickly due to uprising trends and technologies and the competitive landscape demands adaptation of techniques in order to enable adaptable assignment customization and project goals. To reduce significant mistakes – or detecting them in time respectively –, flexible reactions to changing requirements and environments is crucial during the software development process. Traditional methods and techniques are often not suitable for development of modern web and mobile applications [3] [4]

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 47–52, 2013.

[5] due to the specific character of cloud-services [6] [7]; for example fast advancement, strong competitors and the need for continuous development through *perpetual beta* or rapid deployment cycles [7] or sharing the Cloud services with many users with the ability of collaborative work accessing the data wherever users are, while desktop programs are primarily used by one user, on one computer (or small local network) using local computer (local network) resources. Classic development models – in comparison to agile processes – are often inadequate due to high risk of refusal by users, the amount of possible errors [8] [9], and demanded development speed. Agile management emphasizes the focus on human resources, quick reaction on changes and on the importance of customer involvement, and therefore supplements the needs of web-based software development.

1.1 Where Current Project Methodologies Fail

Therefore there is a need for project management approach combining such classical methodologies with more complex approaches reflecting necessity of proper business models closely connected to the development, efficiency and performance of rich Internet and mobile applications (RIMA), security issues, etc. Increasing complexity and usability of Web applications connected with mobile applications have changed people's understanding not only of the Internet but the whole development process of such application, communication interface of that application and the approach to people who create the application in the end. These significant questions regarding the software development [10] are now broadly discussed and the process itself has been redefined. The most important change is diverting from purely technological perception of a final software product to the consideration of product's user acceptance. The functionality is not standing on the first place but usability. The change has second major influence on developers. Companies gradually turn their task orientation to developers where they are carriers of the tasks and potentially of a product success. The more a developer is satisfied the more the task will be fulfilled in a satisfactory manner.

According to [11], the rate of the failure is more frequent with projects managed by traditional rigid methods than projects following Agile approach. However, many companies declare Agile approach in software management but in actual fact they just follow slightly customized *Waterfall model*. The Top management often do not have courage enough to accept full impact of Agile management deployment, which leads in certain hybrid solution on practice – customization of Agile approach to the concrete company's environment. The learning curve of the Agile is quick and can be learn e.g. on the Wikipedia site. However, its incorrect grasp may lead to software develop slowdown (or even it is discontinuation). The more tension due to incorrectly goal setting rise, the more classical methods are followed (surveys are available online in details [12] [13]).

However, "pure agile" is not appropriate for all projects. Many developers, analysts and product managers who experienced achievements with agile projects adopted confidence that Agile is always better than the alternative approaches to software development. In spite of the fact they know the agile approach does not work in all cases, these people involved in software creation rather see problem in the organization itself, never a disconnection between agile principles and business needs. Certainly not every project is suitable for Agile management and therefore PRINCE2, full scale RUP or other traditional methodologies start playing the main role. Making it looks like it's the fault of the organizational structure (because of the lack of business responsibility to deliver software to customers early and adapt it to their needs, prioritizing Return of Investment (ROI) above all, freezing on the plan prepared before all of the factors could be considered, etc.) creates needless tension in the team and organization, discontinues important innovations, led to preservation of old well-tried (but not beneficial today) methodologies. The best approaches to use to address that particular problem or opportunity are often overlooked. Agile projects share certain characteristics (a backlog of work, self-organizing teams, rapid, incremental delivery of new functionality to users) that aren't necessarily value-adding for all projects and under all circumstances. [14]

2 LAWA Design

The main aim was to create and approach satisfying needs mentioned in previous chapter. Those requirements are best addressed by Agile approach, particularly by SCRUM process and Agile Unified Process extended by LEAN startup model. Therefore I have chosen these methodologies as a foundation and I modified them by approaches of other established methodologies. I call this modern software development approach LAWA – Lean Agile Web Approach – nevertheless the Web includes both Web and Mobile applications. Traditional methodologies and Agile approach are delimited (e.g. they are focused only on particular life cycle or software development phase). LAWA covers all phases from the very beginning when the idea and project intention was born through analytical phase to final testing and release part, its marketing and business maintenance. The emphasis is put on initial phases (problem identifying, requirements gathering and their validation on the market) and high speed and scalable development process.

Among to other features that were specified for the design of LAWA belongs:

- Independency on programming languages and software platforms.
- Utilization in the development of new software as well as ability to reapply LAWA in existing applications where no methodology has been used yet.

The class of applications for which LAWA is most suited exhibits a regular Web applications and services as well as mobile applications, both developed by a single person or small team, that can be part of bigger structure actually. Examples include enterprise collaborative applications and mobile front-ends of traditional office applications. Since many mobile and Web applications have volatile life cycle that requires frequent updating, some means to routinize and automate both the initial development and subsequent update process is needed.

Table 1 illustrates the usefulness of the LAWA approach for design and development of Web and mobile applications. The two axes representing the complexity (with reference to [15]) and volatility in time are really continuums rather than discrete branches of dichotomy.

Table 1. Usefulness of the LAWA approach for design and development of Web and mobile applications

		Volatility in time	
		Low	*High*
Complexity	*High*	Low usefulness [e.g. server-side systems, bank system]	High usefulness [e.g. collaborative cloud application]
	Low	Not useful [e.g. catalogue]	Medium usefulness [e.g. mobile application]

Applications in the two domains we mentioned – enterprise collaborative applications and mobile front-ends of traditional office applications – have high time volatility and are highly complex, making the LAWA methodology particularly appropriate. Highly complex applications that remain unchanged over a long period of time can benefit from the LAWA methodology during the idea validation, design and construction phases but do not require much maintenance, so that the updating problem is relatively unimportant and the benefit to be gained by the LAWA approach is not as noticeable. Furthermore, applications that have dynamic complexity and high volatility may gain adequate from the use of the LAWA approach. Finally, at the opposite end of the spectrum, a catalogue may not have a readily discernible complexity and usually remains unchanged over time. In this case, LAWA is not applicable very much but still can be used.

LAWA methodology differs from that presented in SCRUM process and Agile Unified Process (AUP) in several dimensions, including better user involvement, additional project intention validation with application of LEAN startup approach, increased emphasis on graphic representations, and a more detailed, recommended step-by-step procedure for software design and development.

2.1 LAWA Life Cycle Phases

Every software project starts with *preparation phase*. Subsequently, we are proceeding in an iterative way similar to AUP for each part of the project. However, it differs in the emphasis on each phase and in the content that each phase generates. Different roles participate in each phase, but the involvement of all team members is highly recommended.

Preparation Phase. It is fundamental stage of the whole development. The overall project plan is predesigned on the basis of its outcomes. It includes a collection of basic requirements, decomposition into separately manageable parts, definition of needed resources and declaration of the basic behavior. Project outcomes are undestood as a vivid apparatus, in which every part is adjusted permanently. In the beginning, we expect to set up just basic project frame. During the identification of a functional - technical part of the project, the business aspects of the final product or service are validated paralelly. Demand verification for particular features can lead to the redefination of the basic requirements. Since this action is undertaken in the very first moments, the resulting cost of this change is marginal. At the same time it maximizes significantly the ROI for the client. Particular problem areas are subjected to qualitative and quantitative questionnaire survey among the defined target customer segments. Those areas that showed a predetermined frequency, proceed to the next step of identification of – the final solution, revenue source for the resulting service, etc. [16] The financial aspect of the project is optional during this step and depends on the degree of involvement of development department (internal vs. made-to-order development). In this part, only primary roles for the project are inolved - manager, customer and product owner / analyst.

Specification Phase. Specification phase is the first iterative phase of an LAWA methodology. Functionality, section and module with the highest priority is always selected for each iteration. At its beginning, the role manager sets other roles to the team for the whole iteration. Key roles for this phase are manager, sponsor, analyst, area expert and user.

Design Phase. On the basis of processed requirements and conceptual analysis document, detailed behavioral description of the system – design document – is created. Therefore it is a more detailed processing of outputs from the specification phase. A data model is prepared parallely with the design document. Manager, customer, product owner / analyst, expert, users and programmers participates in this phase.

Construction Phase. At this stage, proposed functionality is implemented and cyclicaly validated again. This phase is attended by manager, product owner / analyst, programmers and users.

Verification Phase. Verification phase can take place independently in one iteration, as well as in multiple iterations together. During this phase a detailed testing is perfomed, eventually necessary manuals are created as well. Software is deployed or delivered to the customer, users can be trained and change requests are collected.

3 Future Work, Conclusion

Methodology LAWA proceeds iteratively like AUP. Compared to AUP, LAWA uses a simpler means that requiring little overhead to implementation. It uses prototyping (a conceptual documents). It adapts some part of the UML, but uses other means including custom procedures. One of the most significant benefits comes from the awareness about business aspects of the project and therefore an utilization of LEAN approach minimizing the risk of project failure.

In future work on LAWA, the more detailed description of each phase will be prepared including diagrammatic representation and next simplification of the method will be reconsidered within the meaning of SCRUM process in reflection of Web 2.0 and mobile applications needs.

In the end, a validation itself of LAWA will be undertaken during the expert interviews in the practice.

Acknowledgments. This work was supported by the European Union's territorial cooperation program between Austria and the Czech Republic of the under the EFRE grant M00171, project "iCom" (Constructive International Communication in the Context of ICT).

References

[1] Cleland, D.I., Gareis, R.: Global project management handbook. McGraw- Hill Professional (1993)
[2] Jochen, K., Ahmad, S.: Rup reference and certification guide. IBM Press (2007)
[3] Turk, D., France, R., Rumpe, B.: Limitations of Agile Software Processes. In: Proceedings of the Third International Conference on Extreme Programming and Flexible Process in Software Engineering, XP 2002 (2002)
[4] Holovaty, A., Kaplan-Moss, J.: The Definitive Guide to Django: Web Development Done Right, 2nd edn. Apress Series. Apress (2009)
[5] Kacvinský, M.: Agile software development methodologies. Diploma thesis. Brno, Czech Republic: Masaryk University (2012) Supervised by Jaroslav Škrabálek
[6] Hendler, J., Golbeck, J.: Metcalfe's law, Web 2.0, and the Semantic Web. Web Semantics: Science, Services and Agents on the World Wide Web 6(1), 14–20 (2008)
[7] O'Reilly, T.: What is Web 2.0: Design patterns and business models for the next generation of software. O'Reilly Media (2007)
[8] Cohn, M.: Succeeding with Agile: Software Development Using Scrum. Addison-Wesley (2009)
[9] Abrahamsson, P., Salo, O., Ronkainen, J., Warsta, J.: Agile software development methods: Review and analysis. VTT Electronics, Finland (2002)
[10] Pitner, T.: Web 2.0 – příklady, technologie a problémy. In: Proceedings of Tvorba Software Conference, pp. 164–171. VŠB-TU Ostrava (2007)
[11] Ambler, S.W.: Is Agile really that successful? Dr.Dobb's Journal. Editor TechWeb, UBM (August 2007), http://www.drdobbs.com/architecture-and-design/dr-dobbs-agile-newsletter-1107/204200877?cid=Ambysoft
[12] VersionOne: The 7th Annual state Agile Dev Survey, http://www.versionone.net/state_of_agile_development_survey/10/
[13] Ambysoft: Surveys Exploring The Current State of Information Technology Practices, http://www.ambysoft.com/surveys/
[14] Beal, A.: Traditional vs. Agile: The Future is Hybrid. In: Agile Practises, Leading Organizational Change (2011), http://www.bridging-the-gap.com/traditional-vs-agile-the-future-is-hybrid/
[15] Da-Wei, E.: The software complexity model and metrics for object-oriented. In: 2007 IEEE International Workshop on Anti-counterfeiting, Security, Identification. IEEE (2007)
[16] Blank, S.: Why the Lean Start-Up Changes Everything. Harvard Business Review (2013), http://hbr.org/2013/05/why-the-lean-start-up-changes-everything/

UML Diagrams Generation Process by Using Knowledge-Based Subsystem

Audrius Lopata and Ilona Veitaite

Vilnius University, Kaunas Faculty of Humanities,
Muitines St. 8, Kaunas
{audrius.lopata,ilona.veitaite}@khf.vu.lt

Abstract. The main scope of the paper is to present UML diagrams generation process from enterprise model. The paper analyzes the differences between traditional information systems (IS) engineering and knowledge-based engineering. There is proposed to use knowledge-based subsystem for the purpose of avoiding empirically based IS development process.

Keywords: UML Diagrams, enterprise modeling, CASE tool, IS engineering, knowledge-based systems.

1 Introduction

Information systems (IS) are becoming more complicated and textual information with elementary schemes is not sufficient to describe business processes. Specific models are applied for computerized requirements specification and modeling of IS. Enterprise modeling has become an integral part of IS development process. During this process user requirements, business domain knowledge, software architecture, and other essential components are modeled [8]. Recently, the organization's business modeling has become an important phase of modeling design processes [11].

IS engineering phases from modeling to code generation are traditionally implemented empirically. Computerized model of the subject area is formed by experience of the analyst, while traditional CASE system design models are formed on the basis of the information collected on business domain, which is being computerized. The main role is given to the analyst and enhanced knowledge of the subject area is not fully or insufficiently controlled by formalized criteria [6].

Currently, computerized IS engineering, in order to avoid the empiric, is developing based on new knowledge-based methods. Computerized information system in knowledge-based information system engineering, is developed by using stored enterprise knowledge base of the subject area, i.e., enterprise model, the composition of which is defined by formal criteria. Computerized IS software development based on that model is known as knowledge-based IS engineering [6].

A knowledge–based IS engineering offered system modeling and decision–making methods and tools, which helps to develop more accurate and detailed subject area corresponding to the project [2]. IS project developer and (or) programmer is allowed to use not only the knowledge of the project, which is stored in traditional CASE tool

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 53–60, 2013.

repository, but also the knowledge repository, where formal criteria tested subject area knowledge is stored. There are a number of standards and business modeling methodologies [6]. UML is one of the most common software specification – standards. It is a universal IS modeling language applied to a number of methodologists and used in the most popular modeling tools, such as Enterprise Architect, System Architect, MagicDraw and others. The method of UML diagrams generation from enterprise model implements a knowledge-based design phase in the IS development cycle.

Knowledge-based subsystem as CASE tool component with enterprise meta–model and enterprise model inside can solve this question. Enterprise meta–model is a formal structure which ensures more qualified project development process and knowledge base data collection [9]. Enterprise model and enterprise meta–model make UML project models generation process more effective and qualified and ensure lower number of mistakes in the final IS development phase.

2 Traditional IS Engineering and Knowledge-Based IS Engineering

Information systems engineering – is an information system development process, the execution of which is implemented through the IS development life cycle phases. Each phase of the IS development life cycle was implemented independently of the others, because there was no computerized knowledge repository, in which were included all phases of life cycle. Analysts and designers used conventional or object–oriented IS engineering methods. IS projects were implemented non–computerized or by using specially created computerized tools intended for the specific user requirements specifications to solve the problem. Later, computerized IS development methods were implemented. Information system development improved computerized IS development tools – CASE systems which included a part of the IS development life cycle phases – design, documentation and coding. Analysts and designers used CASE system's IS development tools for conceptual and detailed design phases, and in the practical realization phase programmer used the tools designed to generate code. In this case, a computerized system was based on the traditional IS engineering and object–oriented IS engineering methods. With the development of the traditional IS engineering CASE system covered the entire IS development life cycle – from enterprise modeling to code generation. These life cycle phases are implemented with CASE tool components, such as process modeling tools, detailed analysis tools, transformations tools, database/application design tools, application generation tools. Enterprise modeling process is integrated into IS engineering instruments [1].

CASE system applicable not only in IS engineering, but also in enterprise modeling and re-engineering. CASE systems are complemented with enterprise modeling tools and CASE system repository – enterprise model subsystem. System analyst is the main user of enterprise modeling CASE tool component.

Computerized IS development process is based on the design from the model sequence execution, where every other phase of the design model is created interactively in the presence of the analyst, designer and programmer [7].

At that period CASE systems automatically generated only the logical database diagrams, user interface fragments and code, which programmer basically had to adjust. Thus, the traditional computerized IS engineering, IS development life cycle starting from the initial and ending with its last phase takes place empirically and many design models in CASE tools were generated partially. Only analyst could fully implement them on the basis of empirical experience.

The computerized IS engineering specific methods are developed based on common requirements, which systematize the selected methodology. There are some necessary components for knowledge-based IS engineering methodology [4, 5]:

- Core of enterprise modeling theoretical basis is the theoretical enterprise model. Its purpose is to identify the necessary and sufficient business components for IS engineering, which implements enterprise's business management. Theoretical enterprise model is a formalized enterprise management model that identifies business components and their interactions, enterprise management and their interactions.
- Theoretical enterprise knowledge model is enterprise meta–model. Based on the IS development life cycle phases the theoretical model components and their interactions are described as enterprise meta–model. Enterprise meta–model is a structural model, which specifies the necessary and sufficient components of IS engineering enterprise management features and interactions.
- The theoretical basis of knowledge-based IS engineering process is IS engineering process model (methodology) that bases knowledge-based IS development process by using an enterprise model as an additional knowledge source besides the analyst and the user.
- Computerized IS engineering systems development is based on knowledge-based IS engineering tools: enterprise modeling approach, design models, use case methods. Right tools are needed for knowledge-based IS development life cycle phase's implementation, such as practical knowledge-based modeling methods or models sets. Practical knowledge-based modeling methods are intended for the development of functionality of traditional CASE systems by creating a knowledge-based – CASE intelligent system.
- Computerized knowledge-based IS engineering project management basis is CASE system knowledge-based subsystem (Fig. 1). CASE system's knowledge-based subsystem's core component is knowledge base, which essential elements are enterprise meta–model specification and enterprise model for certain problem domain. Knowledge-based subsystem is one more active participant of IS engineering process beside analyst, whose purpose is to verify results of IS life cycle phases.

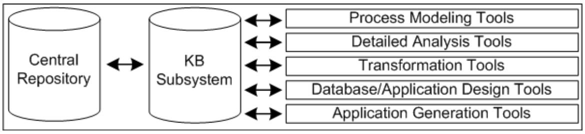


Fig. 1. Knowledge-based CASE Tool Components

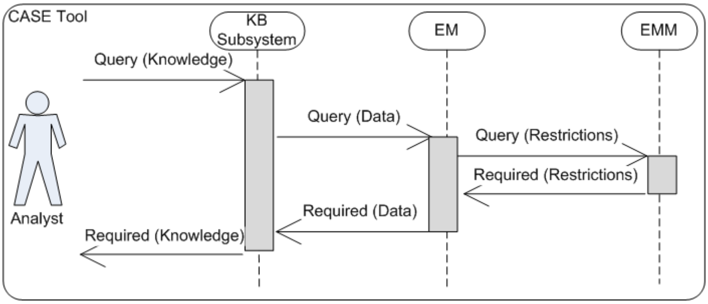


Fig. 2. Knowledge-based subsystem connection to the enterprise model and enterprise meta–model inside CASE tool

Knowledge-based CASE systems containing essential components, which organize knowledge: knowledge-based subsystem's knowledge base, which essential elements are enterprise meta–model specification and enterprise model for certain problem domain. Principal scheme of interaction between CASE tool's knowledge-based subsystem, enterprise model and enterprise meta–model is presented in the figure (Fig. 2).

Traditional IS engineering and knowledge-based IS engineering have qualitative differences. In case of traditional IS engineering – there is an empirical IS engineering, where individual user does not include whole enterprise processes. A knowledge-based IS engineering covers whole enterprise specification (enterprise model), because it specifies the essential enterprise's characteristics. Formalized enterprise model is based on theoretical enterprise's management principles [3].

Information system is developed empirically in traditional computerized IS engineering. In knowledge-based computerized IS engineering information system is developed by using enterprise knowledge repository, where necessary and sufficient computerized knowledge of the subject area is stored [6].

In a knowledge-based computerized IS engineering design there are less logical breaks than in traditional computerized IS engineering. The logical break of a design is, when consistent automated information system design process is terminated to allow the analyst to enter the missing information of IS engineering process. Logical breaks exist not only in theoretical IS engineering models, but also are observable in many CASE tools, such as System Architect, MagicDraw, Enterprise Architect and others. The majority of CASE tool's project models are generated partially and fully implementation possible just by using the systems analyst experience.

In a knowledge-based computerized IS engineering, all project models can be generated interactively by using generation algorithms, if the necessary knowledge will be collected into knowledge repository. A minimum analyst and designer participation is required for ensuring a missing knowledge entry. Knowledge is tested for completeness; it is verified to ensure automatically generated design models and software code quality in knowledge gathering into knowledge repository phase.

The differences in the types of are engineering methods, sources of knowledge, i.e. participants, who are involved in IS development process and necessary software is described below:

- Traditional IS Engineering (non–computerized) – Methods: IS life–cycle stages based design using individual methods; Sources of knowledge (participants)/People: User, Analyst, Designer, Programmer; Software: Database design tools, Development Tools.
- Computer–aided software engineering– Methods: Computer–aided software engineering method includes parts from IS life cycle: Design method, Realization method; Sources of knowledge (participants)/People: User, Analyst, Designer, Programmer Software: User specification tools, Database design tools, User interface design tools, Use Case tools, Development tools.
- Computerized IS Engineering (with enterprise modeling) – Methods: Computerized IS Engineering (CASE) method includes all IS life cycle: Enterprise modeling, Design method, IS realization method; Sources of knowledge (participants)/People: User, Analyst, Designer, Programmer Software: Enterprise modeling tools, User requirements specification tools, Database design tools, User interface design tools, Use Case tools, Development tools.
- Knowledge–based IS engineering – Methods: Knowledge–based CASE method includes: Enterprise knowledge modeling, Enterprise modeling, Design method, Realization method; Sources of knowledge (participants)/People: 1. User, Analyst, Designer, Programmer 2. CASE knowledge-based subsystem Software: Enterprise modeling tools, Knowledge-based subsystem, User requirements specification tools, Database design tools, User interface design tools, Use Case tools, Development tools.

3 Generation Process

The current version of UML is UML 2.5 (beta), released in October 2012. 2.5 tools will have to support a complete UML specification [13]. Computer aided design tools facilitate the work of the designer, but still does not automate enough, because the analyst inclusion is required. The main purpose is to formalize as much as possible model creation process. A variety of methodologies has been designed, but still there are missed opportunities that would improve the IS design methods that are controlled by CASE tools [10].

Information system design methods specify the sequence of systems engineering actions, i.e. how, in what order and what UML diagrams to use in the design process and how to implement the process. Meaning of each diagram can be defined individually, but more important is the fact that each diagram is the projection of the system model [12]. This kind of system description is quite confusing, because most of the information in the diagrams overlap and describe the same things just in different ways. Formalization brings new software and IS design and development opportunities. When the maximum coherence of UML models is reached, models are linked to each other clearly articulated by the rules, expressed stricter and fuller. This greatly facilitates the task of automated software development. Interaction between UML diagrams and enterprise model is realized through the transformation algorithms.

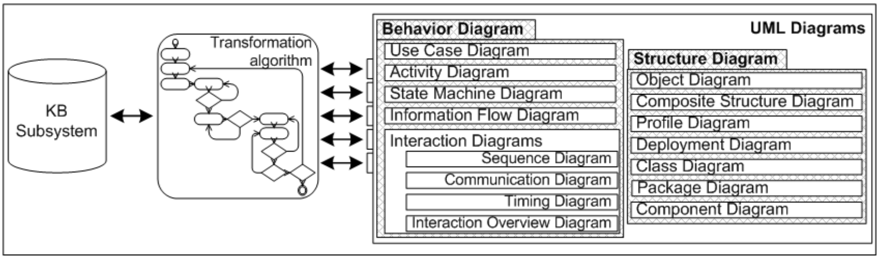


Fig. 3. Generation process of UML diagrams by using the transformation algorithm

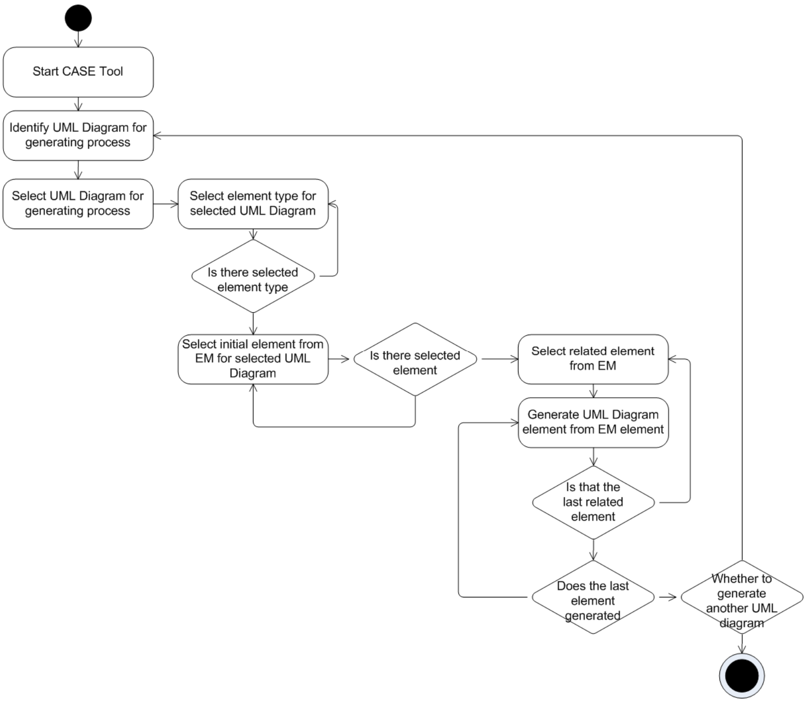


Fig. 4. Algorithm of EM based UML diagrams generation process

Enterprise model as organization's knowledge repository allows to generate UML diagrams after using the transformation algorithms (Fig. 3) Such repository can be used not only for knowledge of the organization accumulation, but also as a tool that minimizes IS reengineering scope of work if changes occur in an organization.

In information system engineering all design models are implemented on the basis of the empirical expert experience. Experts, who participate in the IS development process, do not gain enough knowledge, and process implementation in requirements analysis and specification phases can take a too long time. Enterprise meta–model contains essential elements of business modeling methodologies and techniques, which insures a suitable UML diagrams generation process.

Transformation algorithm (Fig. 4) is top level algorithm for enterprise meta–model based UML diagrams generating process. Main steps for generating process are identifying and selecting UML diagram for generating process, identifying starting elements for the selected UML diagram and selecting all related elements, reflecting enterprise model elements to UML diagram elements and generating the selected UML diagram.

Table 1. Reflections of some basic EM elements to UML diagram elements

Enterprise meta-model elements	UML Diagram	Generated UML diagram element
Material Flow	Activity Diagram	Control Node
Actor	Use Case Diagram	Actor
Process	Activity Diagram	Activity Partition
Process	Use Case Diagram	Use Case
Business Rules	Activity Diagram	Object Node
Information Flow	Activity Diagram	Activity Edge
Function	Use Case Diagram	Use Case
et cetera	*et cetera*	*et cetera*

The second table presents the reflections of some basic EM elements generated to UML diagram elements. An enterprise model with organization's knowledge in the repository allows, by using the transformation algorithms, to generate UML diagrams. This repository can be used not only as knowledge base of the organization's activities collection, but also as a tool that minimizes IS re-engineering scope of work in an organization. Intellectualization of IS engineering design phase means ensuring that UML diagrams generated from enterprise model are correct, match the formal notation and timely accomplished.

4 Conclusions

In Computer aided IS engineering is based on empiric and IS development life cycle phases are implemented on the basis of the experience of experts. A majority of the CASE tools design models are generated only partially, full implementation is possible only non–automatic and with the presence of the expert. Today IS engineering should be based on knowledge.

Knowledge-based CASE Tool contains the necessary knowledge in organizing component – CASE Tool knowledge-based subsystem, which contains enterprise meta–model specification and enterprise model with specific business information, which is compatible with the meta–model's restrictions.

The paper deals with possibility to generate UML diagrams from enterprise model. Every element of UML diagrams can be generated from the enterprise model using CASE Tool knowledge bases subsystem and transformation algorithms. Method of UML diagrams generation process from enterprise model could implement whole knowledge-based IS development cycle design phase.

It is planned to make a detailed enterprise meta–model and UML diagrams meta–models comparative analysis, to show, that the chosen enterprise model composition is sufficient to generate all UML diagrams.

References

[1] FFIECC. Development Procedures. IT Examination Handbook InfoBase (2012), http://ithandbook.ffiec.gov/it-booklets/development-and-acquisition/development-procedures.aspx

[2] Freiberg, M., Baumeister, J., Puppe, F.: Interaction Pattern Categories — Pragmatic Engineering of Knowledge-Based Systems. In: 6th Workshop on Knowledge Engineering and Software Engineering (KESE6) at the 33rd German Conference on Artificial Intelligence, Karlsruhe, Germany (September 21, 2010)

[3] Gudas, S.: Žiniomis grindžiamos IS inžinerijos metodų principai. - Konferencijos pranešimų medžiaga Informacinės technologijos 2004, T.2, Kaunas, Technologija (2005)

[4] Gudas, S.: Enterprise knowledge modelling: Domains and aspects. Technological and economic development of Economy. Baltic Journal on Sustainability, 281–293 (2009)

[5] Gudas, S.: Architecture of Knowledge-Based Enterprise Management Systems: A Control View. In: Proceedings of the 13th World Multiconference on Systemics, Cybernetics and Informatics (WMSCI 2009), Orlando, Florida, USA, July 10 –13, vol. III, pp. 161–266 (2009) ISBN -10: 1-9934272-61-2 (Volume III), ISBN -13: 978-1-9934272-61-9 (Volume III)

[6] Gudas, S.: Informacijos sistemų inžinerijos teorijos pagrindai. Vilnius University (2012)

[7] Gudas, S., Lopata, A.: Žiniomis grindžiama sistemų inžinerija. Mokomoji medžiaga. Vilnius University. Kaunas Faculty of Humanities (2011)

[8] Kerzazi, N., Lavallée, M., Robillard, P.N.: A Knowledge-Based Perspective for Software Process Modeling. e-Informatica Software Engineering Journal 7 (2013)

[9] Lopata, A., Gudas, S.: Meta-Model Based Development of Use Case Model for Business Function. Information Technology and Control 36(324) (2007)

[10] Perjons, E.: Model-Driven Process Design. Aligning Value Networks, Enterprise Goals, Services and IT Systems. Department of Computer and Systems Sciences, Stockholm University. Sweden by US-AB, Stockholm (2011) ISBN 978-91-7447-249-3

[11] Sommerville, I.: Software engineering, 9th edn. Pearson Education, Inc., Addison-Wesley (2011) ISBN-13: 978-0-13-703515-1

[12] UML Diagrams. Diagrams characteristics (2012), http://www.uml-diagrams.org

[13] OMG UML. Unified Modeling Language version 2.5. Unified Modeling (2012), http://www.omg.org/spec/UML/2.5/Beta1/

Using Ontologies to Integrate Multiple Enterprise Architecture Domains

Gonçalo Antunes[1], Artur Caetano[1,2],
Marzieh Bakhshandeh[1], Rudolf Mayer[3], and José Borbinha[1,2]

[1] Instituto Superior Técnico, University of Lisbon,
Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal
[2] Information Systems Group, INESC-ID,
Rua Alves Redol 9, 1000-029 Lisboa, Portugal
[3] SBA Research, Favoritenstraße 16, 1040 Wien, Austria
{goncalo.antunes,artur.caetano}@ist.utl.pt

Abstract. A goal of enterprise architecture is to align the business with the underlying support systems. An enterprise architecture description encompasses an heterogeneous spectrum of domains, such as business processes, application components, metrics, people and technological infrastructure. Architectural views express the domain elements and their relationships from the perspective of the system stakeholders. As a result, a view needs to be expressed using a domain language that addresses the specific concerns of its stakeholders. However, enterprise architecture description languages are often based on generic or broad meta-models that cross-cut distinct architectural domains. But describing each domain through a specialized language and then integrating it with the other domains raises challenges at the level of traceability and consistency. This paper proposes using ontologies to specify different enterprise architecture domains and to integrate and analyse these models. This goal is realized through a domain-independent language that is extended by domain-specific languages, each focussing on a set of specific domain concerns. The approach contributes to the alignment of the different domains while ensuring traceability between then concepts. The proposal is demonstrated through an evaluation scenario that uses ArchiMate as the domain-independent language extended with a set of domain-specific languages. The demonstration shows that the architecture domains can be integrated and analysed through the use of ontologies.

Keywords: enterprise architecture, ArchiMate, ontology, OWL, alignment, separation of concerns.

1 Introduction

Enterprise architecture (EA) is defined by Lankhorst as "a coherent whole of principles, methods, and models that are used in the design and realization of an enterprise's organizational structure, business processes, information systems, and infrastructure" whose models "focus on alleviating the infamous

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 61–72, 2013.

business–information technology alignment problem" [1]. Alignment results from applying models, methods, patterns and best practices to the specification and governance of the different domains of an organisation [2,3,4,5]. Managing the dependencies between these concepts is fundamental for supporting the communication between the different stakeholders and to maintain the consistency at model and meta-model level [6,7]. Moreover, EA governance requires the ability to analyse artefacts [1,8] and is also required to assist business analytics [9].

Despite the efforts for developing comprehensive approaches to enterprise architecture, such as TOGAF [10], a "one language fits all" approach seems to be unable to address specific domains of an organization [11,12]. The specific needs of different organizations place particular demands on the required EA artefacts. As such, the development of an architecture description language entails ensuring the consistency and traceability between the language concepts [13,1]. On the other hand, creating a consistent and comprehensive architecture description language that deals with specific domains is a challenge despite existing situational method approaches to enterprise architecture [14].

The ISO 42010 standard suggests describing a system's architecture through multiple views to address the specific interests of the stakeholders on the system. An architecture description should therefore aggregate multiple views, materialized as a set of models, that are formulated according to viewpoints expressing the concerns of the stakeholders of the system-of-interest [15]. In this way, an architecture works as a communication agent between stakeholders, as each is presented with is own view over the system of interest. But creating different viewpoints may actually require using different meta-models, tools, and validation mechanisms. The integration and extension of models and underlying meta-models is common [16,17], but raises challenges at the level of traceability and consistency because it is difficult to trace concepts between different languages and domains, a problem that is aggravated as the models evolve [18]. Moreover, the integration of different meta-models poses multiple challenges [19].

This paper is concerned with integrating multiple enterprise architecture domains while preserving traceability between the interrelated concepts. The goal is to integrate multiple enterprise architecture description languages as a means to assist alignment. Specifically, we investigate whether ontology technologies (OWL-DL in particular) can be used to specify, integrate and analyse multiple description languages.

Ontologies describe a domain model by associating meaning to its terms and relations. A more formal and widely used definition is that of Grüber who defines an ontology as a "formal specification of a conceptualisation" [20]. The importance of this technology is evidenced by the growing use of ontologies in a variety of application areas [21,22] and, especially, by their role on the Semantic Web initiative [23,24]. Ontology technologies are also used in the field of enterprise architecture to formalize organizational artefacts and to assist with model analysis [25,26,27,28,29,30]. In fact, there is a wide body of knowledge that may improve the practice of EA, including ontology matching [31], and model extension and validation [32]. Ontologies facilitate the construction of complex models and can assist model analysis by depicting the consequences of a model. Formal ontology technologies also contribute to viewing and understanding the implicit consequences of explicit statements and can help ensuring that a model is consistent [33].

This paper posits that modelling the different enterprise architecture domains with a set of integrated description languages contributes to their alignment because consistency and traceability become ascertained. The approach entails using ontologies to represent and integrate the multiple architecture description languages and to analyse the resulting models. We argue for the integration of ontologies and associated technologies as mechanisms for developing consistent enterprise architecture models. The combination of formally specified models with their analysis via automatic mechanisms contributes to aligning the heterogeneous domains of an EA. One example is the impact analysis of changes from the business on the IT infrastructure and vice-versa. The main contribution of this paper is thus proposing an architecture based on the use of ontologies with the purpose of enhancing the extensibility with domain-specific aspects while enforcing consistency. We demonstrate the applicability of the proposal through the application of formal ontologies to model a set of different EA domains and through the consistent integration of these domains. In particular, we develop an ontology to specify the ArchiMate 2.0 meta-model and then create traceable maps to it from a set of domain-specific languages. We also describe an example that maps the sensor technology domain to ArchiMate in the context of a real-world scenario. This demonstration shows that the application of ontologies to enterprise architecture modelling effectively assists consistently aligning and analysing different domains.

The rest of this paper is organized as follows: section 2 describes a ontology-based proposal to integrate and analyse enterprise architecture models; section 3.1 describes the realization of the proposal; section 3 evaluates the solution using a scenario; finally, section 4 concludes the paper.

2 Enterprise Architecture Domain Integration

This paper proposes an ontology-based framework to formalize and integrate different domains of an enterprise architecture. The design of this artefact adheres to the following architectural principles:

- **Concern orientation.** The architecture represents the concepts that address an explicit set of concerns as a meta-model. The meta-model does not support any concepts that are not derived from the stakeholders' concerns.
- **Viewpoint-orientation.** The architecture supports defining views over subsets of its concepts. This facilitates communication because viewpoints act as a separation of concerns mechanism. Viewpoints facilitate addressing multiple concerns and can improve decision-making by isolating certain aspects of the architecture according to the needs of stakeholders.

- **Expressiveness.** The architecture represents a set of unambiguous domain concepts. This entails defining the minimum set of types and relationships to describe the domain.
- **Extensibility.** The architecture supports the integration of multiple domain-specific and domain-independent meta-models while minimizing coupling.
- **Modularity.** The architecture observes high-cohesion and low-coupling. Observing these qualities contributes to the expressiveness and extensibility of the architecture with the goal of minimizing the impact of adding new domain-specific concepts.

The ontology-based framework uses a meta-model to formalize the upper-level or core concepts. This meta-model is formalized as an upper-level ontology and is designated as *domain-independent ontology* (DIO). The design goal of the DIO is to represent the set of concepts pertaining to the central modelling domain. The DIO concepts are extended by defining a variable number of domain-specific meta-models, each depending on a particular system concern. Each domain-specific meta-model is formalized as a *domain-specific ontology* (DSO). Thus, a DSO represents a domain-specific language that addresses a particular set of concerns, and should also have the minimum set of concepts required to describe the domain. Therefore, separation of concerns, low-coupling and high-cohesion are the primary qualities that affect DIO and DSO design.

Ontology integration is required to link concepts from the DIO to each DSO. Integration combines different ontologies while ensuring consistency and maximum coverage of the domain being addressed. The simplest case is that of integrating the DSO concepts with the core concepts represented in the DIO. Cross-DSO integration occurs whenever more expressiveness is required to model a specific domain. Ontology integration makes use of model transformation, which involves defining a mapping strategy from a source to a target model [34,35]. Figure 1 depicts the types of transformation mapping strategies between the DIO and the DSOs. Ideally, a map defines a one-to-one correspondence between each pair of concepts from a source and destination. But three types of mapping deficiencies may occur [36]: a source concept may map to more than one destination concept resulting in *overload*, a source concept may not be mappable to any destination resulting in *deficit*, or several source concepts may map to the same destination concept leading to *redundancy*. Deficiencies can be addressed by revising the DSO so that a one-to-one correspondence is achieved. If not possible, the deficiencies are addressed when querying or reasoning with the ontologies.

One of the goals of integrating the multiple ontologies is to analyse the resulting model instances. The analysis is performed by querying the models or by using a reasoner to infer the model's properties and relations. Four types of reasoning are possible with this architecture:*DIO reasoning* when inference is limited to the DIO concepts, *Single DSO reasoning*, when inference is limited to the concepts of a single DSO, *Cross-DSO reasoning*, when inference use concepts from more than one DSO, and *DIO-DSO reasoning*, when inference uses concepts from the DIO and one or more DSOs.

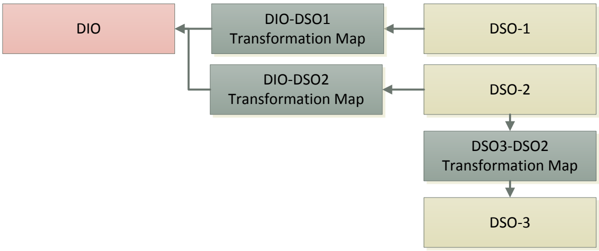


Fig. 1. Mappings between domain-independent and domain-specific ontologies

3 Application to Enterprise Architecture

This section describes the specification of a DIO, a DSO, the transformation maps and an application of the integrated meta-models. The evaluation scenario concerns a civil engineering safety authority that monitors structures such as hydroelectric power dams, reservoirs and bridges. Each structure has different sensors that measure physical phenomena and produce data that is analysed to assess its dynamics. The business process that deals with structure assessment includes activities to acquire data and to analyse data. Instance of this process are long-running as they may be active for decades, from the early construction phases until structure disposition. Part of this process is supported and automated by an information system that provides the following functions:

– Instrumentation: manages sensor installation, configuration and deployment.
– Transformation: manages the algorithms that transform sensor raw data into information.
– Observation: manages geodetic data, visual inspections data, and the data acquired from monitoring systems.
– Analysis: manages data analysis, visualization and reporting.
– Synchronization: synchronizes data between multiple geographic locations and logical systems.

The authority is required to acquire and preserve the monitoring data during the structure's life cycle. Therefore, capturing and preserving the information about the acquisition processes and supporting technological infrastructure is fundamental to attest the provenance and authenticity of the monitoring data. Moreover, historic data can be used to analyse and predict the behaviour of the structure. In this setting, enterprise architecture plays a valuable role to assist with the specification, evolution and the alignment of these processes with the supporting technology.

This scenario was modelled with the ArchiMate 2.0 enterprise architecture modelling language [37]. Although ArchiMate is able to specify the different domains of the scenario at a high-level of abstraction, it has not the expressiveness to model the domain-specific concerns pertaining to sensors and data acquisition. The first step was to create an ArchiMate model of the scenario. The model was produced with the Archi tool [1]). Figure 2 depicts an overview of the acquisition

[1] `http://archi.cetis.ac.uk/`

process and the services supporting it. Figure 3 depicts an overview of the application components and underlying technological infrastructure. The ArchiMate model was then exported from the Archi tool and automatically converted to OWL with a tool developed for that purpose.

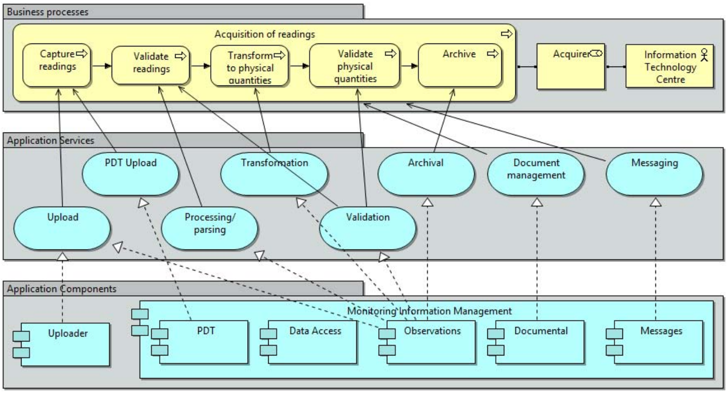


Fig. 2. Business processes and application infrastructure

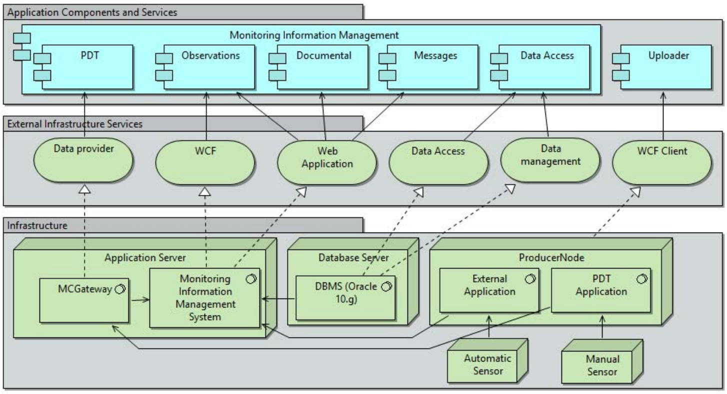


Fig. 3. Technological infrastructure

3.1 The ArchiMate Domain-Independent Ontology

ArchiMate describes the core concepts pertaining to enterprise architecture. The DIO is therefore a specification of the ArchiMate meta-model using OWL-DL. OWL-DL enables taking advantage of existing inference and querying mechanisms to analyse the models and assessing their consistency. The ArchiMate ontology was mainly developed according to the ontology engineering methodology defined by Horridge [38]. The steps include:

1. Identification of the concepts and concept hierarchy.
2. Identification of the disjoint concepts.
3. Modelling composition.
4. Addition of all the relationships between concepts.
5. Identification of definitions.
6. Addition of annotations.
7. Refinement of the ontology through various iterations of the above steps.

The resulting DIO represents the ArchiMate concepts as OWL *Classes* and relations as OWL *ObjectProperties*. Restrictions were added to the properties, such as *InverseObjectProperties* and *SuperObjectProperties*, so that ArchiMate's derived relationships are correctly inferred by the reasoner. Figure 4 depicts a partial OWL-DL specification of ArchiMate's *Business Function* as displayed in Protégé 4.3.

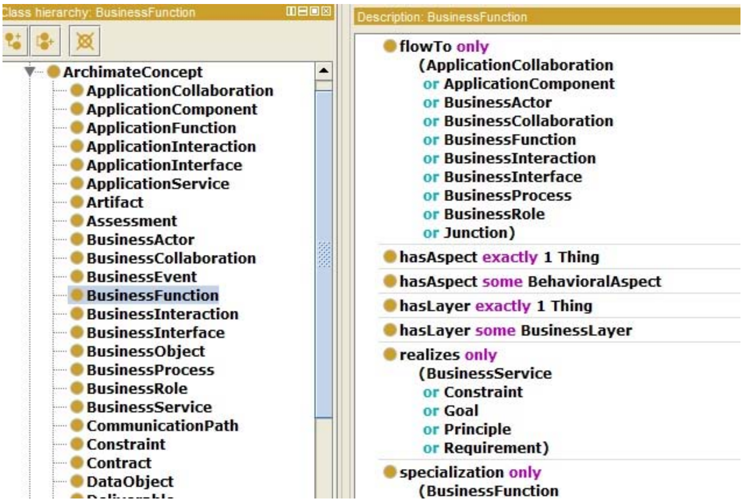


Fig. 4. Partial OWL-DL specification of the ArchiMate Business Function

3.2 The Domain-Specific Ontology

In this scenario, the organizational stakeholders required modelling and analysing specific information about sensors. However, it is out of scope of the ArchiMate language providing the expressiveness to capture the specifics of this particular domain. Sensors measure values that are processed to perform structural analysis. There are sensors for making different types os measurements, which have specific transformation algorithms and calibration parameters. Some sensors are georeferenced and others capture data according to dynamic acquisition rates. As such, the particularities of this domain imply defining a domain-specific language. The organization evaluated different sensor modelling languages, such as

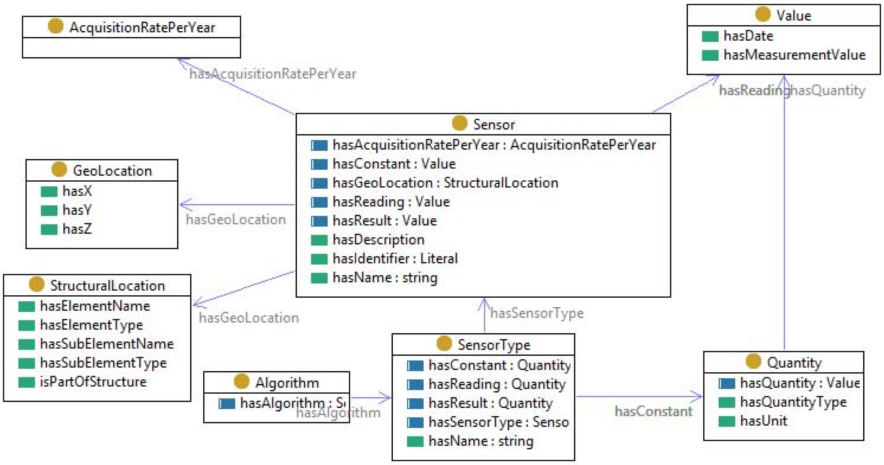


Fig. 5. Structure of the sensor DSO

SensorML [2] and TransducerML[3], but suggested a language based on SensorML to address its specific concerns. As a result, we developed the sensor DSO using the ontology engineering methodology described earlier. The core concepts of the sensor DSO are depicted in figure 5. The transformation map between the sensor DSO and the ArchiMate DIO contains the relations described on Table 1.

Table 1. Transformation map between the Sensor DSO and the ArchiMate DIO

Sensor DSO	**ArchiMate DIO**
Sensor	Node
GeoLocation	Location
StructuralLocation	Location
Algorithm	ApplicationComponent
Value	Data Object
AcquisitionRatePerYear	Data Object

3.3 Model Analysis

One of the stakeholder concerns relates to the technological infrastructure elements that support the acquisition process. This concern can be addressed through DIO reasoning, i.e. via the ArchiMate meta-model. Figure 6 depicts the question formalized as an OWL-DL query along with the *ObjectProperty* chains that identify the 19 instances that support the acquisition process. An example of intra-DSO reasoning is depicted in figure 7 that depicts the sensors that are able to make temporal readings. Finally, figure 8 shows the result of DIO-DSO reasoning where the integrated models are queried about which ArchiMate

[2] http://www.ogcnetwork.net/SensorML
[3] http://www.ogcnetwork.net/infomodels/tml

Fig. 6. Intra-DIO query results

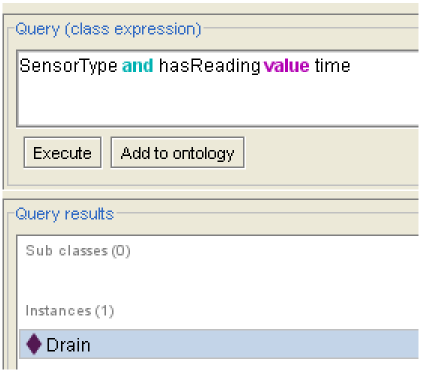


Fig. 7. Intra-DSO query results

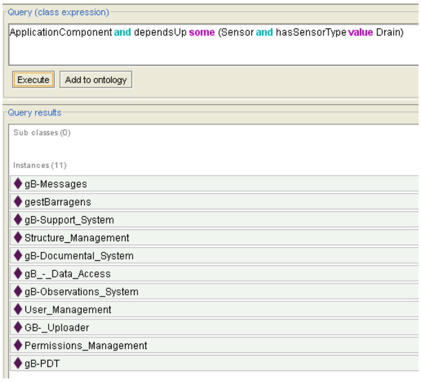


Fig. 8. Cross DIO-DSO query results

Application Components rely on the reading of the sensors of type *Drain*. The reasoner uses the mappings between the sensor DSO and the ArchiMate DIO to infer the reasoning chains and thus to answer the query.

4 Conclusions

This paper proposes using ontologies to integrate different enterprise architecture domains and to analyse the resulting model instances. This goal is realized through the specification of a core domain-independent language that is extended by multiple domain-specific languages, each focussing on a set of specific concerns. The approach contributes to the alignment of the different domains while ensuring traceability, consistency and extensibility. As observed from the case study, ontologies can enhance the quality of meta-modelling due to their automated analysis capability that can be used to assess meta-model consistency as well as model conformance. Moreover, ontologies positively contribute to enterprise architecture alignment because multiple meta-models can be integrated and represented in such a way that its information can be traced and analysed while the reasoning consequences are exposed. The proposal was evaluated using ArchiMate as the DIO. To do that, we converted the ArchiMate meta-model to OWL-DL. A scenario was modelled using the ArchiMate DIO and its domain-specific aspects were modelled using a set of integrated DSOs. This paper partially described one of the DSOs, the sensor DSO, and exemplified different analysis types that can be accomplished using this approach. This demonstration shows that the application of ontologies to enterprise architecture modelling effectively assists aligning and analysis different domains.

Our current work focuses on extending the analysis capabilities to support the validation of models and the assessment of models and meta-models. We are also working on a set of automated and semi-automated extractor and process mining tools to instantiate the domain-specific and domain-independent ontologies with operational data to test the conformance of the "should-be" models towards the actual "as-is" models.

Acknowledgements. This project is partially supported by the European Commission under the 7th Framework Programme (FP7/2007-2013) under grant agreement 269940, TIMBUS project (`http://timbusproject.net`) and by COMET K1, FFG, Austrian Research Promotion Agency. Gonçalo Antunes is partially supported by FCT, Fundação para a Ciência e a Tecnologia, through project PEst-OE/EEI/LA0021/2011 and grant SFRH/BD/69121/2010.

References

1. Lankhorst, M.: Enterprise Architecture at Work: Modeling, Communication, and Analysis, 3rd edn. Springer (2013)
2. Pereira, C.M., Sousa, P.M.A.: Enterprise architecture: business and it alignment. In: Haddad, H., Liebrock, L.M., Omicini, A., Wainwright, R.L. (eds.) SAC, pp. 1344–1345. ACM (2005)
3. Winter, R., Fischer, R.: Essential layers, artifacts, and dependencies of enterprise architecture. In: 10th IEEE International Enterprise Distributed Object Computing Conference Workshops, EDOCW 2006, p. 30. IEEE (2006)
4. Wegmann, A., Balabko, P., Lê, L.S., Regev, G., Rychkova, I.: A method and tool for business-it alignment in enterprise architecture. In: Proceedings of the CAiSE, Citeseer, vol. 5, pp. 113–118 (2005)
5. Caetano, A., Silva, A.R., Tribolet, J.M.: Business process decomposition - an approach based on the principle of separation of concerns. Enterprise Modelling and Information Systems Architectures 5(1), 44–57 (2010)
6. Galster, M.: Dependencies, traceability and consistency in software architecture: towards a view-based perspective. In: ECSA 2011 Proceedings of the 5th European Conference on Software Architecture: Companion Volume (2011)

7. Romero, J., Jaen, J.I., Vallencillo, A.: Realizing correspondences in multi-viewpoint speci cations. In: Proceedings of the 2009 IEEE International Enterprise Distributed Object Computing Conference (2009)
8. Buschle, M., Ullberg, J., Franke, U., Lagerstrom, R., Sommestad, T.: A tool for enterprise architecture analysis using the PRM formalism. In: Soffer, P., Proper, E. (eds.) CAiSE Forum 2010. LNBIP, vol. 72, pp. 108–121. Springer, Heidelberg (2011)
9. Binz, T., Leymann, F., Nowak, A., Schumm, D.: Improving the manageability of enterprise topologies through segmentation, graph transformation, and analysis strategies. In: 2012 16th IEEE International Enterprise Distributed Object Computing Conference, EDOC (2012)
10. The Open Group: TOGAF 9 - the open group architecture framework version 9 (2009)
11. Saat, J., Franke, U., Lagerstrom, R., Ekstedt, M.: Enterprise architecture meta models for it/business alignment situations. In: 2010 14th IEEE International Enterprise Distributed Object Computing Conference (2010)
12. Buckl, S., Matthes, F., Schweda, C.M.: Conceptual models for cross-cutting aspects in enterprise architecture modeling. In: 2010 14th IEEE International Enterprise Distributed Object Computing Conference Workshops (2010)
13. Braun, C., Winter, R.: A comprehensive enterprise architecture metamodel and its implementation using a metamodeling platform. Enterprise Modelling and Information Systems Architectures, 64–79 (2005)
14. Buckl, S., Schweda, C.M., Matthes, F.: A design theory nexus for situational enterprise architecture management. In: 2010 14th IEEE International Enterprise Distributed Object Computing Conference Workshops (2010)
15. ISO: ISO/IEC/IEEE 42010:2011 - systems and software engineering - architecture description (2011)
16. Burmester, S., Giese, H., Niere, J., Tichy, M., Wadsack, J.P., Wagner, R., Wendehals, L., Zundorf, A.: Tool integration at the meta-model level: the fujaba approach. Int. J. Softw. Tools Technol. Transfer 6, 20–218 (2004)
17. Kapsammer, E., Reiter, T., Schwinger, W.: Model-based tool integration - state of the art and future perspectives. In: Proceedings of the 3rd International Conference on Cybernetics and Information Technologies, Systems and Applications (CITSA 2006) (2006)
18. France, R., Rumpe, B.: Model-driven development of complex software: A research roadmap. In: Future of Software Engineering, FOSE 2007 (2007)
19. Almeida, J.P.A., Iacob, M.E., van Eck, P.: Requirements traceability in modeldriven development: Applying model and transformation conformance. Inf. Syst. Front 9, 327–342 (2007)
20. Grüber, T.: A translation approach to portable ontology specifications. Knowledge Acquisition 5(2), 199–220 (1993)
21. Happel, H.J., Seedorf, S.: Applications of ontologies in software engineering. In: 2nd International Workshop on Semantic Web Enabled Software Engineering (SWESE 2006), held at the 5th International Semantic Web Conference (ISWC 2006) (2006)
22. Bürger, T., Simperl, E.: Measuring the benefits of ontologies. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM-WS 2008. LNCS, vol. 5333, pp. 584–594. Springer, Heidelberg (2008)
23. Shadbolt, N., Hall, W., Berners-Lee, T.: The semantic web revisited. IEEE Intelligent Systems Journal 4, 96–101 (2006)
24. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web, pp. 29–37. Scientific American (May 2001)

25. Uschold, M., Uschold, M., King, M., King, M., House, S.B.R., Moralee, S., Moralee, S., Zorgios, Y., Zorgios, Y.: The enterprise ontology. The Knowledge Engineering Review 13, 31–89 (1995)
26. Geerts, G.L., McCarthy, W.E.: An ontological analysis of the economic primitives of the extended-rea enterprise information architecture. International Journal of Accounting Information Systems 3(1), 1–16 (2002)
27. Kang, D.: An ontology-based enterprise architecture. Expert Systems with Applications 37, 1456–1464 (2010)
28. Wagner, G.: Ontologies and rules for enterprise modeling and simulation. In: 2011 15th IEEE International Enterprise Distributed Object Computing Conference Workshops (2011)
29. Azevedo, C.L.B., Almeida, J.P.A., van Sinderen, M., Quartel, D., Guizzardi, G.: An ontology-based semantics for the motivation extension to archimate. In: 2011 15th IEEE International Enterprise Distributed Object Computing Conference (2011)
30. Almeida, J.P.A., Guizzardi, G.: An ontological analysis of the notion of community in the rm-odp enterprise language. Computer Standards & Interfaces 35(3), 257–268 (2013)
31. Shvaiko, P., Euzenat, J.: A survey of schema-based matching approaches. In: Spaccapietra, S. (ed.) Journal on Data Semantics IV. LNCS, vol. 3730, pp. 146–171. Springer, Heidelberg (2005)
32. Santos Jr., P.S., Almeida, J.P.A., Guizzardi, G.: An ontology-based semantic foundation for organizational structure modeling in the aris method. In: 2010 14th IEEE International Enterprise Distributed Object Computing Conference Workshops (2010)
33. Pidcock, W.: What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model? (2002), `http://infogrid.org/wiki/Reference/PidcockArticle` (accessed April 2013)
34. Guizzardi, G.: Ontological foundations for structural conceptual models. PhD thesis, University of Twente, Enschede, The Netherlands (2005)
35. Rosemann, M., Green, P., Indulska, M.: A reference methodology for conducting ontological analyses. In: Atzeni, P., Chu, W., Lu, H., Zhou, S., Ling, T.-W. (eds.) ER 2004. LNCS, vol. 3288, pp. 110–121. Springer, Heidelberg (2004)
36. Bunge, M.A.: Treatise on Basic Philosophy. Ontology I - The Furniture of the World, vol. 3. Kluwer Academic Publishers (1977)
37. The Open Group: ArchiMate 2.0 specification (2012)
38. Horridge, M., Knublauch, H., Rector, A., Stevens, R., Wroe, C.: A practical guide to building owl ontologies using the protégé-owl plugin and co-ode tools edition 1.2, The University of Manchester (2009)

New-Generation Managers' Business/IT Alignment Perspective for a More Business-Driven IS Design

Jörg Hauke[1], Jörg Hans Mayer[2], Reiner Quick[1], and Hannes Feistenauer[1]

[1] Darmstadt University of Technology, Chair of Accounting and Auditing, Hochschulstrasse 1, 64289 Darmstadt, Germany
{joerg.hauke,feistenauer}@stud.tu-darmstadt.de,
quick@bwl.tu-darmstadt.de

[2] University of St.Gallen, Institute of Information Management, Müller-Friedberg-Strasse 8, 9000 St.Gallen, Switzerland
joerg.mayer@unisg.ch

Abstract. Digital natives are increasingly populating organizations' management. As they have higher expectations of IS accommodating their non-functional user preferences, the frontend application design of management support systems (MSS)—as managers' direct user-interface—plays an increasingly dominant role. By means of using new-generation managers' perspective of a business/IT alignment, this article proposes a method for a more business-driven IS design. We obtain both a set of business-driven MSS requirements and—applying this approach—initial design guidelines. We evaluate our approach with respect to the state of the art and propose avenues for future research.

Keywords: Corporate management, digital natives, management support systems (MSS), self-service IS, frontend application design, requirements analysis, literature review, principle of economic efficiency, multiple-case study.

1 Introduction

Henderson and Venkatraman [1] expose the term "*business/IT alignment*" to indicate that a company's IT is designed in terms of its business value. Business/IT alignment barely plays a role in management support systems (MSS) design [2, 3].

Besides its relatively stable content given by legal regulations and internal guidelines (functional MSS perspective), the present moment seems favorable for redesigning MSS from managers' *non-functional perspective*. First, digital natives in-creasingly populate organizations' management along with digital immigrants, who learned to engage with MSS over the years [4]. These new-generation managers more naturally accept IS, but have higher expectations about how MSS should accommodate their (new) user preferences. Second, technical progress has been made in recent years, so that even senior managers should be able to operate MSS themselves [5].

We focus on *MSS frontend applications*. They play a growing role in MSS design, as they are managers' direct IS user interface and evolved over the last 20 years from text fields, buttons, menus, and tiny icons into a palette of IS functionalities, such as

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 73–88, 2013.

exception reporting, collaboration, and context-sensitive "do more" functions [6, 7]. Simplifying IS handling, they currently have to provide information that is more at users' fingertips, and besides stationary IS use, they have to accommodate the growing importance of mobile IS use situations [8].

Applying a new-generation managers' business/IT alignment perspective helps to make MSS considerably easier to use and, thus, determine managers' IT acceptance [9]. The objective of this article is to propose a *method for using the potential of business/IT alignment* for a more business-driven IS design. We do so from a new-generation manager perspective and obtain both a set of business-driven MSS requirements and—applying this approach—initial design guidelines to start discussions.

We follow design science research (DSR) in IS [10] and apply Peffer et al.'s [11] six-step process model. We suggest that an appropriate frontend application design makes MSS easier to use (Sec. 1). After laying the foundations (Sec. 2), we propose a multidimensional framework for our *literature review* (Sec. 3). We describe the most important publications reflecting the state of the art and derive a research agenda (Sec. 4). As a rigorous starting point for MSS design, we propose a set of business-driven requirements (Sec. 5). We apply our proposal in a *case study* (Sec. 6) and arrive at initial design guidelines for MSS design. Our method is evaluated in terms of lessons learned contrasted against the state of the art (Sec. 7). The article concludes with a summary, limitations of our work, and avenues for future research (Sec. 8).

2 Importance of Frontend Applications in MSS Design

Both the terms MSS [12] and decision support systems (DSS, [13]) have been proposed as labels for IS intended to support managers. Since DSS evolved from a specific concept that complemented management information systems (MIS) and overlapped in the late 1980s with executive information systems (EIS), we refer to our object of study as *MSS*. This term covers MIS, DSS, EIS, and, more recently, knowledge management systems (KMS) and business intelligence (BI) systems for managers [14].

ISO 9241-110 [15] defines MSS as a combination of software and hardware components that receive input from managers and communicate output in order to support them in performing their tasks. *Frontend applications* "[…] are what users see and work with to use a product" [16]. They complement data warehouses (DW) and combine different software components to handle MSS in different use situations. We focus on MSS frontend application design for two reasons. First, IS use is *not* mandatory for managers and, in the worst case, they refuse to use MSS because of their position within the firm per se [17]. Second, the domain of MSS frontend applications is in a state of flux as vendors like SAP recently bought Business Objects (BO) in a USD 6.3 bn deal [18] to combine their DW and business application technology with smart frontend application design—and IBM did the same with Cognos (USD 5 bn) [19]. However, research on frontend application design is underrepresented in literature [20]. *Mobile* MSS are defined as IS offering "[…] all the necessary IS services when managers are traveling" [21].

3 Framework for Literature Systematization

A literature review should be concept-centric [22]. We structure our framework according to elements of IS design theories and the research approach used.

3.1 Elements of IS Design Theories

IS design theories consist of three elements [23]: (1) *User requirements* are prerequisites, conditions, or capabilities needed by users of IS to solve a problem or achieve an objective [24]. They delineate the specification of IS both from the functional and non-functional perspective [25]. Functional requirements address “what” IS are supposed to or must do (purpose). Non-functional requirements, in contrast, reflect “how well” IS performs within its environment fulfilling its function (quality) [26].

(2) *Guidelines* go beyond mere requirements. They serve as predefined actions that bring IS to life [27] and cover the span from a generic type such as Hevner et al’s Proposal [10] to more in-deepth IS specification we propose in this paper. Models outline IS features, or combinations of these [28]. Complementary methods describe the process of building MSS frontend applications [29].

(3) Managers’ IS use factors are gaining importance and MSS must broaden its scope beyond mere deployment [30]. Leveraging findings from HCI research [31], *IS design for use* is specified in two ways. An IS user analysis segments user groups and different user-group characteristics that influence how managers use MSS. Furthermore, the effects that frontend application design has on managers IS use is examined [32].

3.2 Research Approach

(4) Papers with a *behavioral focus* explain phenomena from practice and rely on observations as well as empirical methods. We differentiate between case studies to learn from singel implementations, experiments, surveys, and structural equation models (SEMs), e.g., IS success model [33]; technology acceptance model (TAM, [34, 35]); combined models [36].

(5) *Design approaches* involve ideas and frameworks for IS recommendations for creating a better world [23]. We distinguish between single items and broader list approaches, which specify sets of requirements, and frameworks that focus on the relationship between requirements and design guidelines.

4 Literature Analysis

4.1 Search Strategy

Following vom Brocke et al.’s four step process [37], we started with a journal search. We focused on leading IS research outlets that reflect their ranking and impact factor, provided by the London School of Economics [38].[1] Conducting a multidimensional review, we included journals from HCI[2] [39], computer science,[3] system and software engineering[4] [38], and added publications from ICIS and ECIS [40].

To access the journals we used EBSCOHost, ProQuest, the ACM Digital Library, and AIS Electronic Library. Our keyword search (Table 1) yielded a total of 383 hits. After qualifying them, we end up with 72 hits in total. A final back and forward search revealed *91 publications* to be relevant for our purpose.

Table 1. Keyword search string

		OR					
AND		Management support system	MSS	Executive information system	EIS	Business intelligence	Data warehouse
AND	OR	Dashboards	Frontends	User-interface	Manager	Mobile	Requirements
AND	OR	Antecedents	Criteria	Determinants	Elements	Evaluation	

4.2 Results: Current Shortcomings and Future Research Agenda

Fig. 1 systemizes the examined 91 publications identified as relevant within the framework we derived before (Sec. 3.1). The most revealing publications disclose three major gaps that should be overcome in future MSS frontend application design.

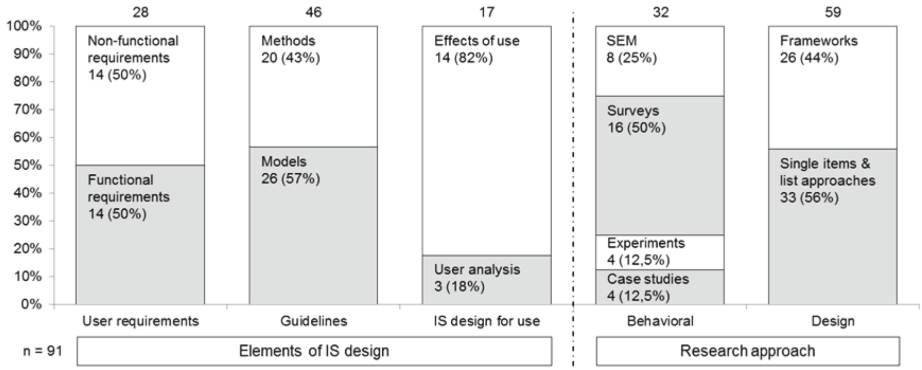


Fig. 1. Classification of the publications within our framework

---

1 This catalog incorporates not only mainstream IS journals, but also social studies. We chose the five top journals from each set, namely: MISQ, ISR, I&M, JMIS, and DSS, as well as EJIS, Information & Organization, ISJ, J. of Organizational and End User Computing, and JIS.

2 Int. J. of Human-computer Studies/Man-machine Studies, HCI, Int. J. of HCI, ACM-TCHI.

3 ACM Transactions on Computer Systems, IEEE Transactions on Computers, J. of Computer and System Sciences, JIT.

4 IEEE Transactions on Software Engineering, ACM Transactions on SE and Methodologies, J. of Systems and Software, IEEE Software, Information and Software Technology.

Focusing on the element of IS design we examined three shortcomings in the state of the art: *(1) User requirements: lack of sound requirements analysis:* 14 of 91 publications focus on functional requirements and an equal number on non-functional requirements. For example, Yigitbasioglu and Velcu [41] recommend that frontend applications ensure flexibility allowing users to switch between different presentation formats. Cheung and Babin [42], and Watson et al. [43] postulate quick IS adaptability towards different data sources. Kimball [44] states that frontend applications should allow “drill downs” to ask about KPI deviations, for example. Regarding “mobile,” most publications focus on layouts for desktop PCs [45]. However, Gebauer et al. [46] classify mobile use contexts in terms of the level of distraction, connection quality, and mobility and propose a single source-authoring approach to avoid manual adaption displayed on stationary and mobile devices.

However, the existing studies fail to provide an applicable set of requirements with a rigorous basis. The publications we reviewed either examine attributes of a single software component [44] or provide applicable lists of requirements without a rigorous basis [7, 47, 48]. The SEMs [33] are rigorous research methods, but their direct application faces obstacles such as the need for extensive surveys or they focus more on organizational impacts rather than on applicable criteria [49].

(2) *Guidelines: Lack of IS designs focusing on mobile use:* 46 of 91 publications cover either methods or models for designing MSS. Warmouth and Yen [50] structure IS software components into information presentation, dialog control, and analytical functions that managers can operate by themselves. Few [51] presents a design concentrating on the impact of different visual elements (e.g., bar charts).

Addressing the growing importance of mobile, Tarasewich et al. [52] identify several issues in this area: a changing context, limited user attention, users’ occupied hands, high mobility, and IS interaction while in motion. Gebauer and Shaw [53] refer to the value of mobile applications, as they allow a quick information retrieval even when “on the road.” Wixom and Watson [5] identify web browsers as a first step towards pervasive MSS, whereas Yuan et al. [54] state that apps promise new freedom to managers accessing and disseminating information beyond the desktop PC.

However, we found just one publication identifying information desired by managers in stationary situations, but not addressing situations that are in motion and, thus, an appropriate way to visualize such information at a glance.

(3) *IS design for use: lack “grasp” to apply in practice:* Various authors explain usage determinants or user satisfaction for different display types [55], functions (e.g., drill-down, filter) [50], reliability such as information access, and performance [56]. However, only 3 of 91 articles perform a user analysis. Articles on “IS design for use” remain at a generic level [57]. Mayer et al. [58] propose a configuration model accommodating the growing range of managers’ different working styles.

Summarizing our findings from literature review, (1) a ready-to-use set of business-driven requirements with a rigorous selection of criteria and its distinct clustering is lacking. Furthermore, (2) initial guidelines for new-generation managers’ MSS, especially focusing on an (3) IS design for use, are missing.

5 Alternative Method

5.1 Principle of Economic Efficiency

The principle of economic efficiency is a paradigm in business research which addresses the ratio between benefit and cost [59]. Thus, *business-driven* evaluation criteria for frontend applications are oriented towards what is economically feasible (positive benefit-cost ratio), and not what is conceptually or technically feasible.

Even if the IS costs by nature and amount can be identified, the ability to quantify IS value is limited [60]. We apply the "black-box method" [61] and differentiate between the basic criteria of IS output and IS input. *Solution capabilities* (IS output) refer to the relevance of MSS frontend applications to support managerial information needs. *Resource requirements* (IS input), in turn, refer to the input needed to generate the IS output, such as primary information and manpower in terms of cost and time.

5.2 First Level of Specification: Design Criteria

IS output is specified with various information and IS characteristics that are relevant to specific users. Based on our literature research (Sect. 4), we take the *user interface design* as our first design criterion. We propose that the quality of information presentation [62, 63] and an intuitive dialog control influence the "IS look&feel" [50].

We split the *functions* of frontend applications into five items. Various display formats are needed to present information at a glance and display even complex dependencies in an easy-to-understand manner [51, 64]. Specific requirements cover individual data manipulation on the frontend application layer such as financial rounding [65]. The more cooperative work increases [66], the more IS need to incorporate collaboration capabilities [67]. Frontend applications have to handle more and more managerial mobile use situations (Sec. 3). Other requirements include a user-concept for access rights, print functions, and advanced analysis [62]. A final criterion addressing the solution capabilities (IS output) is *information management.* Accommodating the growing range of manager working styles, use cases, and access modes [58], flexibility covers criteria to customize the frontend [41]. The degree of integration finally subsumes the capability to cover different data sources and output formats [42].

Resources required (IS input) are specified in terms of the *effort* to generate the IS output described before: cost adequacy and time adequacy.

5.3 Second Level of Specification: Evaluation Criteria

The design criteria remain at a general level. Thus, we specify them by means of evaluation criteria and complement them in our case study with the experience from new-generation managers (EC 13, 17). The findings are summarized in Table 2.

The quality of information presentation can be specified by the *graphical user interface design (GUI,* EC 1) measuring consistency and its "look& feel" [62, 68, 69]. Specifying the dialog control (Sec. 5.2), a *self-service user guidance* (EC 2) [9, 70, 71] with a consistent menu structure, an intuitive navigation, and *different types of dialog control* (EC 3) such as drill-functionalities, filter, and sorting mechanisms should enable further IS interactions, even for managers [44, 51, 72, 73].

We propose *spreadsheet capabilities* (EC 4) to provide a familiar environment for new-generation managers [64, 74]. Different *graphic types with features for interaction* (EC 5) allow freedom for information presentation [7, 41, 51]. Tooltips provide information on demand [41]. Frontend applications must offer *business concepts "out of the box"* such as economic value-added or portfolio techniques (EC 6) [75]. To gather information at a glance, frontends need to cover *exception reporting* (EC 7) [76, 77].

Table 2. Set of business-driven requirements for frontend application selection

Principle of economic efficiency	Design criteria		Evaluation criteria		Description
Solution capabilities (IS output)	User-interface design	Information presentation	EC 1	Graphical user interface (GUI) [62, 68, 69]	How is the first "look&feel" and is the basic screen design consistent?
		Dialog control	EC 2	Self-service user guidance [9, 70, 71]	How intuitive is the user guidance and is the menu logical and consistent?
			EC 3	Different types of dialog control [44, 51, 72, 73]	Are all drill-functionalities, filter, and sorting mechanisms supported?
	Functions	Various display formats	EC 4	Spreadsheet capabilities [64, 74]	How is the support of already known spreadsheet functionalities?
			EC 5	Graphic types with interaction [7, 41, 51]	How is the variety of visualizations and the IS interaction support?
			EC 6	Business concepts "out of the box" [75]	Are business concepts "out of the box" supported?
			EC 7	Exception reporting [76, 77]	Is it possible to define exceptions and to visualize them?
		Specific requirements	EC 8	Data manipulation layer [65, 78]*	Is there a layer between the DW and the frontend?
			EC 9	Copying hierarchies from the DW [78, 79]	Is it possible to copy already defined hierarchies from DW?
		Collaboration	EC 10	Comments [67, 80]	Is it possible to add comments to support collaboration across the company?
		Mobile	EC 11	Adapting (stationary) report designs to smart devices [8, 81]	How comfortable is the mobile support (e.g., report transformation for smart devices)?
			EC 12	Different information media [5, 54]	Are there different information media (PDF, web, app) available to fit characteristics of different smart devices?
		Other requirements	EC 13	Advanced print functions**	Is it possible to define printing areas (e.g., graphics, graphics and tables)?
			EC 14	Import of authorization concepts [82]	Is it possible to import permissions from DW? Is there a separate layer?
			EC 15	Data mining [62, 83]	Are data mining functionalities available to identify data patterns?
	Information management	Flexibility	EC 16	Customizing function for the GUI [8, 84]	How flexible is the frontend to accommodate individual working styles?
			EC 17	Programming language for add-ons*	Can individual add-ons be added?
		Integration	EC 18	Degree of integration [42, 85]	Are all required export and import formats supported?
Resource requirements (IS input)	Effort	Cost adequacy	EC 19	Cost adequacy [86]	What is the amount of money to be spent for implementation?
		Time adequacy	EC 20	Time adequacy [86]	How much time needs to be spent for implementation?

* requirement from BI Team; ** requirement from new-generation managers
The classification of the publications of this table is detailed in the appendix (Table A1).

Besides the *data manipulation layer* (EC 8) [65, 78], frontend applications must incorporate a function for *copying hierarchies from the DW* (EC 9) [78, 79]. *Comments* (EC 10) are another requirement for frontend applications [67, 80]. *Adapting (stationary) report designs to smart devices* (EC 11) [8, 81] and supporting *different information media* (EC 12) are important [5, 54]. *Advanced print functions* (EC 13) cover graphics or graphics and tables prints. An *import function for authorization concepts* (EC 14) to take over existing user-based rules for data access from the DW concept [82] and *data mining* (EC 15) can help to identify data patterns, leading to insights [62, 83].

Accommodating working styles, use cases, and access modes [58], a *customizing function for the GUI* (EC 16) is another requirement [8, 84] and—according to our case study—an easy-to-use *programming language for individual add-ons* (EC 17) is important. The *degree of integration* (EC 18) finally subsumes capabilities for covering different data sources and data output formats [42, 85].

Detailing the effort of frontend applications, we finally differentiate between *cost adequacy* (EC 19, e.g., license costs) and *time adequacy* (EC 20) [86].

We finally recommend using a method to apply weights for the different criteria, which reflect their importance. We considered an analytic hierarchy process which delivers distinct results, but it was too complex to handle [87]. We performed a *scoring model* with weights determined from the experience of new-generation managers which was much easier to handle and delivered quite distinct results (Fig. 2).

6 Demonstration and First MSS Frontend Application Design Guidelines

In order to lay out first MSS design guidelines, we applied our set of requirements at an automotive supplier. Separated into an automotive and rubber division, it had revenues amounting to USD 49.95 bn, and 164,000 employees at 200 production sites (2011).

The objective of the project was to enable managers to access their monthly group reporting by themselves and to establish a mobile MSS access—both with the help of a MSS frontend application. We specified our set by means of a description for each evaluation criterion (using the references we researched in Sec. 4.2) and introduced a five-point Likert scale to evaluate eight frontend applications from seven vendors.

From August to December 2012, two researchers (authors of this paper) and two experts from the group BI team of the company tested each application with two representatives of each vendor in a three-hour live-demonstration session. In doing so, the lines of Fig. 2 show the examined evaluation criteria, the columns cover the results from the evaluation ("as-is" profile). We recorded a to-be profile for MSS from new-generation managers perspective (Fig. 2) in order to measure the difference per criterion and calculated the median (as-is > to-be were set to zero). Implementation gaps (median of difference per criterion) of 2 and more were considered highly relevant and, thus, used to develop design guidelines for MSS frontend applications (Fig. 2, shaded in gray). Furthermore, the maximum observed deviation per line (Fig. 2, gray bars) indicates that solutions show large deficiencies in regard to evaluation criteria which are solved by most other applications (e.g., advanced print functions).

The result is a proposal which reflects MSS frontend applications that "best" match new-generation manager requirements. Starting with a business blueprint and a proof-of-concept in April 2013, this frontend will be implemented by summer 2013.

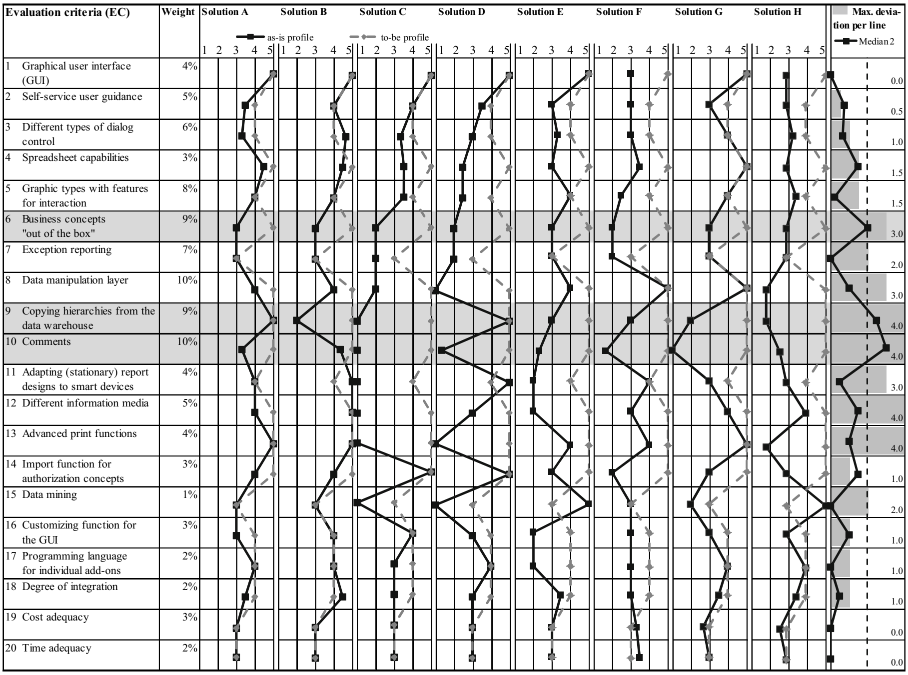


Fig. 2. Business-driven requirements for frontend application design (case example)

Three design guidelines for MSS frontend applications were established:

(1) Business concepts "out of the box" are missing (EC 6). In order to align business concepts with IT capabilities that new-generation managers expect from their alignment perspective, important business concepts should be available "out of the box." Currently, none of the evaluated applications supports concepts such as value-driver trees, so that they need to be developed manually. The changes associated with re-compositing such concepts entail an extensive adaption process.

(2) Copying data hierarchies from the DW needs improvement (EC 9). In order to avoid repetitive work at different IS layers, our case study suggests the need to copy business hierarchies from the DW to the frontend application. Furthermore, it would be convenient to manipulate some elements (e.g., change order) along the hierarchy.

(3) Comment function is inappropriate (EC 10). Business success can be improved by collaboration. Responding to increasing collaborative work, we recommend easing and accelerating collaboration by both enabling annotations at the report level for general discussions, and comments at the KPI level, in order to facilitate more detailed discussions.

7 Evaluation

Comparing our lessons learned with the state of the art, we reveal value propositions for our method as follows. The principle of economic efficiency enjoys broad acceptance in business-management research. As a reliable, frequently applied design principle, it is an accepted starting point for a *business-driven IS design*–especially for structuring a series of evaluation criteria. Compared with single-criteria approaches, the method on hand yields a set of requirements with a high level of completeness and distinctiveness. The approach is more hands-on, while SEMs are rigorous, but barely directly applicable. However, the impact of using a Likert scale is that most of the more complex statistical methods are not applicable. Another limitation is that the evaluation always entails some subjectivity.

The initial design guidelines should be valid as they are based on capabilities of leading MSS frontend applications. However, given that their development was not the main objective of the article, they are merely a starting point for future discussions.

For *research* purposes, the method on hand provides future MSS frontend application design with a rigorous starting point. The findings are applicable to other IS domains as well and thus contribute to improving IS design in general. In terms of business *practice*, the set of requirements and the design guidelines are useful to both software evaluation and serve as guidelines for application design. They can be applied at companies and software vendors wishing to improve their products.

8 Avenues for Future Research

Taking the design of MSS frontend applications as an example, this article proposes a new method for a more business-driven IS design. Using new-generation managers' business/IT alignment perspective, it covers both a rigorous set of business-driven requirements and initial MSS design guidelines to start discussions.

One avenue for future research is to complement the as-is profiles with results from live demonstrations, and the to-be profiles with findings from a multi-company case. Besides the examined software perspective of IS, managers' end-user device selection (hardware perspective) determines their IS acceptance. It should not be difficult to define a company's choice, but how to examine patterns for such a selection is likely to be more challenging. Furthermore, managers' IS use situations will gain importance for a proper IS design and their working styles should be captured more in detail. Gender, age, temperament, self-efficacy in IS knowledge, level of expertise, and prior IS experience might worth considering, as well as cultural factors. Finally, the functional perspective on MSS design should be specified in detail, especially whether there are changes due to the 2008/2009 economic crisis and the ongoing financial turbulences in Europe [88].

References

1. Henderson, J.C., Venkatraman, N.: Strategic Alignment: Leveraging IT for Transforming Organizations. IBM Syst. J. 32, 4–16 (1993)
2. Gluchowski, P., Gabriel, R., Dittmar, C.: Management Support Systems und Business Intelligence - Computergestützte Informationssysteme für Fach- und Führungskräfte. Springer, Heidelberg (2008)

3. Schulze, K.-D., Besbak, U., Dinter, B., Overmeyer, A., Schulz-Sacharow, C., Stenzel, E.: BI-Studie 2009. Steria Mummert Consulting AG, Hamburg (2009)
4. Vodanovich, S., Sundaram, D., Myers, M.: Digital Natives and Ubiquitous Information Systems. Information Systems Research 21, 711–723 (2010)
5. Wixom, B.H., Watson, H.J.: The BI-Based Organization. International Journal of Business Intelligence Research 1, 13–28 (2010)
6. Tidwell, J.: Designing Interfaces. O'Reilly, Köln (2005)
7. Eckerson, W., Hammond, M.: Visual Reporting and Analysis - Seeing Is Knowing. TDWI Best Practices Report. TDWI, Chatsworth (2011)
8. Mayer, J.H.: Using The Kano Model To Identify Attractive User-Interface Software Components. In: George, J.F. (ed.) Proceedings of the 20th Inter. Conf. on Information Systems (ICIS), Orlando, USA, vol. 314 (HCI), pp. 1–17 (2012)
9. Pijpers, G.G.M., Bemelmans, T.M.A., Heemstra, F.J., van Montfort, K.A.G.M.: Senior Executives' Use of Information Technology. Information and Software Technology 43, 959–971 (2001)
10. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28, 75–105 (2004)
11. Peffers, K., Tuunanen, T., Gengler, C.E., Rossi, M., Hui, W., Virtanen, V., Bragge, J.: The Design Science Research Process: A Model for Producing and Presenting Information Systems Research. In: 1st International Conference on Design Science in Information Systems and Technology, pp. 83–106. Claremont (2006)
12. Clark Jr., T.D., Jones, M.C., Armstrong Curtis, P.: The Dynamic Structure of Management Support Systems: Theory Development, Research Focus, and Direction. MIS Quarterly 31, 579–615 (2007)
13. Arnott, D., Pervan, G.: Eight Key Issues for the Decision Support Systems Discipline. Decision Support Systems 44, 657–672 (2008)
14. Carlsson, S.A., Henningsson, S., Hrastinski, S., Keller, C.: An Approach for Designing Management Support Systems: The Design Science Research Process and Its Outcomes. In: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology, pp. 1–10. Malvern (2009)
15. ISO: Ergonomics of Human-System Interaction—Part 110: Dialogue Principles (ISO 9241-110:2006), Berlin (2008)
16. Hackos, J.T., Redish, J.C.R.: User and Task Analysis for Interface Design. John Wiley & Sons, New York (1998)
17. Young, D., Watson, H.J.: Determinates of EIS Acceptance. Information & Management 29, 153–164 (1995)
18. SAP AG, `http://www.sap.com/corporate-en/press.epx?pressid=8360`
19. IBM, `http://www-03.ibm.com/press/us/en/pressrelease/22572.wss`
20. Sumita, U., Yoshii, J.: Enhancement of E-commerce via Mobile Accesses to the Internet. Electronic Commerce Research and Applications 9, 217–227 (2010)
21. Puuronen, S., Savolainen, V.: Mobile Information Systems—Executives' View. Information Systems Journal 7, 3–20 (1997)
22. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly 26, 13–23 (2002)
23. Walls, J.G., Widmeyer, G.R., El Sawy, O.A.: Building an Information System Design Theory for Vigilant EIS. Information Systems Research 3, 36–59 (1992)
24. Kotonya, G., Sommerville, I.: Requirements Engineering: Processes and Techniques. John Wiley & Sons, New York (1998)
25. Sommerville, I.: Software Engineering. Pearson, München (2010)

26. Paech, B., Kerkow, D.: Non-functional Requirements Engineering - Quality Is Essential. In: Regnell, B., Kamsties, E., Gervasi, V. (eds.) 10th Requirements Engineering: Foundation for Software Quality (REFSQ), pp. 237–250. Universität Duisburg-Essen, Riga (2004)
27. Hoogervorst, J.A.P.: Enterprise Governance and Enterprise Engineering. Springer, Diemen (2009)
28. Gregor, S.: The Nature of Theory in Information Systems. MIS Quarterly 30, 611–642 (2006)
29. March, S.T., Smith, G.F.: Design and Natural Science Research on Information Technology. Decision Support Systems 15, 251–266 (1995)
30. Marchand, D.A., Peppard, J.: Designed to Fail: Why IT Projects Underachieve and What to Do about It. IMD International, Lausanne (2008)
31. Zhang, P., Benbasat, I., Carey, J., Davis, F., Galletta, D.: Human-Computer Interaction Research in the MIS Discipline. Communications of the Association for Information Systems 9, 334–355 (2002)
32. Benbasat, I., Nault, B.R.: An Evaluation of Empirical Research in Managerial Support Systems. Decision Support Systems 6, 203–226 (1990)
33. DeLone, W.H., McLean, E.R.: The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. Journal of Management Information Systems 19, 9–30 (2003)
34. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quartely 13, 318–340 (1989)
35. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 27, 425–478 (2003)
36. Wixom, B.H., Todd, P.A.: A Theoretical Integration of User Satisfaction and Technology Acceptance. Information Systems Research 16, 85–102 (2005)
37. Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In: 17th European Conference on Information Systems, Verona, pp. 2206–2217 (2009)
38. Willcocks, L.P., Whitley, E.A., Avgerou, C.: The Ranking of Top IS Journals: A Perspective from the London School of Economics. European Journal of Information Systems 17, 163–168 (2008)
39. AIS, `http://ais.affiniscape.com/displaycommon.cfm?an=1&subarticlenbr=432`
40. WKWI: WI-Liste der Konferenzen, Proceedings und Lecture Notes. Wirtschaftsinformatik 50, 155–163 (2008)
41. Yigitbasioglu, O.M., Velcu, O.: A Review of Dashboards in Performance Management: Implications for Design and Research. International Journal of Accounting Information Systems 13, 41–59 (2012)
42. Cheung, W., Babin, G.: A Metadatabase-Enabled Executive Information System (Part A): A Flexible and Adaptable Architecture. Decision Support Systems 42, 1589–1598 (2006)
43. Watson, H.J., Watson, R.T., Singh, S., Holmes, D.T.: Development Practices of EIS: Findings of a Field Study. Decision Support Systems 14, 171–184 (1995)
44. Kimball, R.: Drill Down to Ask Why, Part 1. DM Review 18, 6 (2008)
45. Wu, C.-S., Cheng, F.-F., Yen, D.C., Huang, Y.-W.: User Acceptance of Wireless Technology in Organizations: A Comparison of Alternative Models. Computer Standards & Interfaces 33, 50–58 (2011)

46. Gebauer, J., Shaw, M.J., Gribbins, M.L.: Task-Technology Fit for Mobile Information Systems. Journal of Information Technology (Palgrave Macmillan) 25, 259–272 (2010)
47. Few, S.: Dashboard Design: Beyond Meters, Gauges, and Traffic Lights. Business Intelligence Journal 10, 18–24 (2005)
48. Papageorgiou, E., de Bruyn, H.: Creating Strategic Value through Executive Information Systems: An Exploratory Study. Electronic Journal of Information Systems Evaluation 13, 57–76 (2010)
49. Mayer, J.H., Marx, F.: Systematic Development of Business-Driven Requirements – Using Next-Generation EIS Design as an Example. In: Winter, R., Zhao, J.L., Aier, S. (eds.) DESRIST 2010. LNCS, vol. 6105, pp. 506–521. Springer, Heidelberg (2010)
50. Warmouth, M.T., Yen, D.: A Detailed Analysis of Executive Information Systems. International Journal of Information Management 12, 192–208 (1992)
51. Few, S.: Information Dashboard Design: The Effective Visual Communication of Data. O'Reilly Media, North Sebastopol (2006)
52. Tarasewich, P., Gong, J., Nah, F.F.-H., Dewester, D.: Mobile Interaction Design: Integrating Individual and Organizational Perspectives. Information KMS 7, 121–144 (2008)
53. Gebauer, J., Shaw, M.J.: Success Factors and Impacts of Mobile Business Applications: Results from a Mobile E-procurement Study. International Journal of Electronic Commerce 8, 19–41 (2004)
54. Yuan, Y., Archer, N., Connelly, C.E., Zheng, W.: Identifying the Ideal Fit between Mobile Work and Mobile Work Support. Information & Management 47, 125–137 (2010)
55. Benbasat, I., Dexter, A.S., Todd, P.: An Experimental Program Investigating Color-Enhanced and Graphical Information Presentation: An Integration of the Findings. Communications of the ACM 29, 1094–1105 (1986)
56. Chatterjee, S., Chakraborty, S., Sarker, S., Sarker, S., Lau, F.Y.: Examining the Success Factors for Mobile Work in Healthcare: A Deductive Study. Decision Support Systems 46, 620–633 (2009)
57. Winter, R.: Design of Situational Artefacts—Conceptual Foundations and Their Application to IT/Business Alignment. In: Pokorny, J., Repa, V., Richta, K., Wojtkowski, W., Linger, H., Barry, C., Lang, M. (eds.) Information Systems Development, pp. 35–49. Springer, New York (2011)
58. Mayer, J.H., Winter, R., Mohr, T.: Situational Management Support Systems. Business & Information Systems Engineering 4, 331–345 (2012)
59. Samuelson, P.A.: Foundations of Economic Analysis. Harvard Press, Cambridge (1983)
60. Patas, J., Bartenschlager, J., Goeken, M.: RBV in Empirical IT Business Value Research–An Evidence-Based Literature Review. In: Proceedings of the 2012 45th Hawaii International Conference on System Sciences, pp. 5062–5071. IEEE Computer Society (2012)
61. Todd, P., Benbasat, I.: Process Tracing Methods in Decision Support Systems Research: Exploring the Black Box. MIS Quarterly 11, 493–512 (1987)
62. March, S.T., Hevner, A.R.: Integrated Decision Support Systems: A Data Warehousing Perspective. Decision Support Systems 43, 1031–1043 (2007)
63. Hung, S.-Y., Wu, H.-L.: Factors Influencing User Acceptance of Web-Based Decision Support Systems. Journal of Computer Information Systems 52, 70–77 (2012)
64. Eckerson, W.W.: Performance Dashboards: Measuring, Monitoring, and Managing Your Business. John Wiley & Sons, Hoboken (2010)
65. Power, D.J., Sharda, R.: Model-Driven Decision Support Systems: Concepts and Research Directions. Decision Support Systems 43, 1044–1061 (2007)

66. Rosen, E.: The Culture of Collaboration: Maximizing Time, Talent and Tools to Create Value in the Global Economy. Red Ape Pub., San Francisco (2007)
67. Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R., Carlsson, C.: Past, Present, and Future of Decision Support Technology. Decision Support Systems 33, 111–126 (2002)
68. Howson, C.: Ease of Use and Interface Appeal in Business Intelligence Tools. Beye Network, Boulder (2010)
69. Sankar, C.S., Nelson Ford, F., Bauer, M.: A DSS User Interface Model to Provide Consistency and Adaptability. Decision Support Systems 13, 93–104 (1995)
70. Ziff Davis Enterprise: Reaping the benefits of next generation dashboards. Beye Network, Newton (2009)
71. Donlon, B.S.: Designing Next-Generation Dashboards. DM Review 17, 39–39 (2007)
72. Chi, R.T., Turban, E.: Distributed Intelligent Executive Information Systems. Decision Support Systems 14, 117–130 (1995)
73. Salmeron, J.L.: EIS Data: Findings from an Evolutionary Study. Journal of Systems and Software 64, 111–114 (2002)
74. Fest, G.: The Future of Spreadsheets in BI? Bank Technology News 20, 33–33 (2007)
75. Galloway, D.L.: Achieving Accurate Metrics Using Balanced Scorecards and Dashboards. Performance Improvement 49, 38–45 (2010)
76. Wu, J.: Harnessing the Power of Exception Reporting. DM Review 15, 16–68 (2005)
77. Houghton, R., El Sawy, O., Gray, P., Donegan, C., Joshi, A.: Vigilant Information Systems for Managing Enterprises in Dynamic Supply Chains: Real-Time Dashboards at Western Digital. MIS Quarterly Executive 3, 19–35 (2004)
78. Schneider, M.: A General Model for the Design of Data Warehouses. International Journal of Production Economics 112, 309–325 (2008)
79. Talwar, K., Gosain, A.: Hierarchy Classification for Data Warehouse: A Survey. Procedia Technology 6, 460–468 (2012)
80. Parr Rud, O.: Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy. John Wiley & Sons, Hoboken (2009)
81. Andersson, B., Henningsson, S.: Developing Mobile Information Systems: Managing Additional Aspects. In: ECIS 2010 Proceedings. Pretoria (2010)
82. Fernández-Medina, E., Trujillo, J., Villarroel, R., Piattini, M.: Access Control and Audit Model For the Multidimensional Modeling of Data Warehouses. Decision Support Systems 42, 1270–1289 (2006)
83. Kantardzic, M.: Data Mining: Concepts, Models, Methods, and Algorithms. John Wiley & Sons, Hoboken (2011)
84. Gebauer, J., Schober, F.: Information System Flexibility and the Cost Efficiency of Business Processes. Journal of the Association for Information Systems 7, 122–146 (2006)
85. Wixom, B.H., Watson, H.J.: An Empirical Investigation of the Factors Affecting Data Warehousing Success. MIS Quarterly 25, 17–41 (2001)
86. Schober, F., Gebauer, J.: How Much to Spend on Flexibility? Determining the Value of Information System Flexibility. Decision Support Systems 51, 638–647 (2011)
87. Saaty, T.L.: The Analytic Hierarchy Process. McGraw-Hill, New York (1980)
88. Mayer, J.H.: Current Changes in Executive Work and How to Handle Them by Redesigning Executive Information Systems. In: Mancini, D., Vaassen, E., Dameri, R.P. (eds.) Accounting Information Systems for Decision Making, pp. 134–155. Springer, Heidelberg (2012)

Appendix

Table A1. Classification of publications utilized for the development of the business-driven requirements in Table 2

No.	Author(s)	Year	Title	Publication	Elements of IS design	Research approach
[5]	Wixom, B., and Watson, H. J.	2010	The BI-Based Organization	International Journal of BI Research	Model	Framework
[7]	Eckerson, W., and Hammond, Mark	2011	Visual Reporting and Analysis	TDWI Best Practices Report	Method	Survey
[8]	Mayer, J. H.	2012	Using the Kano Model to Identify Attractive User Interface Software Components	International Conference on Information Systems	Functional requirements	Case study
[9]	Pijpers, G. G. M. et al.	2001	Senior Executives' Use of Information Technology	Information and Software Technology	Model	Structural equation model
[41]	Yigitbasioglu, O. M., and Velcu, O.	2012	A Review of Dashboards In Performance Management: Implications for Design and Research	International Journal of Accounting Information Systems	Non-functional requirements	Single items & list approach
[42]	Cheung, W., and Babin, G.	2006	A Metadatabase-Enabled EIS (Part A): A Flexible and Adaptable Architecture	Decision Support Systems	Functional requirements	Single items & list approach
[44]	Kimball, R.	2008	Drill Down to Ask Why	DM Review	Functional requirements	Single items & list approach
[51]	Few, S.	2006	Information Dashboard Design	Book	Non-functional requirements	Single items & list approach
[54]	Yuan, Y. et al.	2010	Identifying the Ideal Fit Between Mobile Work and Work Support	Information & Management	Model	Framework
[62]	March, S. T., and Hevner, A. R.	2007	Integrated Decision Support Systems: A DW Perspective	Decision Support Systems	Functional requirements	Framework
[64]	Eckerson, W.	2010	Performance Dashboards: Measuring, Monitoring, and Managing Your Business	Book	Method	Single items & list approach
[65]	Power, D. J., and Sharda, R.	2007	Model-Driven Decision Support Systems: Concepts and Research Directions	Decision Support Systems	Model	Framework
[67]	Shim, J. P. et al.	2002	Past, Present, and Future of Decision Support Technology	Decision Support Systems	Method	Single items & list approach
[68]	Howson, C.	2010	Ease of Use and Interface Appeal in Business Intelligence Tools	Beye Network	Non-functional requirements	Single items & list approach
[69]	Sankar, C. S., Ford, N., and Bauer, M.	1995	A DSS User Interface Model to Provide Consistency and Adaptability	Decision Support Systems	Model	Single items & list approach
[70]	Ziff, Davis	2009	Reaping the Benefits of Next Generation Dashboards	Ziff Davis Enterprise	Non-functional requirements	Single items & list approach
[71]	Donlon, B.	2007	Designing Next-Generation Dashboards	DM Review	Non-functional requirements	Single items & list approach
[72]	Chi, R. T., and Turban E.	1995	Distributed Intelligent Executive Information Systems	Decision Support Systems	Model	Framework
[73]	Salmeron, J. L.	2002	EIS Data: Findings From An Evolutionary Study	The Journal of IS and Software	Functional requirements	Single items & list approach
[74]	Fest, G.	2007	The Future of Spreadsheets in BI?	Bank Technology News	Functional requirements	Single items & list approach
[75]	Galloway, D. L.	2010	Achieving Accurate Metrics Using BSC and Dashboards	Performance Improvement	Functional requirements	Single items & list approach

Table A1. (*continued*)

[76]	Wu, J.	2005	Harnessing the Power of Exception Reporting	DM Review	Functional requirements	Single items & list approach
[77]	Houghton, R. et al.	2004	Vigilant Information Systems for Managing Enterprises in Dynamic Supply Chains: Realtime Dash-boards At Western Digital	MIS Quarterly Executive	Method	Single items & list approach
[78]	Schneider, M.	2008	A General Model for the Design of Data Warehouses	International Journal Production	Model	Framework
[79]	Talwar, K., and Gosain, A.	2012	Hierarchy Classification for Data Warehouse: A Survey	Procidea Technology	Functional requirements	Single items & list approach
[80]	Parr Rud, O.	2009	BI Success Factors	Book	Model	Framework
[81]	Andersson, B., Henningsson, S.	2010	Mobile IS: Managing Additional Aspects	European Conference on Information Systems	Model	Framework
[82]	Fernández-Medina, E. et al.	2006	Access Control for the Multidimen-sional Modeling of DW	Decision Support Systems	Functional requirements	Single items & list approach
[83]	Kantardzic, M.	2011	Data Mining: Concepts, Models, Methods, and Algorithms	Book	Method	Single items & list approach
[84]	Gebauer, J., and Schober, F.	2006	Information System Flexibility and the Cost Efficiency of Processes	Journal of the Association for IS	Model	Framework
[85]	Wixom, B., and Watson, H. J.	2001	An Empirical Investigation of the Factors Affecting DW Success	MIS Quarterly	Method	Survey
[86]	Schober, F.	2011	How Much to Spend on Flexibility? Determining the Value of IS Flexibility	Decision Support Systems	Model	Experiment

Policy Conflict Handling as a Monitoring Activity of Hospital Information Systems

Raik Kuhlisch[1] and Kurt Sandkuhl[2]

[1] Fraunhofer Institute for Open Communication Systems FOKUS
Kaiserin-Augusta-Allee 31, 10589 Berlin, Germany
raik.kuhlisch@fokus.fraunhofer.de

[2] University of Rostock, Institute of Computer Science
Albert-Einstein-Str. 22, 18059 Rostock, Germany
kurt.sandkuhl@uni-rostock.de

Abstract. Alignment of business and IT is a serious challenge in enterprises due to continuously changing business environments and at the same time changing organizational IT infrastructures. The same challenges can be detected in health information technology accompanied by domain-specific information security demands regarding the access to patient-related information and medical data. The paper addresses a specific aspect in this area, which is of high relevance for business and IT alignment: how to define and apply policies as means to translate organizational requirements into guidelines and rules in IT management. The scope of the paper is limited to hospital information systems and policies in information security. The main contributions of this paper are (1) to present a case study from hospital information security confirming the need for supporting policy implementation, (2) to identify and describe the problem of policy conflict management as part of IT and business alignment, and (3) to define the research design for addressing this problem from a design science perspective.

Keywords: E-health, hospital intra-enterprise policy conflict, policy compliance verification, information security, design science, case study.

1 Introduction

Alignment of business and information technology (IT) often is considered as serious challenge in enterprises since the business environment continuously changes and so does the IT in an enterprise, but the pace of change and the time frame needed to implement changes is different in both areas. The reasons for this are manifold [1]: legal aspects, regulations, business requirements, economic factors, etc. Technological trends such as Software as a Service, virtualization, cloud computing or service-oriented architectures are influencing the way in which IT services are rendered [2]. Model-based approaches and IT governance are potential candidates to bridge evolving business contexts and IT, in order to adapt the provisioning of IT for business needs [3]. Additionally, business professionals and IT professionals often have different backgrounds, use

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 89–99, 2013.

different terminologies and set different priorities. Reaching an agreement about how to set priorities for the enterprise in this context is difficult and adds one more level of complexity to the task of business and IT alignment.

The above challenges can also be found in healthcare. The health information technology (HIT) in a hospital is multifaceted and can be divided into two main parts, the clinical information systems and administrative information systems [4]. Clinical information systems primarily serve as support and documentation of the medical treatment process, including specialized systems for, e.g., radiology, care, laboratory or medical imaging. The administrative systems include patient data management, accounting, resource planning, asset management, and management information systems. This complexity from a systems' point of view is accompanied by a variety of user groups and demands regarding the access to patient-related information and medical data. The managing of the involved information systems has become a need that should not be underestimated since organizational problems and change management aspects expose technical problems from before the 1990s [5].

The paper addresses a specific aspect in health information technology, which is of high relevance for business and IT alignment: how to define and apply policies as means to translate organizational requirements into guidelines and rules in IT management. The scope of the paper is limited to hospital information systems and policies in information security. This area was selected since many laws and – partly conflicting – regulations exist, which make development and implementation of a conflict-free set of policies a conceptual and organizational challenge. As a support for implementation and monitoring of changing policies in organizations, we propose the development of a conflict model aligning the given security policies, hospital-internal compliance rules, and privacy aspects of patients. This conflict model in combination with adequate methods for detection and management of potential conflicts is supposed to be an instrument for IT management. The main contributions of this paper are (1) a case study from hospital information security confirming the need for supporting policy implementation, (2) to identify and describe the problem of policy conflict management as part of IT and business alignment, and (3) to define the research design for addressing this problem from a design science perspective. The work presented in this paper is research in progress.

The remainder of this paper is structured as follows. Section 2 gives an overview to hospital information management as a background for this paper. In Section 3, we discuss the research strategy for gathering new information in this area and the development of solutions for the problem. Section 4 presents the case study in a large hospital in Germany and the experiences made. Section 5 discusses the planned steps to find solutions for the problem described above. Section 6 summarizes this paper.

2 Hospital Information Management

A good practice to guide the overall information management of hospitals is to prepare a strategic information management plan that breaks down the hospital information processing into the three typical management disciplines, namely strategic, tactical, and operational information management, whereas the strategic information management considers the overall picture of a hospital information system (HIS). Individual HIS components that are developed and integrated by means of projects are assigned as tactical. The latter discipline refers to the actual operating of a HIS. Each level is accompanied by tasks that plan, control and monitor related activities to the HIS. The information management plan contains, among other things, principles and standards to be applied for information management in an organization [6]. Figure 1 provides a brief overview of typical information systems that are part of a HIS and integrated with a communication server which is not unusual in Germany.

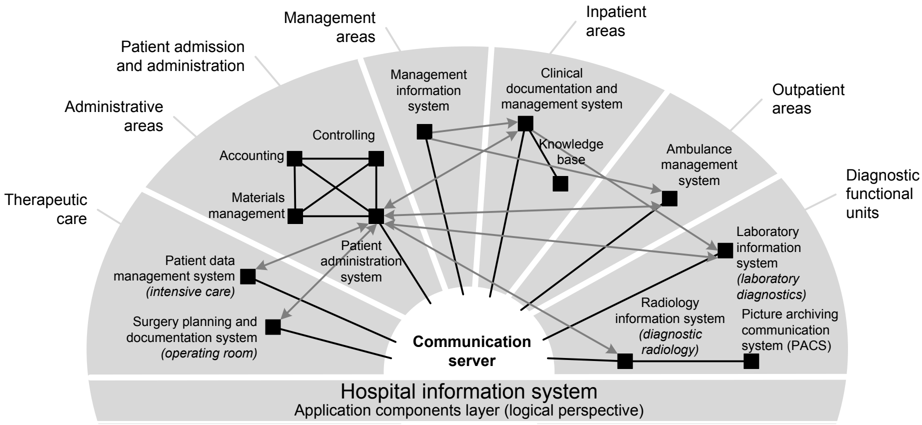


Fig. 1. Typical components of a HIS integrated via a communication server (Note: Black lines indicate a physical connection between application components while gray arrows represent cross-functional, logical data flows among the application components)

Access to electronic medical records (EMR) is realized by IT services. Information technology service management (ITSM) focusses on customers which consume reliable IT services. ITSM is an appropriate method for controlling and monitoring the HIS. It might benefit from formal guidance since rules for information security to be applied to the HIS are directly extractable or at least derivable from the information management plan. Policies are the formal container for such rules.

According to the definition in [7] policies are the result of implemented management goals, that help to define a desired (system) behavior. The general notion of a policy is to regulate the system behavior by solely applying the respective policy instead of re-implementing the system or its features. Contrary, policies might simply be applied to management or system domains in order to enforce requirements. This reveals the versatility of policies what is observable in hospitals, too. This paper addresses compliance and access control rather than network security as the two main applications in security management.

Nowadays, patient data is primarily processed by computer-based information systems. Any data access features a multitude of legal and regulatory constraints which are subsumed under the term compliance. These concerns consist of security policies, hospital-internal compliance rules, and privacy aspects of patients and contribute to requirements of data protection and data security in information processing systems.

The "need-to-know" principle integrates/focusses such concerns. This principle describes the minimal access to information granted to the staff in order to fulfill necessary tasks for patient treatment [8]. In practice, this principle is split into a need-to-use and mediates-access relationship (i.e., medical staff depend on applications or systems that enable the access to medical data or EMR). Without the implementation of rules behind this principle, physicians would be bound to screen all the documents of a patient (patient history, closed cases etc.) in order to make a medical decision. This approach would lead to infeasible and non-sustainable efforts.

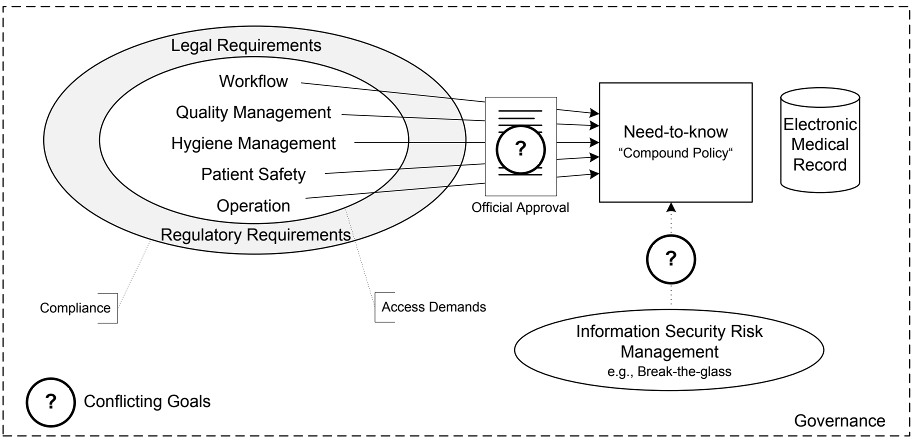


Fig. 2. Interfaces with conflicting goals with regard to the need-to-know principle

Nevertheless, need-to-know must be officially approved (e.g., by means of organizational privacy policies) to prevent unauthorized access. Exceptions arise from the information security risk management that balances need-to-know according higher pillars of information security: For instance, in order to realize patient safety with regard to emergency situations it is a common practice to implement "break-the-glass" activations and break the security concept "confidentiality" accordingly. Abuses cannot be prevented, but must be traceable and in an accepted range. However, goals of the organization of labor and risk management may conflict each other.

In addition, patients may restrict default permissions of existing authorization policies or privacy policies in consents, and thus rule their disclosure of data sovereignly (e.g., implemented by means of the "Basic Patient Privacy Consents"

(BPPC) [9] developed by IHE—an initiative that procures the improvement of the way computer systems in healthcare share information). Consequently, the access to patient data is characterized by an intersection of multiple privacy or security-related policies that are applied at the same time. These are legal directives, organizational privacy policies, and constraints stated by patients in their consent (i.e., patient privacy policy). The highly heterogeneous protection objects, priorities, and specific concerns of those policy types may cause processing conflicts when applied by a system (cf. Fig. 2). This holds for the design time as well at the runtime of policies.

3 Research Approach

Information systems (IS) are the link between IT and organizations and also a discipline of applied research in which interdisciplinary theories are applied. Design is a usable and valuable research methodology respected by engineering disciplines, too [10]. It is to be noted that research is a problem driven formal work and should not be methodology driven [11]. Hence, the used methods should be judged accordingly. In the IS research the two complementary paradigms behavioral science and design science are distinguished [12]. Design science fits very well because its goal of the development of a technology-based solution for open business needs matches excellent to the problem space outlined in the introduction.

To gain more powerful insights in the business needs within the triangle people, organizations, and technology, behavioral science provides valuable contributions. Behavioral science aims at developing and justifying theories that help understand complex interrelationships that relate to the business needs. This knowledge ideally allows predictions of organizational and human phenomena (in the context of information systems as well). It is derived from natural science and social science. In contrast, design science is basically a discipline to find solutions for business needs or IS-related problems. In this way, design science works on the description of a design problem, the iterative development of a solution and its evaluation [12].

The research approach used for this work is design science. [13] provides a detailed description of what design science deals with: Design science serves as a blueprint for researchers on how to build and evaluate design artifacts in technology-driven organizations. These two processes are complemented by design or IT artifacts that deliver guidance to heretofore unsolved organizational problems. Such innovations might be constructs, models, methods, and instantiations, or some combination of such. Constructs provide the vocabulary and symbols that enables IT researchers and practitioners to communicate both the problem and the solution accordingly. Models use the vocabulary and simplify the real world situation by means of abstract representations of relevant aspects of the identified problem. Among methods are hidden algorithms and practices that enable stakeholders to investigate the problem and to seek for solution options. Finally, instantiations demonstrate practicability of the three above-named artifacts in an IT system through their simultaneous application.

For the successful application of design-science research guidelines are proposed [13]. If a researcher follows these guidelines, the utility that the IT artifact provides could be approved. The guidelines deliver information on: problem relevance, research rigor, design as a search process, design as an artifact, design evaluation, research contributions, and research communication. These guidelines indicate underlying processes to be run through.

In the context of this work, the expected outcome of design science is, on the one hand, a description of the gap between actual and the desired status in the compensation of security-related policies for patient data accesses. On the other hand, the developed IT artifact gives guidance for stakeholders in hospital's IT departments on how to achieve the desired status (i.e., to spot and prevent conflicts between policies in the healthcare domain).

4 Case Study: Information Management in a German Hospital

In order to demonstrate the relevance and importance of the aforesaid problem areas in policy-based accesses to medical data, the environment of the security policy as well as practices and relevant IT application systems need to be investigated further. This allows for a proper assessment of (working) processes and defines an actual state.

Hence, a qualitative data collection should light on the introductory description of the business problem and investigate the real life context. For this purpose, a case study was conducted in a maximum care hospital in Munich, Germany. The case strategy fits well to IS research since organizational issues are more of interest than technical ones [14] and it is accepted as an established research strategy within the IS research community [15].

4.1 Subject of Investigation

The case study was primarily concerned with the data access by healthcare professionals to personal medical data of patients treated in the hospital (i.e., primary care). The case study distinguished from paper files so that only electronically stored patient data and associated security policies were considered.

Use Cases. To obtain a representative result a cross-section through various service facilities of the hospital should be formed. Therefore, requests for medical information on the admission, treatment and discharge of a patient are of interest.

Emergency scenarios were excluded from the case study because it is assumed that in such situations no standardized requests can be guaranteed. In addition, special types of medical activity (i.e., attending and company physicians) were not considered, as they are based on a differentiated authorization concept: Partially, patient records must be kept separate in their own responsibility of the files of the hospital administration. Also the external, inter-enterprise processing of medical data according to the German Federal Data Protection Act was not discussed, since the passing of computer-stored personal data to others relates to information systems outside the hospital.

4.2 Case Study Design

This section gives a description of the procedures used for conducting the case study. Initially, characteristics of the case study are presented.

Classification. According to the classifications outlined in [16] the case study's objective is primarily exploratory since we were seeking new insights in the variety of security policies or rather the access to patient data in hospitals. For more insights into the contextual understanding of conflicting security policies mainly qualitative data should be used what would lead to the fact that no generalization of the findings can be done. The focus is laid on an intensive analysis of the above mentioned use cases. Recapping, an inductive approach rather than theory testing was pursued with this case study.

Data Collection Methods. Techniques for data collection are distinguished in the three categories of "directly", "independent" and "indirectly" [17]. From the first two categories the following techniques were intended to be used for gathering the data:

- Direct: interviews, polls, on-site observations
- Independent: organizational and technical documentation analysis.

Polls have the disadvantage that it is not sure who actually fills out the survey and to what extent all the questions are properly understood. For this reason, interviews should be carried out primarily.

Sample Design. As mentioned before, interviews should be conducted with subject matter experts. The interviewees were the data protection officer and the chief information officer. The two interviews were conducted by means of a structured interview guide—the interviews itself were unstructured to elicit information to the fullest. Interview questions to ask were dedicated to the following topics: corporate governance and compliance, labor organization, IT infrastructure, and implemented security measures. Beside the interviews, the data collection is accompanied by provided documents. These documents are a current snapshot of the access matrix and a patient consent form.

4.3 Analysis of the Data

The objective of the analysis was to document relevant policies that have a direct impact on authorized access to patient data. Hence, the answers of the interviews are analyzed whether the access to medical data is somehow policy-driven. More precisely, it is of interest whether access policies are derived from abstract (organizational security) policies. On this assumption, a policy hierarchy must be in place.

An outcome of the interviews as well as the provided data is that the access to patient data is based on other principles that deviate from the management activities outlined in Section 2. The concept can be described in brief as follows: Departments with their special field of work are internally mapped to nation-wide codes published by the National Association of Statutory Health Insurance Physicians. These codes are assigned structural roles which represent permissions to the EMR. The medical staff is assigned such role in a requirements-driven manner. That is, if an access by medical staff is required, someone has to make an application for a special structural role at the human resource department. After verification of his professional qualification the role will be assigned. The IT department configures necessary roles (i.e., access groups) to internal organization units, which happens rather rarely. After a defined deadline, together with the data protection officer, the IT department checks unnecessary authorizations and removes them where applicable.

An anomaly exists with radiological data and imagery from the radiology information system (RIS). The RIS images are made available via the EMR. However, they make the connection to the picture archiving and communication system (PACS) via the communication standard DICOM and confront the PACS as a consumer. Privileges for internal use of the RIS and PACS are configured separately. The general idea in mind is that the requests should run all over the EMR. Otherwise the question is always how to ensure that the requests from internal and external point are exactly the same view of the file.

4.4 Findings

Although being a maximum care hospital, the interviews yielded that no formal framework for information (security) management has been set up. The access to medical data is constrained through a central hospital EMR that is linked to the entire subsystems of the HIS. This EMR is protected by a subject-specific access right concept. Conflicts with regard to EMR accesses may arise when the patient disagrees with special access groups. Policy conflicts are hard to identify since policies that define why access groups are authorized are intrinsically not existent. The access matrix serves for a lookup table for currently given access rights. However, an organizational privacy policy needs to be in place that justifies such configured authorizations. This policy type in conjunction with legal/regulatory policies and patient consents allows for a proper conflict analysis. A feasible way seems to constitute such policy with good-practices recommendations.

5 Conclusion from the Case Study

The case study revealed that the hospital lacks formal organizational (privacy) policies, which might lead to unauthorized access since no official approval for need-to-know is present. Moreover, conflicts between other policy types cannot be precluded. This does not, however, mean that compliance regulations are not present but exist implicitly within the organization of labor. It's understood that this phenomena cannot be generalized but ad-hoc authorizations seem to be a common procedure in clinical practice. To provide utility for monitoring activities of the HIS, policy compliance verification is considered indispensable.

Based on the findings of the case study that indicate problem relevance the next steps of the design science guidelines are applied.

5.1 Design as a Search Process

The questions that must be cleared up in this phase of design science regard issues laws, means and ends of the solution or IT artifact. Laws are constants or forces in the environment, means stand for actions to take and resources to use in order to construct the solution whereas ends represent satisfied goals, utility or constraints on the solution [13].

An IT artifact should provide guidance for a conflict-free compensation of security-related policies. This goal is achieved amongst other things by a step-by-step analysis of the involved policies. For simplification a recently published recommendation by national data protection officers for the design and operation of hospital information systems in Germany [18] is examined as a starting point for further investigation. This recommendation gives detailed guidance for mapping privacy compliant accesses to the service processes in a hospital and thus indirectly represents a universal target model. In the ideal case this model needs only be concretized by other policies and thus further restricted. With regard to the abstract policy types that are involved in the overall patient data access this recommendation combines legal and organizational policy concerns. However, such regulations should be compliant with personal or patient-related concerns.

Detection and resolution of conflicts is only possible if the concepts of the policies match. That is why the model of roles and organization of work defined in the guidance needs to be analyzed and compared to other policy concepts that apply to the process of accessing medical data. Such consolidated model is a first contribution and a premise for the conflict model. The conflict model is compared to contributions with respect to policy analysis and policy conflict handling e.g. in [19] and [20]. With the help of the conflict model three handling strategies are then analyzed:

1. All policy concepts are converted into a common policy language, which recognizes conflicts (e.g., by means of policy algebra [21]).
2. From all policy concerns a compound policy is created. If this is not possible, it is a conflict. One issue to compound policies is how to deal with changes on policy concerns and how to enforce these changes.
3. There are general rules of conflict that can be applied independently to individual policy concepts to show absence of conflicts.

5.2 Design as an Artifact

Currently, there is no certainty what kind of IT artifact will emerge since the research is still in progress. It should not be excluded that a toolkit (i.e., an instantiation) to discover potential conflicts with healthcare policies at design time will be developed. The toolkit would allow for the proposition of alternative policy statements for the resolution of policy conflicts. This fits very well to patient consents. Alternatively, a framework with guidance or methods for policy compliance verification might be an outcome. Nevertheless, it must be integrated into monitoring activities of a hospital information system. That is, the IT artifact must provide a valuable contribution for (existing) methods that are applied in tactical and operational level.

5.3 Design Evaluation

To demonstrate utility of the IT artifact it must be evaluated within the HIS infrastructure. Hence, it should be examined whether the IT artifact correspond to the conditions of the internal hospital information management. According to the design evaluation methods outlined in [13], observational, experimental, or descriptive methods should be chosen for this approach. Reasonable candidates of these categories are case study, simulation, and scenarios.

To see utility, it can be checked whether the actual access rights configuration of exemplary information subsystems in the hospital match with the organizational justifications. The result can be evaluated against applied best-practice approaches (e.g., methods provided by the Information Technology Infrastructure Library (ITIL) [22]).

6 Summary and Future Work

Based on the case of a German hospital the paper investigated a specific problem of business and IT alignment: the need for adjusting information systems and IT infrastructure to information security regulations and the frequent changes in this area. The case demonstrated the existence of conflicts between various regulations in the field when refining them for implementation purposes or when considering daily practice of clinicians. As a contribution to business and IT alignment, the paper proposed the use of policies and the development of a semantic model for capturing the specifics of relevant policies and for exposing areas of conflict between them.

Future research will include theoretical and practical parts. From a theoretical perspective, the policy language for the conflict model has to be defined and the model as such has to be developed. In this context, decision will be made whether to accompany application of the conflict model by providing a method or a toolkit, or both. From a practical perspective, the aim is to use the policy language, conflict model and accompanying artifacts in hospitals for validation purposes. Based on the knowledge and experiences gained in such case, a rigorous validation of the developed artifacts from a design science perspective will have to follow, which also will include communication of the research results.

References

1. Seigerroth, U.: Enterprise Modeling and Enterprise Architecture: The Constituents of Transformation and Alignment of Business and IT. IJITBAG 2(1), 16–34 (2011)

2. Woitsch, R., Karagiannis, D., Plexousakis, D., Hinkelmann, K.: Business and IT alignment: the IT-Socket. e & i Elektrotechnik und Informationstechnik 126(7–8), 308–321 (2009)
3. Krogstie, J.: Model-Based Development and Evolution of Information Systems - A Quality Approach. Springer, London (2012)
4. Haas, P., Kuhn, K.A.: Hospital Information Systems. In: Kramme, R., Hoffmann, K.-P., Pozos, R.S. (eds.) Springer Handbook of Medical Technology, pp. 1095–1118. Springer, Heidelberg (2011)
5. Haux, R.: Health information systems–past, present, future. International Journal of Medical Informatics 75(3-4), 268–281 (2006)
6. Brigl, B., Ammenwerth, E., Dujat, C., Gräber, S., Große, A., Häber, A., Jostes, C., Winter, A.: Preparing strategic information management plans for hospitals: a practical guideline. SIM plans for hospitals: a guideline. International Journal of Medical Informatics 74(1), 51–65 (2005)
7. Wies, R.: Policies in Network and Systems Management – Formal Definition and Architecture. Journal of Network and Systems Management 2(1), 63–83 (1994)
8. Caumanns, J., Kuhlisch, R., Pfaff, O., Rode, O.: IHE IT-Infrastructure White Paper: Access Control. IHE International (September 2009)
9. IHE IT Infrastructure Technical Framework: Volume 1 (ITI TF-1): Integration Profiles. IHE International, 9.0 (August 2012)
10. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24(3), 45–78 (2007)
11. Flyvbjerg, B.: Case Study. In: Denzin, N.K., Lincoln, Y.S. (eds.) The Sage Handbook of Qualitative Research, 4th edn., pp. 301–316. Sage Publications Ltd. (2011)
12. Winter, R.: Design science research in Europe. European Journal of Information Systems 17(5), 470–475 (2008)
13. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28(1), 75–105 (2004)
14. Benbasat, I., Goldstein, D.K., Mead, M.: The Case Research Strategy in Studies of Information Systems. MIS Quarterly 11(3), 369–386 (1987)
15. Klein, H.K., Myers, M.D.: A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. MIS Quarterly 23(1), 67–88 (1999)
16. Runeson, P., Höst, M.: Guidelines for conducting and reporting case study research in software engineering. Empirical Software Engineering 14(2), 131–164 (2009)
17. Singer, J., Sim, S.E., Lethbridge, T.C.: Software Engineering Data Collection for Field Studies. In: Shull, F., Singer, J., Sjberg, D.I.K. (eds.) Guide to Advanced Empirical Software Engineering, pp. 9–34. Springer London, London (2008)
18. Conference of the Data Protection Commissioners of the Federation and the Federal Länder: Orientierungshilfe Krankenhausinformationssysteme, in Datenschutzkonforme Gestaltung und Nutzung von Krankenhausinformationssystemen, Würzburg (2011)
19. Moffett, J.D., Sloman, M.S.: Policy Conflict Analysis in Distributed System Management. Journal of Organizational Computing 4(1), 1–22 (1994)
20. Kempter, B., Danciu, V.: Generic Policy Conflict Handling Using a priori Models. In: Schönwälder, J., Serrat, J. (eds.) DSOM 2005. LNCS, vol. 3775, pp. 84–96. Springer, Heidelberg (2005)
21. Bonatti, P.A., di Vimercati, S.D.C., Samarati, P.: An Algebra for Composing Access Control Policies. ACM Transactions on Information and System Security (TISSEC) 5(1), 1–35 (2002)
22. Cabinet Office, IT Infrastructure Library, Official ITIL® Website (May 11, 2012), `http://www.itil-officialsite.com/` (accessed: March 26, 2013)

Configurative Alignment of Business and Application Services: A Work Systems Perspective

Peteris Rudzajs, Marite Kirikova, and Renate Strazdina

Institute of Applied Computer Systems, Riga Technical University, Latvia
{peteris.rudzajs,marite.kirikova,renate.strazdina}@rtu.lv

Abstract. There are different ways of defining business and IT alignment. On one end it is possible to look at the alignment from the strategic perspective; on another end we can go down to the small granularity manual and software functions to see how they are related via human-computer interfaces. In this paper we take a work systems perspective in addressing the alignment between business and IT. In particular, we consider service systems by describing the services at several levels of granularity where they can be instantiated in manual, semi-automatic, and automatic modes of functioning. This perspective helps to address some design challenges for flexible work systems that require intention- and situation-dependent alignment between human and artificial work to manage multiple options of heterogeneous service configurations. The proposed alignment approach is illustrated by education demand and offer correspondence monitoring services.

Keywords: service system, work system, service composition, business-IT alignment, intentional approach, monitoring services, semi-automatic services.

1 Introduction

Business and IT alignment is a broad term, which may address strategic issues, requirements engineering, IT management [1], [2] as well as, at a smaller scale, the distribution of work between human actors and artificial systems. The scope of this paper concerns the challenges of multi-level granularity alignment issues that arise in situations when one and the same service can be performed either by a human actor or software application, or by both depending on the intentions of human actors and situation-dependent configuration possibilities of the service system. Currently these issues are scarcely researched. However, they have to be addressed to achieve the most effective alignment of human and software work in cases where multiple intensions and multiple ways of combination of human and software work are possible. For this purpose the paper proposes and intension oriented configurative alignment of heterogeneous services, which gives an opportunity to achieve business and IT alignment flexibly and neatly.

The approach is discussed using concepts of work systems [3] in general and service systems [4] in particular. The applicability of the approach is analyzed with respect to the following service science challenges:

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 100–111, 2013.

- Use a broadly applicable definition of service.
- Use a broadly applicable definition of service system.
- Replace “the customer” with clear distinctions between various customer groups and other stakeholders whose different interests may conflict.
- Clarify assumptions about the nature and intentions of service system participants.
- Maintain analytical rigor without losing the spirit of service.

These are five of eight challenges identified by Alter [3]. Other three challenges, namely, “Highlight customer and provider responsibilities for value creation”, “Treat servitizing and productizing symmetrically”, and “Locate service systems within value constellations” are not considered in this paper and are matter of further research.

The paper is organized as follows: In Section 2, a particular business and IT alignment problem is defined in the context of work and service systems; definitions of basic concepts used in the paper are presented. Related work is discussed in Section 3. The proposed approach is discussed in Section 4. Brief conclusions are stated in Section 5. In the paper, the proposed approach and concepts are illustrated by education demand and offer correspondence monitoring services [5].

2 Heterogeneous Service System

To illustrate the problem that arises in situations where multiple intensions and multiple ways of combination of human and software work are possible, we start with an example of education demand and offer (d/o) correspondence monitoring service system EduMON [5]. EduMON is aimed at providing mechanisms to support education d/o correspondence monitoring process by the following activities (see Fig. 1):

- Obtaining information from different types of textual information sources (e.g., Web sites, databases, documents, XML-based files) reflecting information about knowledge, skills, and competences demanded and offered.
- Processing information by extracting relevant information such as document structure and relevant keywords encoding knowledge, skills, and competences and analyzing the correspondence between them.
- Distributing current and operative information regarding to knowledge, skills, and competences demanded in labor market and offered in education institutions.

In each above-mentioned activity, various stakeholders (see Fig. 1) are involved. Stakeholders interact via, with, and within EduMON to fulfill specific intentions, e.g., from available information sources obtain corresponding information to their information source (e.g., get university courses where the knowledge, skills, and competences required in particular job position could be obtained), and retrieve cleaned information from newly added information source to be available for further processing. We denote these intentions as intentional services that are captured at business level focusing on the intention they allow to achieve [6]. Intentional services are composed of different other services that belong to Information source management, Information retrieval, Information extraction, Change management, Storage management, Analysis, Notification, and Presentation service classes. In Fig. 1 we illustrate the service classes of EduMON services (according to supplying, processing, and distributing information activities) and examples of intentional services and stakeholders (both internal and external).

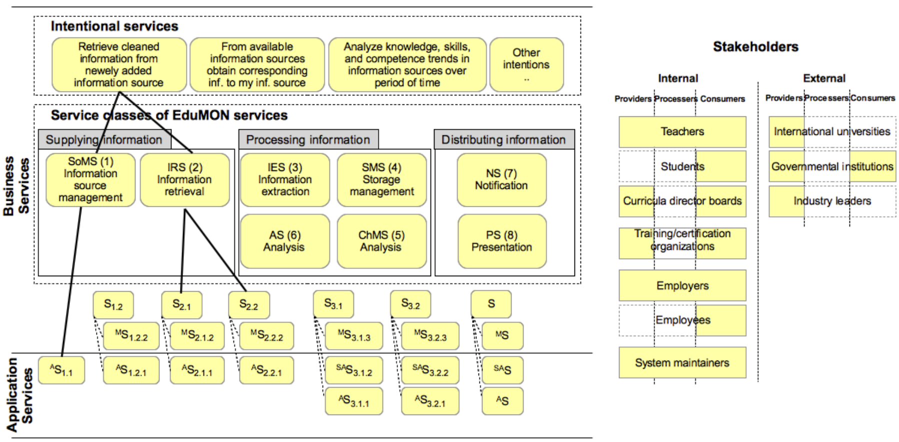


Fig. 1. Overview of EduMON services classes and examples of several intentional services and stakeholders. Bolded lines between services indicate composition of several services, dotted lines - variation of particular service.

As already mentioned, intentional services can be composed of other services from different service classes available in EduMON. For instance, consider the intentional service "Retrieve cleaned information from newly added information source" with the goal to retrieve information from available information source and perform information cleaning (e.g., if information is retrieved from Web sources, then non-relevant information and HTML code should be removed). Suppose that to fulfill this intentional service, in EduMON system three subsequent services $S_{1.1.\text{Get saved source access information}}$, $S_{2.1.\text{Get information from source}}$, and $S_{2.2.\text{Clean information}}$ should be performed where the first number in subscript after letter S denotes to which class the service belongs and the second number is the service number in the class (see Fig. 1 and numbers of service classes and other service examples in Table 1). One and the same service can be performed either by a human actor (manually) or by a software application (automatically), or by both (semi-automatically). This indicates the heterogeneity of services. In the aforementioned example, service $S_{2.1.\text{Get information from source}}$ has variations with two possible modes of functioning, namely, ${}^{A}S_{2.1.1.\text{Get inf. from source by specialized software}}$ (automatic), ${}^{M}S_{2.1.2.\text{Get inf. from source manually}}$ (manual) (the third number in subscript after letter S denotes the number of service variant and letter in the superscript before letter S of service denotes the mode of functioning: A - automatic, M - manual, SA - semi-automatic). When performing an intentional service one or both of these service modes of functioning can be chosen depending on the situation at hand.

Table 1. Examples of some EduMON services with and without variations (A, M, and SA denote mode of functioning)

Service class	Service
SoMS (1)	Without variation: ${}^{A}S_{1.1}$Get saved source access information
	With variation: $S_{1.2}$Find new source $v({}^{A}S_{1.2.1}$Find new source by specialized software, ${}^{M}S_{1.2.2}$Find new source manually)
IRS (2)	With variation : $S_{2.1}$Get information from source $v({}^{A}S_{2.1.1}$Get inf. from source by specialized software, ${}^{M}S_{2.1.2}$Get inf. from source manually)
	With variation: $S_{2.2}$Clean information $v({}^{A}S_{2.2.1}$Clean information by specialized algorithms, ${}^{M}S_{2.2.2}$Clean information manually)
IES (3)	With variation: $S_{3.1}$Extract document structure $v({}^{A}S_{3.1.1}$Extract document structure by specialized algorithms, ${}^{SA}S_{3.1.2}$Extract document structure by manually annotating, ${}^{M}S_{3.1.3}$Extract document structure manually)
	With variation: $S_{3.2.}$Extract document keywords $v({}^{A}S_{3.2.1}$Extract document keywords by specialized algorithms, ${}^{SA}S_{3.2.2}$Extract document keywords by annotating, ${}^{M}S_{3.2.3}$Extract document keywords manually)

Taking into consideration the variety of possible combinations of different services depending on the particular intentions of work, it is necessary to have an approach that ensures business and IT alignment not only at the level of intentions but also at the level of each particular service. The proposed approach that satisfies these requirements is discussed in Section 4 after review of related work in Section 3. The following basic concepts are used in the remainder of the paper.

- Services are acts performed for other entities, including the provision of resources that other *entities* will use [7]. As we should deal with heterogeneous services, it requires sticking to this broad definition provided by Alter. We treat entities as *stakeholders* and *other services*; thus we expand the definition to "*Services* are acts performed for other services or stakeholders, including the provision of resources that other services and stakeholders will use". The resources include data, information, and knowledge.
- *Business service* is a service which is consumed by at least one human actor (or that fulfills a business need for an external or internal service system's customer [8]).
- *Intentional service* is a service that is captured at business level focusing on the intention it allows to achieve [6]. Intentional service is a business service
- *Application service* is a service which exposes automated behavior [8]. It may have automatic and semi-automatic modes of functioning.
- *Atomic service* is a service that can be executed by software service [6].
- *Functioning mode of service*:

  - *manual* - service is performed by a human actor (perhaps, using some office software, but not specific services or tools included in the service system).
  - *automatic* - service is performed by specially designed software tool that does not require human actor intervention.
  - *semi-automatic* - service is performed by some specially designed software tool or service subsystem that requires human input, e.g., a human should prepare and input data into the tool, review and approve data processed and generated by the tool or service subsystem.
- *Service with variation* – service that can be instantiated in different functioning modes.
- *Service system* is defined as a value-coproduction configuration of people, technology, other internal and external service systems, and shared information [4]. Service system is composed of intentional, business and application services.
- *Heterogeneous service system* is a service system in which at least two services can be performed in different functioning modes. At a higher level of granularity, the heterogeneous service system or sub-system may be considered as a composite *heterogeneous service*. In heterogeneous service system we may need to distinguish between data, information, and knowledge flows between the services.
- *Work system* - A work system is a system in which human participants and/or machines perform work (processes and activities) using information, technology, and other resources to produce specific products/services for specific internal and/or external customers. This definition is taken from [9]. Service system is type of work system.
- *Stakeholders* are people who perform work within the work system, including both users and non-users of IT [9]. We describe stakeholders in terms of their *types*, namely, 1) internal and external stakeholders and 2) data, information, and knowledge providers, processers, and consumers (see particular examples in Fig. 1). Stakeholders can belong to one or more organizations or institutions.
- *Configurative alignment* is a composition of heterogeneous service system, sub-system, or service.

3 Related Work

The closest related area to configurative alignment of services is service composition, because it involves the selecting of appropriate services to satisfy some intention (as in the example of EduMON system where the need for service composition is vital). Service composition can be defined as the process of combining and linking existing services (atomic or composite) to create new working services [10]. Chen H-M in [11] claims that rapid movement towards service orientation on both the business and technology fronts offers unprecedented opportunities for business and IT alignment. He outlines that service composition is a powerful concept for business: composite services may be recursively combined to produce even more services, offering new functionality and satisfying the diverse requirements of potential service consumers.

[11] presents a framework where several business and IT alignment approaches are integrated, namely, alignment via architecture, alignment via governance, and alignment via communication. The approach proposed in this paper is described from the point of view of alignment via architecture. It has some conceptual similarities with [11], however the approach is different since [11] does not focus on configurative alignment of heterogeneous services. For selecting the method for heterogeneous service composition several surveys of service composition methods where analyzed. Most of them refer to Web service composition and mainly focus on automatically configurable services. In [10] the following service composition approaches are presented: 1) *Artificial Intelligence Planning* including Finite State Machines, Simulation Calculus, Hierarchical Task Networks, and Petri Nets; 2) *Semantic Web Approaches* including Semantic Annotation, Rule Based Approaches, and Knowledge based Composition; 3) *Middleware approach* including Mobile Agents, Input/Output Dependency, and Policy Based Approaches (see also [2] on service-oriented middleware); 4) *Other approaches* including Composition based on Patterns and Manual Service Composition based on Modeling. In [12] several automated or semi-automated methods of generation of composition process model are surveyed. They are grouped as Methods that use Workflow Techniques and Methods that use Artificial Intelligence Planning. With respect to workflow-based composition methods authors distinguish between static and dynamic workflow generation. The static methods use an abstract process model that has to be built before composition planning starts. The process model includes a set of tasks and their data dependencies. The dynamic composition both creates process model and selects atomic services automatically. Use of this method requires the register to specify constraints, e.g., user preferences. In the category Artificial Intelligence Planning, additionally to the methods listed in [10], the methods based on planners input standards, rule-based planning, theorem proving, and interactive semi-automatic web service composition are included. Static and dynamic service composition are discussed also in [13].Formal methods for service composition are surveyed in [14]. They include Syntactic Service Composition, based on specific XML languages; and Semantic composition based on ontologies or conceptual structures. In [15] the distributed data-flows service composition method is proposed that supports direct data exchange between services. This approach clearly distinguishes between and transparently models both control flow and data flow. The runtime model is called centralized control flow distributed data flow model. Separation between data and control flows is addressed also in [16].

The number of aforementioned methods characterizes the high variety of approaches that can be used for service composition. However, these approaches mainly consider different ways of composing application services with automatic mode of functioning where the alignment only between services with this mode of functioning could be considered. In this paper we are addressing different types of services from the point of view of their modes of functioning and are trying to assure that their composition is configured so that it aligns to the business intention (external alignment) as well as the alignment is ensured at each connection point of services in any permitted configuration of services (internal alignment inside the composition).

The different modes of functioning also stress the necessity to deal not only with data flow but also with information and knowledge flows, because human actors may be involved in service execution. Internal business and IT alignment becomes important in acting in accordance with the work systems theory [9]. This theory treats human participants as part of the systems that generate business results, not just as users of technology; and business processes - as part of work systems, not just the context within which technology is used.

Compliance with work systems perspective requires to carefully consider the design of appropriate interaction interfaces with service performers (artificial or human) thus aligning the output of one service to the input of other service in a different or the same functioning mode. It is important that these interfaces shall support information and knowledge flows, not just data flows, in the service system [17]. Since functioning mode can influence particular configurations in a composite service that helps fulfill a particular business intension, we had to choose or develop a service composition/configuration approach that is flexible enough to show all possible variations at the composite level. The requirement to consider three different types of flow (data, information, and knowledge flows; control flow is out of the scope of this paper) did not permit, at this stage of research, to directly use any of the aforementioned methods of application (with automatic mode of functioning) service composition. Therefore we adopted the method from intention-driven service oriented computing. This approach gives an opportunity to consider interfaces between services in terms of business goals that the services allow to fulfill instead of defining signatures of basic operations that can be invoked on class objects [6]. This permits to take into consideration information and knowledge flows in the service system. The proposed approach is based on the idea of "intentional service" defined in Section 2. From [6] we use the simple notation for composite and variant services (i.e. services with variation) that allows expressing the paths of service configurations. The principles of intentional services and composition of services form the basis of the approach proposed in the next section.

4 Configurative Alignment of Heterogeneous Services

In this section we propose an approach of configuring heterogeneous services in the way that alignment between business and application services is considered with respect to any stakeholder. Internal alignment inside the composition is considered at each connection point of smallest granularity services with various modes of functioning. The concepts used in the approach are described in Section 2. The approach is rooted in the intentional approach of service engineering [6] and the concept of service composition discussed in Section 3.. The proposed approach differs from the approach discussed in [6] in the understanding of the notion of smaller granularity service. In [6] it is atomic service that, in terminology of this paper, is a service ${}^{A}S$ with automatic functioning mode only. Hereby the smallest granularity service can be the service with any of the three modes of functioning (automatic, semi-automatic, and manual). This way it is possible to describe a situation-dependent alignment between human and artificial work and handle multiple options of heterogeneous service configurations. To illustrate the approach we refer to the scenario of EduMON service system where the intentions of stakeholders and composition of heterogeneous services is vital. The approach presented here helps to describe such kind of systems. The proposed approach consists of the following steps:

1. *Identify the set of intentional services.* For illustration we provide two intentional services of EduMON:

   - $S_{\text{Retrieve cleaned information from information source}}$ with the goal to retrieve information from the available information source and perform information cleaning (e.g., if information retrieved from Web sources, then particular page parts and HTML code should be removed).
   - $S_{\text{Get related documents to my new information source}}$ with the goal of adding new information sources to the EduMON system, performing information processing and finding related documents (e.g., get related university courses to new job position).

2. *For each intentional service identify the stakeholders the service is dedicated to.* For instance, for service $S_{\text{Retrieve cleaned information from information source}}$ the stakeholders are of the type "Information providers" (e.g., Teachers, Employers). For service $S_{\text{Get related documents to my new information source}}$ stakeholders are of the type "Information consumers" (e.g., Teachers, Students, Employers and others; see all stakeholders types in Fig. 1).
3. *For each service identify the composition of aggregate or smallest granularity services.* Aggregation of services can involve variants (i.e., smallest granularity services with multiple possible modes of functioning) and other composite services:

   - Composition indicates a sequential order between services and is denoted by "●". Consider, e.g., intentional service "Retrieve cleaned information from information source": $S_{\text{Retrieve cleaned information from information source}} =$

     $$= \bullet\ ({}^{A}S_{\text{1.1 Get saved source access information}}, S_{\text{2.1 Get information from source}}, S_{\text{2.2 Clean information}})$$

   - Variants are of three types, namely, *alternative* (corresponds to XOR relationship between services involved; denoted by "⊗"), *choice* (corresponds to OR relationship; denoted by "*v*"), and *multipath* (when several compositions allow achieving the same intentional service; denoted by "∪"). In the case of EduMON, we use variants to capture multiple modes of functioning. Mode of functioning is assigned to smallest granularity services that are in composition or in variant. ${}^{A}S$, ${}^{M}S$, or ${}^{SA}S$ denotes that the mode of functioning of the service is automatic, manual, or semi-automatic. Decomposition up to the smallest granularity of the above-presented service composition is as follows (see also Fig. 2):

$$S_{\text{Retrieve cleaned information from information source}} =$$

$$= \bullet\ ({}^{A}S_{1.1\ \text{Get saved source access information}},$$
$$v({}^{A}S_{2.1.1\ \text{Get inf. from source by specialized software}},\ {}^{M}S_{2.1.2\ \text{Get inf. from source manually}})$$
$$v({}^{A}S_{2.2.1\ \text{Clean information by specialized algorithms}},\ {}^{M}S_{2.2.2\ \text{Clean information manually}})\ )$$

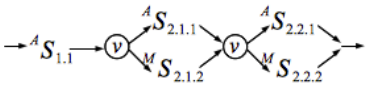


Fig. 2. Intentional service "Retrieve cleaned information from information source"

4. *For each smallest granularity service identify the types of stakeholders* that will perform this service, i.e., internal and external stakeholders and data, information, and knowledge providers, processers, and consumers. For instance, performers for service ${}^{M}S_{2.1.2\ \text{Get inf. from source manually}}$ are of the type "Information providers" (e.g., Teachers, Employers).
5. *To get particular options for service configurations, for each intentional service provide a multipath variant of smallest granularity services* (see the multipath intentional service in Fig 3).

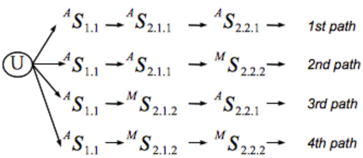


Fig. 3. Multipath intentional service

For instance, the second path in multipath intentional service represented in Fig. 3 has the following steps:

- The system stores URL address to some university course description page (it is "saved source access information"), the service ${}^{A}S_{1.1\ \text{Get saved source access information}}$ obtains (selects) just this URL address (this is only data in form of URL address).
- Then next service ${}^{A}S_{2.1.1\ \text{Get inf. from source by specialized software}}$ uses this data (URL address) to perform automatic retrieval of all the content from this URL address (we get whole web page content - just the data).
- Then a human actor interprets the data (the web page content) and removes non-relevant content by performing ${}^{M}S_{2.2.2\ \text{Clean information manually}}$ service. As a result we get cleaned content - just the course content without HTML tags, banners and non-relevant information (human knowledge is applied to interpret the data, thus we have to deal with information flow here).

When multiple options for service configurations are determined (paths in intentional service), we should focus on transitions between services of different modes of functioning to provide internal (inside the composition) business and IT alignment between the services. Therefore, each service should be analyzed with respect to need for (multiple) input and output interfaces or communication possibilities to be designated or provided (see Table 2) for passing and retrieving data, information, and knowledge to/from service with particular mode of functioning.

Table 2. Interfaces between services of various modes of functioning

Modes functioning	Interface involved	Human actor involved	Possible flows (except of control flow)
^{A}S - ^{A}S	Application level service interfaces should be established between services in transition	-	Data flow Information flow (in case of available meta-data)
^{A}S - ^{M}S	User interface should be established allowing particular stakeholder of ^{M}S to review the output of ^{A}S	+	Data flow Information flow Knowledge flow
^{M}S - ^{A}S	User interface should be established allowing the preparation of the result of ^{M}S for input into ^{A}S	+	Data flow Information flow (in case of available meta-data)
^{M}S - ^{M}S	No specific application level interfaces are required, business level communication must be possible	+	Data flow Information flow Knowledge flow
^{SA}S involved	Depending on the specifics of the ^{SA}S, it may require only application level interface, only business level communication, or both.	+	Data flow Information flow Knowledge flow

The approach discussed in this section facilitates the alignment between business and application services at two levels: external alignment to business intention and internal alignment inside the composition. The approach is applicable in the cases where the smallest granularity services can be performed automatically, semi-automatically, and manually. The validity of the approach can be discussed via its ability to meet the challenges of service science stated in [3]:

- The challenge “use a broadly applicable definition of service” is met fully, because the definition of service introduced in Section 2 can cover all types of services considered in the approach proposed in this section.
- The challenge “use a broadly applicable definition of service system” is met fully, because the definition of service system adopted in Section 2 is applicable to all configurations of related services considered in the description of the approach.
- The challenge “replace “the customer” with clear distinctions between various customer groups and other stakeholders whose different interests may conflict” is met partly. Classes of stakeholders are introduced and used for intentional and smallest granularity services, however no analysis of conflicts of their interest is provided.
- The challenge “clarify assumptions about the nature and intentions of service system participants” is partly met by utilizing work systems perspective where work (service) systems participants are human actors as well as IT solutions.
- The challenge “maintain analytical rigor without losing the spirit of service” is met by adopting and adapting the formalism used in intentional approach to service engineering [6]. Application of the formalism at several levels of service granularity and for all three functional modes of the services gives an opportunity to have formal description of the service configuration without the need to

decompose semi-automated services to the level of detail where the components cannot be treated as meaningful services. Descriptions of interfaces are not yet included in the formalism, however this is one of directions of planned further research.

At this stage of development the approach is partly validated. To estimate the level of the possible business and IT alignment that can be achieved by applying the proposed approach the specific assessment method should be developed in the further research. This assessment method could be used before implementing new smallest granularity services or services with variations to determine if the planned services will allow achieving the desired state. If the assessment results show that alignment cannot be achieved atomic services should be redefined at a more detailed level or replaced by other services

5 Conclusions

In this paper business and IT alignment is addressed from the point of view of opportunities offered by service orientation in both the business and IT [11]. In particular, it is investigated how to configure business and application services so that their alignment is achieved at different levels of granularity. The solution is based on (1) use of work systems theory that puts a human actor inside the system, (2) treating service system as a work system, (3) use of concept of intentional service for business services; (4) consideration of three modes of service functioning, and (5) careful analysis of data, information and knowledge flows among the smallest granularity services.

The proposed approach helps meet five of the eight service systems challenges [3] at least partly. Further research concerning service interface specification shall give an opportunity to analyze data, information, and knowledge production in service system at different levels of granularity and perhaps, in turn, to identify sub-configurations that can be treated by well-known service composition methods and to separate them from sub-configurations, where only work systems theory based treatment is applicable. Other directions of further research concern the analysis of services, a service system, and service sub-systems from the value creation and value networks perspectives, and developing assessment method of possible alignment that can be achieved by applying the proposed approach.

Acknowledgments. This work has been supported by the European Social Fund within the project "Support for the implementation of doctoral studies at Riga Technical University".

References

1. Baïna, S., Ansias, P., Petit, M., Castiaux, A.: Strategic Business/IT alignment using goal models. In: Proceedings of the 3rd International Workshop on Business/IT Alignment and Interoperability (held in conjunction with CAiSE 2008), pp. 31–43 (2008)

2. Regev, G., Favre, J., Hayek, E., Wilson, P., Wegmann, A.: Business/IT Alignment in Practice: Lessons Learned from a Requirements Project at P&G. In: Salinesi, C., Pastor, O. (eds.) CAiSE Workshops 2011. LNBIP, vol. 83, pp. 93–101. Springer, Heidelberg (2011)
3. Alter, S.: Challenges for Service Science. Journal of Information Technology Theory and Application 13, 22–37 (2012)
4. Spohrer, J., Maglio, P., Bailey, J., Gruhl, D.: Steps Toward a Science of Service Systems. Computer 71–77 (2007)
5. Rudzajs, P.: Towards Automated Education Demand-Offer Information Monitoring: the System's Architecture. In: Niedrite, L., Strazdina, R., Wangler, B. (eds.) BIR Workshops 2011. LNBIP, vol. 106, pp. 252–265. Springer, Heidelberg (2012)
6. Rolland, C., Kirsch-Pinheiro, M., Souveyet, C.: An intentional approach to service engineering. IEEE Transactions on Service Computing 3, 292–305 (2010)
7. Alter, S.: Viewing systems as services: a fresh approach in the IS field. Communications of the Association for Information Systems 26, 195–224 (2010)
8. The Open Group: ArchiMate 2.0 specification (2012), `http://pubs.opengroup.org/architecture/archimate2-doc/`
9. Alter, S.: Work System Theory: Overview of Core Concepts, Extensions, and Challenges for the Future. Journal of the Association for Information Systems 14, 72–121 (2013)
10. Kapitsaki, G., Kateros, D., Foukarakis, I., Prezerakos, G., Kaklamani, D., Venieris, I.: Service Composition: State of the art and future challenges. In: Proceedings of the Mobile and Wireless Communications Summit, pp. 1–5 (2007)
11. Chen, H.: Towards service engineering: service orientation and business-IT alignment. In: Proceedings of the 41st Hawaii International Conference on System Sciences (2008)
12. Rao, J., Su, X.: A survey of automated web service composition methods. In: Cardoso, J., Sheth, A.P. (eds.) SWSWPC 2004. LNCS, vol. 3387, pp. 43–54. Springer, Heidelberg (2005)
13. Beek, M., ter, B.A., Gnesi, S.: A survey on service composition approaches: From industrial standards to formal methods. In: Proceedings of the 2nd International Conference on Internet and Web Applications and Services, ICIW 2007 (2006)
14. Beek, M., Ter, B.A., Gnesi, S.: Formal methods for service composition. Annals of Mathematics, Computing & Teleinformatics 1, 1–10 (2007)
15. Liu, D., Peng, J., Law, K., Wiederhold, G., Sriram, R.: Composition of engineering web services with distributed data-flows and computations. Advanced Engineering Informatics 19, 25–42 (2005)
16. Monsieur, G., Snoeck, M., Lemahieu, W.: Managing data dependencies in service compositions. Journal of Systems and Software 85, 2604–2628 (2012)
17. Kirikova, M., Rudzajs, P.: Multimode Information Logistics for Conceptual Correspondence Monitoring. In: Proceedings of 5th Workshop on Information Logistics and Knowledge Supply (Held in Conjunction with BIR 2012), pp. 31–42 (2012)

TOGAF Adaption for Small and Medium Enterprises

Rebekka Alm and Matthias Wißotzki

The University of Rostock, Institute of Computer Science, Chair of Business Information Systems, Albert-Einstein-Str. 22, 18059 Rostock, Germany
{rebekka.alm,matthias.wissotzki}@uni-rostock.de

Abstract. Enterprise Architecture Management (EAM) includes the planning, management, control and improvement of enterprise architecture. The Open Group Architecture Framework (TOGAF) is the best known and most trusted enterprise architecture standard used to improve the operational efficiency. It brings many benefits, but is also associated with costs. While the applicability to large enterprises is beyond question, the application in small and medium enterprises (SME) in consideration of costs and benefits is controversial. Scientific literature gives no satisfactory recommendation to SME. A survey with TOGAF experienced users was performed to get a better understanding of the most important parts of the framework and the possibility of adapting TOGAF to the requirements of SME.

Keywords: TOGAF, Enterprise Architecture Management (EAM), small and medium enterprises (SME).

1 Introduction

Enterprise Architecture Management (EAM) includes the planning, management, control and improvement of enterprise architecture [5, 12]. The Open Group Architecture Framework (TOGAF[1]) is the best known and most trusted enterprise architecture standard [10]. It includes a recognized methodology and a framework for enterprise architecture. TOGAF is used by leading companies worldwide in order to improve the operational efficiency[2].

There are four architecture layers in EAM that can be distinguished, which are referred to in the context of TOGAF as business architecture, data architecture, application architecture and technology architecture. The development of these four interconnected architecture layers is fully supported by TOGAF. The framework is divided into seven parts as shown in Fig. 1.

The Architecture Development Method (ADM) is the core of TOGAF as shown in Fig. 1. The ADM is illustrated not only centered, but is connected

[1] http://www.togaf.info/

[2] An overview of the institutions and companies that use TOGAF: http://www3.opengroup.org/togaf/users-by-market-sector

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 112–123, 2013.

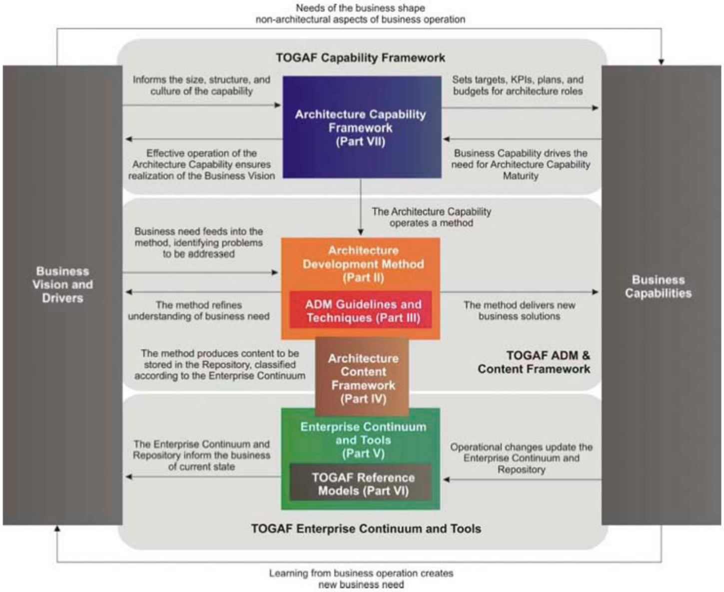


Fig. 1. TOGAF Content Overview [9]

to all the other parts. It describes the process of developing and managing the enterprise architecture life cycle.

The development process is divided into eight stages. It is triggered by the preparation phase and monitored by the requirements management. The architectural requirements management is a continuous phase, ensuring requirements changes are met through appropriate governance processes and are considered in all other phases. Additionally all phases are divided into steps and each phase contains well-defined inputs and outputs, which are processed or generated by these steps. According to the standard document the numbering of the phases is not fixed, but rather exemplary. The architect should change the order in consideration of the requirements of the company, existing repositories and the tools used [9].

The large size and complexity of such a framework makes it more difficult to handle for the user [6]. This problem is well known and often cited as generalized reason why a framework driven EAM is not profitable for SME. However, the problem itself is barely addressed. The TOGAF documentation refers to the adaptability of the framework [9]. Several reasons for adaption are mentioned, one being the need for a customized, possibly reduced form of the framework for SME. However, no indication is given what parts should be reduced in this case.

There are a variety of reasons for implementing EAM, such as summarized by Weinberger [15]:

– Supports delivery of the business strategy
– Effective management and exploitation of information is key to business sucess and competitive advantage

– Manage stakeholder concerns that needed to be addressed by IT systems
– Manage complexity and changes to business/IT
– Enables the right balance between IT efficiency and business innovation
– Improvement of transparency and manage risks
– Optimizes the (fragmented) legacy processes (manual and automated) to an integrated environment

The applicability to large enterprises is beyond question, as The Open Group's own website lists some familiar names[3]. Yet there is no sufficient exploration on whether the implementation of TOGAF in SME could be reasonable.

To answer this question two research approaches were used that will be introduced in Sect. 2. The approach and results of the literature analysis are described in Sect. 3. Here the problem of Enterprise Architecture Management (EAM) in SME is explained and pros and cons of TOGAF are compared. Since the problem has not been addressed satisfactorily by the literature sources, a survey was performed. The contents and results of the survey are presented in Sect. 4 and limitations are assessed in Sect. 5.

2 Research Approach

The first research approach used by this paper is a methodological review of research results. The approach aims to be transparent, conclusive, and repeatable for the audience. The purpose of this paper is to summarize knowledge on TOGAF and to identify its applicability for SME. We have conducted a systematic literature review based on the guidelines of Kitchenham et al. [13]:

1. Formulation of the research questions to define the important topics and relevant research fields.
2. Identification of literature sources covering TOGAF and EAM
3. Selection of papers for inclusion in the analysis
4. Data extraction from selected papers
5. Presentation of results
6. Interpretation of results

Results of the literature review are summarized in Sect. 3. A preliminary overview has shown a limited amount of existing scientific papers with relevance to this topic, as well as limited availability and access to the few sources we identified. Therefore we decided to validate our conclusions through a survey.

A survey is defined in the following way [8]:

– The purpose of a survey is to produce statistics, that is, quantitative or numerical descriptions of some aspects of the study population.
– The main way of collecting information is by asking questions; their answers constitute the data to be analyzed.

[3] http://www3.opengroup.org/togaf/users-by-market-sector

– Generally information is to be collected from only a fraction of the population, that is a sample, rather than from every member of the population.

Relevant research questions were identified through the literature review. As target population a sample of TOGAF certified experienced users was chosen. The survey was performed web-based[4]. Approach and results are represented in Sect. 4.

3 State of the Art

This chapter provides an overview of the approach in the search for literature on the subject of TOGAF in SME. The effort associated with the development of an enterprise architecture by applying TOGAF should produce appropriate benefits. SME do not have the same means such as large companies and applying EAM can result in noteworthy costs. It should be answered, what changes to the approach could significantly reduce the effort for SME. It would be helpful to analyze case studies of small and medium enterprises, which successfully use TOGAF or had reason to cancel a planned deployment. The list of publications[5] on The Open Group website contains some case studies. These, however, refer to global players, therefore large companies. For the systematic literature review there were used the academic search engines Springerlink and Bielefeld Academic Search Engine (BASE). Due to the lack of relevant results the general search engine Google was used as well.

Springerlink. The advanced search in full text for "TOGAF" returned 348 documents. The limitation of the search on title and abstract reduced the number to 22 hits. From this list only six titles were available from the network of the University of Rostock [1–4, 7, 14]. These articles were not applicable to the context of SME.

Bielefeld Academic Search Engine (BASE). The default search with the keyword "TOGAF" in BASE provided a set of 62 articles. This search yielded no relevant results, since none of the articles were based on small and medium enterprises, or if so, only in an indirect way.

Google. The search on Google Scholar[6] was also discouraging. Following keywords were used in the search: "TOGAF experiences", "TOGAF in SME", "TOGAF publication", "TOGAF small medium business" and "TOGAF". This search yielded few relevant new results. To achieve further reference litereráture it was sought for general "EAM in SME". The main result of this search was the book by Keuntje and Barkow [11], in which the relevance of EAM for SME is analyzed. Since the literature results on Google Scholar were still not satisfactory, the same approach was repeated for Google's general search engine. Result of this search were some interesting articles with instructions for use and adaptation of TOGAF in general and SME in particular, however, no scientific publications.

[4] LimeSurvey: http://www.limesurvey.org/de

[5] https://www2.opengroup.org/ogsys/jsp/publications/PublicationsFree.jsp

[6] http://scholar.google.de/

Results. Keuntje and Barkow investigate in [11] the relevance of EAM for the middle class and analyze it regarding two drivers for EAM they identified:

1. To support the systematic, holistic alignment of IT activities to current and future business requirements
2. Management of complexity of the application landscape and associated infrastructure

The first point is for any company of interest, for SME as for large companies alike. Critically is the question of the need for specific methods, models or possibly software solutions to implement. Keuntje and Barkow refer to the possibility and the need to adapt, to find a balanced compromise between semantics and completeness of the result with a justifiable effort. The second point that focusses on the complexity of the company is strongly motivated by the two authors. In large companies high complexity is already given by the variety and heterogeneity of the applications used. Critically, however, be noted that the complexity in this case would likely be managed instead of actually mastered. In (growing) SME "chaos" and complexity can be prevented with little effort by early use of EAM. The necessary changes are according to [11] implemented faster and with less effort than in large companies. The assessment of the company situation can be created within a short time and thus timely and accurate. Moreover, in a small company with a flat structure, all relevant employees are reached simultaneously. Thus, an SME in the right circumstances and with the necessary motivation can much more quickly achieve a high level of maturity. A key success factor in this context is the acceptance and cooperation of stakeholders [15]. The corporate culture must be established and demonstrated by the Executive Board.

To assess the relevance of EAM for SME Keuntje and Barkow prepared a checklist that includes the following areas: documentation, efficiency and security of the IT organization, competitive importance of IT, environment and market dynamics of the development of the company. The checklist is based on the extend of the dynamic of the environment in which the company is located. Is the company in a near optimal state without pressure to change, the benefits of EAM can indeed be questioned. Otherwise EAM is recommended, regardless of the size of the company. The choice of a concrete EAM approach is left to the reader.

Dube and Dixit [6] compare different EAM approaches and evaluate them according to criteria in three areas: "Higher Order Goals", "non-funcional requirements" and "Inputs and Outcomes". TOGAF is winner of every category and receives above-average ratings for all criteria.

In summary it can be said that the issue of the applicability of TOGAF for SME is not considered sufficiently in the literature. However, the hypothesis can be deduced that TOGAF is certainly of interest for some SME. EAM is particularly advisable if the SME is in a dynamic environment, or in growth and if a longer "lifetime" of the company is aimed at. To reduce the cost and complexity of the framework different measures can be taken. This measures are regrettably not clarified. According to the TOGAF documentation [9] all parts of the framework act together as a whole, yet it is still possible within a company to implement only part of it. In the introduction of the documentation it is even explicitly recommended to focus on certain components and to classify other for later viewing when companies use TOGAF for the first time. A specific recommendation on which components to focus is lacking so far.

4 SME Survey Results

The survey was designed to compensate the unsatisfactory results of the literature review. The questions aim to identify the main components of the framework, and in particular the TOGAF Architecture Development Method as the core of the framework. This, and an open question at the end of the survey was designed to answer which parts of TOGAF are the most valuable assets and can be suggested to small and medium enterprises.

The survey results are to answer the following research questions:

- RQ1: Which components of TOGAF are particularly useful?
- RQ2: What phases of the ADM are particularly useful?
- RQ3: What steps in the respective phases are particularly useful?
- RQ4: Does it make sense to apply TOGAF to small and medium enterprises?

Respondents. Survey participants were acquired in two ways. First, certified TOGAF users[7] were asked by mail to participate. Furthermore, the survey was published in an international TOGAF user group on XING. The survey was answered 53 times in total. The return rate was 10%. Most participants were consultants (43%), followed by modelers and architects (25%). The respondents can be assumed to be experienced with TOGAF. More than a third (42%) stated to have at least 5 years experience with TOGAF, almost a third (32%) had 3-4 years experience. Among the remaining responses, most participants have 2 years experience (17%) and only 8% of the participants have 0 or 1 year experience with TOGAF.

Survey Contents. The survey was divided into four parts. The first part queried for general data about the person and the company. In the second part the participants were challenged to select five phases of ADM they would run if they could only perform five of the ten ADM phases. This type of questioning was used to identify key components of TOGAF. The answer to this question conditioned which phase were displayed to the participant in the following subquestions. In addition, the participants were requested to evaluate the seven TOGAF components regarding their usefulness. There was no neutral response option available, so participants were forced to specify a tendency to useful or not useful. The third part is similar in design to the second part, the content depended on the responses from the second part: The participant received all

[7] TOGAF Directory of Certified People:
`http://www.opengroup.org/togaf9/cert/cert_archlist-short.tpl`

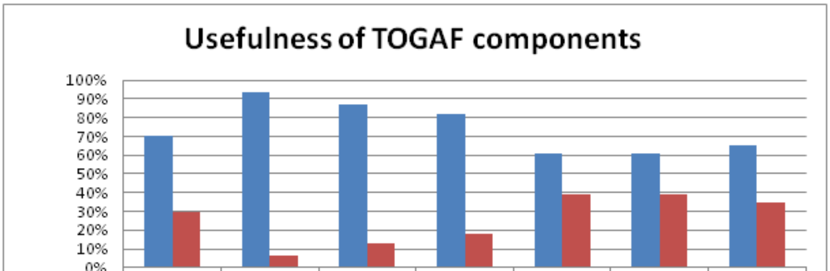
the steps of the selected phase, and had to choose which steps to perform, if he could only carry out half of the steps. At last the participants were asked an open question querying about the applicability of TOGAF in SME:

> "Do you believe TOGAF should/could be applied in small to medium enterprises? Please answer and give reasons for your opinion."

RQ1: Which Components of TOGAF are Particularly Useful? Participants were challenged to rate the 7 components each on a scale from "very useful", "useful" to "occasionally useful" and "not useful". In Fig. 2 the answers were summarized to be "useful" and "not useful". For this purpose the following formula was prepared:

$$percentage_{useful} = \frac{count_{veryuseful} * 1 + count_{useful} * 0,5}{count_{total}} \tag{1}$$

An equivalent formula was used for the non-useful answers.



	Introduction	ADM	ADM Guidelines	Architecture Content Framework	Enterprise Continuum and Tools	TOGAF Reference Models	Architecture Capability Framework
useful	70%	93%	87%	82%	61%	61%	65%
not useful	30%	7%	13%	18%	39%	39%	35%

Fig. 2. Usefulness of TOGAF components

Figure 2 clearly shows that the ADM and its guidelines are considered the most important component of TOGAF. This confirms the approach to ask detailed questions about ADM.

RQ2: What Phases of the ADM are Particularly Useful? To identify the "most useful" or main phases, the question was asked, what phases the respondent would perform if he could run only 5 of the 10 phases. The phases of business architecture (83%), and Architecture Vision and Information System Architecture with each 70% were the most voted phases. It should be noted that 100% would have been achieved if each participant had executed the respective phase. The diagram in Fig. 3 shows the results in more detail.

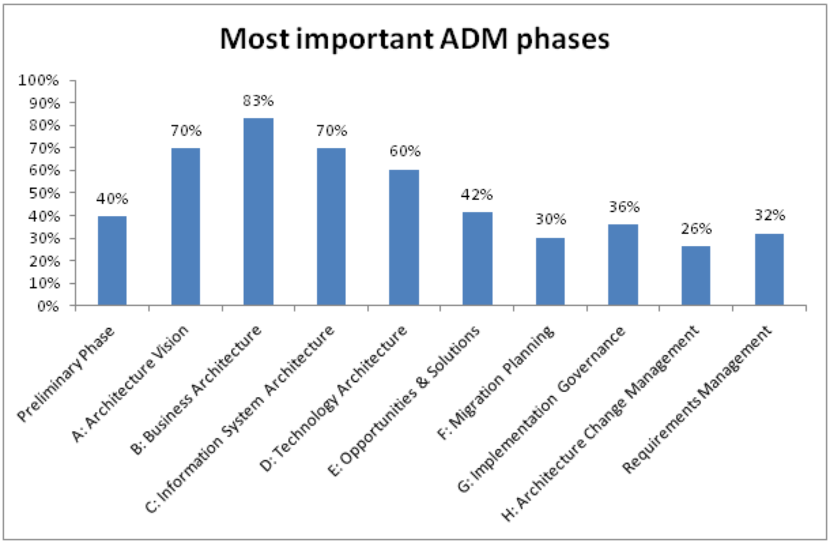


Fig. 3. The most useful ADM phases

RQ3: What Steps in the Respective Phases are Particularly Useful? Another result of the survey was an assessment of the participants about the importance of each step of each phase. We will discuss the similarities and striking differences we identified. Due to the high degree of detail and scope needed for an expressive presentation, the specific results for each phase shall be omitted. Exemplary the specific results for the most useful[8] phase B (Business Architecture) will be presented. Phase B was initially chosen by 44 of the 53 participants as one of the most useful phases. The answers of the respondents are visualized in Fig. 4.

The steps “Develop Target Business Architecture” (100% as a baseline), “Develop Baseline Business Architecture” (85%) and “Perform Gap Analysis” (81%) are leading. Interestingly, the step “Conduct Stakeholder Review” (52%) is a short distance before any further steps. The same step in the phases C & D (Information System Architecture & Technology Architecture), only got 27% and 22% of the votes. Therefore, the step seems to be of particular importance in phase B.

Another striking difference between the phases can be determined for the steps regarding confirmation. The confirmation takes a central role in the Architecture Vision. The particular step is ”Confirm and Elaborate Business Goals, Business Drivers, and Constraints”. 54% of the participants think this step to be of particular importance. This value is significantly higher than the similar steps of the other phases. This shows the importance of the approval of business objectives.

[8] RQ2: What phases of the ADM are particularly useful?

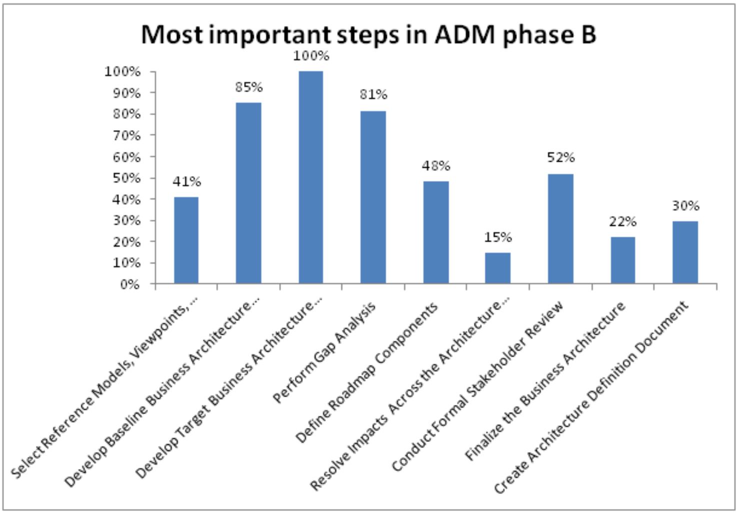


Fig. 4. Most important steps for TOGAF phase B

The step of "Scoping" showed to be particularly important in different phases. Both in the Preliminary phase and in the Architecture Vision, the participants associated this step with a particularly high importance (67% and 76%) .

RQ4: Does it Make Sense to Apply TOGAF to Small and Medium Enterprises? The majority of participants answered "Yes" or "Yes, partially" (84 %). The following points were mentioned in the answers:

- SME should skip steps or implement them less detailed.
- SME should minimize Documentation and Tracking Steps.
- TOGAF is suitable for medium-sized companies, but not suitable for small businesses.
- The common language and the vocabulary by TOGAF is useful for any company.
- TOGAF is a scalable framework.
- Depending on the use case, the framework needs to be adjusted differently.
- At least the ADM can be useful in SME.
- TOGAF is useful for all companies in which much change occurs, and not useful in static businesses regardless of size.
- It was repeatedly mentioned that the size of the company is not the determining factor for the application of TOGAF.
- The desire for a TOGAF framework "Lite" with reduced overhead for SME has been mentioned several times.
- One participant believed that many users confuse TOGAF with a method, rather than seeing it as customizable framework.
- It is recommended that adjustment of TOGAF was to be carried out with assistance of an external expert with many years experience.

Participants answering "No", mentioned the following reasons:

- TOGAF is too complex and abstract.
- TOGAF creates too much overhead.
- It is questionable whether using TOGAF would benefit the company.

The negative foundations seem very generalizing, so it can be assumed that the positive responses are more deliberate and accurate.

Summary. The research questions were largely answered by the survey results. The first question regarding the most helpful components of TOGAF could confirm existing assumptions. It turned out that the ADM is especially useful while the reference models are less useful (Fig. 2). This seems logical and is supported by a high number of participants. The second question was about the most important phases in TOGAF. The results are satisfactory and can also be read directly from the survey results (Fig. 3). For the third research question concerning the steps of each phase, the most frequently selected steps were rated as most important.

For the fourth research question we collected evidence in form of opinions of experienced users. A complete answer to the question is missing, however, because not only employees of SME were registered as participants for the survey.

5 Limitations

The survey was approached open minded and the questions were formulated neutral. More practical experience with TOGAF could have improved the survey.

The questions were deliberately chosen to challenge the experts, which TOGAF phases and steps are most useful and correspondingly which phases could be paid less attention to reduce the effort for SME. Some participants did have great objections to the notion of omitting phases, so they refused to participate in the survey. Another participant had difficulty answering the questions, since the steps in his view, can be selected generally valid, but only as a function of the implemented project. This is in our view a legitimate criticism, but the problem is difficult to avoid in a survey.

As for the size of company, mostly employees of large organizations with 501 or more employees have responded (64%). From medium-sized companies with 11 to 500 employees were 23% of the responses. Only 13% of the participants were from small businesses with 1 to 10 employees. Based on the sales 27% of the participants can be described as small (2 million euros), an additional 18% of the organizations can be classified as a medium on account of having a turnover of 2 to 50 million euros. The majority (55%) of the the companies can be described in terms of sales as large. It is to be noted that many of the participants were consultants, and thus possibly only the consulting firm was an SME, but TOGAF is actually applied in a large company. Unfortunately, the design of the survey did not allow to make a distinction here.

6 Conclusion and Outlook

This paper deals with the issue of the applicability of TOGAF in small and medium enterprises. For this purpose we introduced the framework and analyzed existing literature reports. The subject of TOGAF in SME is largely undocumented and opinions on the topic vary widely. The sweeping opinion seems to be that EAM in SME will bring little benefit while causing excessive costs. Selected literature as well as the survey show, however, that TOGAF experts and users see great potential in applying EAM (especially TOGAF) in SME. We agree with the EAM experts who succeed by arguments. The business environment is increasingly dynamic and competitive pressure is increasing. A company cannot afford to pay the price of the consequences of lack of EAM. The measures to reduce the cost and complexity of the application of TOGAF were enumerated in this work, the benefits of TOGAF were outlined as well.

We tried to compensate the lack of available literature on the subject by a survey. The high response rate and the high level of professional knowledge of the participants' responses sufficiently confirmed our hypothesis that EAM or TOGAF is also of interest to SME. Furthermore, we have established that cost reduction by the omission of complete stages is to be applied with caution. Advisable is instead to edit the different phases in a large scale. The weighting of the phases is in this regard an important result of the survey and may serve as a guide for SME, which phases should be edited to what extent. But this only serves as a base. The adjustments of the framework must be done individually according to the nature and objectives of the company. This issue should be explored further based on the findings of the survey.

References

1. Atkinson, C.: Orthographic software modelling: a novel approach to view-based software engineering. In: Kühne, T., Selic, B., Gervais, M.-P., Terrier, F. (eds.) ECMFA 2010. LNCS, vol. 6138, p. 1. Springer, Heidelberg (2010)
2. Balabko, P., Wegmann, A.: Systemic classification of concern-based design methods in the context of enterprise architecture. Information Systems Frontiers 8(2), 115–131 (2006)
3. Czarnecki, A., Orłowski, C.: Ontology as a tool for the it management standards support. In: Jędrzejowicz, P., Nguyen, N.T., Howlet, R.J., Jain, L.C. (eds.) KES-AMSTA 2010, Part II. LNCS, vol. 6071, pp. 330–339. Springer, Heidelberg (2010)
4. Czarnecki, A., Orłowski, C.: It business standards as an ontology domain. In: Jędrzejowicz, P., Nguyen, N.T., Hoang, K. (eds.) ICCCI 2011, Part I. LNCS, vol. 6922, pp. 582–591. Springer, Heidelberg (2011)
5. Dern, G.: Management von it-architekturen. In: Management von IT-Architekturen: Leitlinien für die Ausrichtung, Planung und Gestaltung von Informations Systemen, vol. 1. Vieweg+ Teubner Verlag— GWV Fachverlage GmbH, Wiesbaden (2009) ISBN 978-3-8348-0718-2
6. Dube, M.R., Dixit, S.K.: Comprehensive measurement framework for enterprise architectures. CoRR, abs/1109.1891 (2011)

7. Engelsman, W., Wieringa, R.J.: Goal-oriented requirements engineering and enterprise architecture: Two case studies and some lessons learned. In: Regnell, B., Damian, D. (eds.) REFSQ 2011. LNCS, vol. 7195, pp. 306–320. Springer, Heidelberg (2012)
8. Floyd, J., Fowler, J.: Survey Research Methods. Applied Social Research Methods. SAGE Publications (2001)
9. The Open Group. Open group standard - togaf version 9.1. Web: `http://pubs.opengroup.org/architecture/togaf9-doc/arch/`
10. The Open Group. An introduction to togaf version 9.1. (December 2011) Web: `http://www.opengroup.org/togaf/`
11. Keuntje, J.H., Barkow, R.: Enterprise Architecture Management in der Praxis: Wandel, Komplexität und IT-Kosten im Unternehmen beherrschen. Symposion Publishing GmbH (2010)
12. Kaisler, S.H., Armour, F., Valivullah, M.: Enterprise architecting: Critical problems. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences, HICSS 2005, p. 224b (January 2005)
13. Kitchenham, B.A., Pfleeger, S.L., Pickard, L.M., Jones, P.W., Hoaglin, D.C., El Emam, K., Rosenberg, J.: Preliminary guidelines for empirical research in software engineering. IEEE Transactions on Software Engineering 28(8), 721–734 (2002)
14. van Bommel, P., Hoppenbrouwers, S.J.B.A., Proper, H.A.E., van der Weide, T.P.: Giving meaning to enterprise architectures: Architecture principles with orm and orc on the move to meaningful internet systems 2006. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM 2006 Workshops. LNCS, vol. 4278, pp. 1138–1147. Springer, Heidelberg (2006)
15. Weinberger, D.: ...und am anfang steht die geschäftsanforderung, oder? OBJEKTspektrum, EAM (2010)

The Industrial Practice of ITIL Implementation in Medium-Sized Enterprises

Philipp Schmidtbauer, Kurt Sandkuhl, and Dirk Stamer

University of Rostock, Institute of Computer Science
Albert-Einstein-Str. 22, 18059 Rostock, Germany
{philipp.schmidtbauer,kurt.sandkuhl,dirk.stamer}@uni-rostock.de

Abstract. The Information Technology Infrastructure Library (ITIL) is becoming more and more important for IT service management in the industrial and administrative sector. Service operation as one of the five ITIL core processes provides guidance for delivery and control of IT services at agreed quality levels. Main purpose of this paper is to contribute to the aspect of industrial practice of ITIL implementation in small and medium-sized enterprises. Therefor a case study of an ITIL service operation implementation is explained and analyzed.

1 Introduction

In industrial areas with a highly competitive environment all enterprise departments and functions are expected to contribute to efficient operations and an economic cost structure. In particular the IT-budgets of organizations have been under pressure during the last decade with a clearly expressed expectation towards IT-departments to provide solutions and services of high quality tailored to business demands. Furthermore, many enterprises consider outsourcing of such IT-services as an option to reduce IT-related costs, which can be classified as commodities [1]. In particular in decision processes for new information systems with take-make-buy situations this is an option.

In the context of the above situation, IT-departments are increasingly forced to show the value of IT-services for the overall enterprise. Technology-centric attitudes towards IT have to be substituted by a customer-oriented perspective. Service-orientation as proposed and propagated by ITIL (see section 2) offers a widely acknowledged approach to establish continuous service improvement. The overall aim of the ITIL reference framework is to provide dependable IT-services in accordance to the customers' demands with respect to functionality, quality of service and transparency. The ITIL framework also includes best practices supporting the transformation of function-oriented to service-oriented organizations as seen in Fig. 1 [2].

However, ITIL is considered as complex, quite large and connected to high implementation efforts. The motivation for this paper is to investigate, whether ITIL is suitable for medium-sized enterprises as it is or needs adaptation. The focus is put on a case study at the wind turbine manufacturer Nordex from Germany. Main contributions of the paper are (1) experiences from process modeling in the context of

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 124–135, 2013.

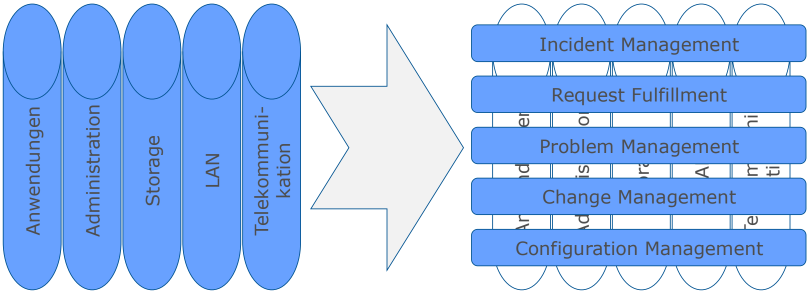


Fig. 1. – Process-oriented alignment of the IT functions (see [2])

IT-service management, (2) a case study for ITIL implementation at a medium-sized enterprise and (3) conclusions regarding the suitability of ITIL for this size of enterprises.

The remaining part of the paper is structured as follows: section 2 gives an introduction to IT service management (ITSM) in general and ITIL in particular. Section 3 discusses the research method used for the case study and the research questions investigated. Section 4 presents results of a literature analysis focusing on approaches for the industrial ITIL implementation in general and with focus on SME. Section 5 presents the industrial case study results (detailed source [3]), before section 6 summarizes the findings based on the research questions.

2 IT Service Management

2.1 ITSM Approach

The services of an IT organization in terms of IT services supporting and enabling business provided through a comprehensible and cost transparent management system can convince business management by adding value to the business processes and operations. A distinction is made between the IT as a supporter or enabler of the business processes of a company or of a customer. [4] [S. 11]

Definitions of IT service as a business term point to a strong focus on customer requirements, since these requirements and demands are the basic reason why the IT service exists, i.e. an essential feature of ITSM is customer orientation [5] [S. 8]. The core idea of the ITSM approach is the transformation of resources into valuable IT services [6] [S. 9]. The IT services of an organization have to be aligned to current and future needs of the customer. The quality of services provided should be continuously improved and the long-term costs of the service delivery should be reduced [7] [S. 30].

In the official ITIL literature ITSM is defined as a set of "[…] specialized organizational capabilities for providing value to customers in the form of services" [6] [S. 11]. This definition disregards the details of the service provision. Bick [7] [S. 30] defines those details as the totality of measures for the creation and provision of IT services to enable a needs-based planning, operation, supervision und control.

Addy provides a broader definition of ITSM [8] [S. 46]. ITSM is defined as „[…] the planned and controlled utilisation of IT assets (including systems, infrastructure and tools), people and processes to support the operational needs of the business as efficiently as possible whilst ensuring that the organisation has the ability to quickly and effectively react to unplanned events, changing circumstances and new business requirements as well as continuously evaluating its processes and performance in order to identify and implement opportunities for improvement."

Table 1. – Reference models for service-oriented IT service delivery (see [9])

Reference model	*Developer*	*Short description*
ITIL	Office of Government Commerce	De-facto standard for service-oriented IT management
IT Service CMM	Vrije Universiteit, Amsterdam	Maturity model for IT service management
Managerial Step-by-Step Plan (MSP)	Delft University of Technology	Stepwise plan for designing information system management
eTOM	Telemanagement Forum	Process model for service provider

Table 1 shows in an overview a selection of reference models that can be used to control the delivery of services in IT organizations. There is a visible trend towards the ITSM alignment. Following Krcmar [9] [S. 362] reference models are particularly suitable for answering the question which business processes are necessary and how these certain processes are required to be designed.

2.2 ITIL

The ITIL reference model is a guide that has been developed by the government agency Office of Government Commerce (OGC) in the UK. The reference model in the current version 3 (ITIL V3) from the year 2007 was published in a series of six books. This series is called the ITIL core publication and is under continuous development.

The ITIL reference model describes the integrated and process-based best practice approach to manage IT services. The non-proprietary license of the ITIL reference model allows for the free distribution of the core ideas as printed books and digital eBooks. The ITIL core publication is distributed through the UK-based privatized government publishing house 'The Stationary Office' (TSO). [8] [S. XXXIX]

In the 2007 edition ITIL V3 was published in a core publication of five books and one introductory volume. The six titles of the ITIL core publication are:

- Introduction to ITIL Service Management,
- Service Strategy,

- Service Design,
- Service Transition,
- Service Operation und
- Continual Service Improvement.

The volume Service Operation consists of the management practices of daily service operation. The service operation lifecycle phase is on the one hand responsible for the coordination and execution of the activities and processes. On the other hand this phase is also responsible for operating the technology needed for the delivery of the IT services as defined in the service level agreements. Compliance of the contracted service levels is a top priority. This includes the availability of the IT service so that the customer can use the service. This also includes the quality of service for example in terms of performance in order to ensure both the usability and utility of the IT service following the requirements of the customer.

Service operation is also responsible for the management of the technology that is necessary for the service delivery and service maintenance. Supervising the performance by monitoring the IT service as well as processing events for example by analyzing log file data plays an important role. Out of the information obtained conclusions have to be made and applied within the incident and problem solution process. A continuous improvement of the IT services is based on the results of the monitoring and evaluation activities of service operation.

Seen from this point of view the focus of service operation becomes visible. Service operation is all about the management processes, the technologies and the responsible and executing people. The focus is also put on the stakeholders connected with the service. This is a key aspect in the service operation phase of the ITIL service lifecycle. [6] [S. 35ff], [10] [S. 93ff]

3 Research Methodology

3.1 Case Study Research

Work presented in this paper uses two main research methods: literature study to investigate the field of IT service management for SME (see section 4) and case study methodology.

In this paper, qualitative case study methodology was applied, which is an approach to research that facilitates exploration of a phenomenon within its context (e.g. an organization of social environment) using a variety of data sources. Baxter and Jack [11] recommend to consider case study design when the focus of the study is to answer "how" and "why" questions; the behavior of those involved in the study cannot be manipulated by the researches performing the study; contextual conditions have to be covered, since they are considered relevant to the phenomenon under study; or the boundaries are not clear between the phenomenon and context.

Case study research ensures that the phenomenon is not explored from one just one perspective, but rather a variety of "lenses" which allows for multiple facets of the phenomenon to be revealed and understood. In this context the case as such is "a phenomenon of some sort occurring in a bounded context", which is "in effect, your unit of analysis". Approaches to case study research seek to ensure that the topic of interest is well explored, and that the essence of the phenomenon is revealed. Most work in this area is contributed by Robert Stake [12] and Robert Yin [13].

Yin distinguishes between different kinds of cases studies including explanatory, exploratory and descriptive case studies. Table 2 is summarizing the most important characteristics of these kinds.

Table 2. – Different kinds of case studies (based on [11])

Case Study Type	*Definition*
Explanatory	This type of case study would be used if you were seeking to answer a question that sought to explain the presumed causal links in real-life interventions that are too complex for a survey or experimental strategies. In evaluation language, the explanations would link program implementation with program effects.
Exploratory	This type of case study is used to explore those situations in which the intervention being evaluated has no clear, single set of outcomes
Descriptive	This type of case study is used to describe an intervention or phenomenon and the real-life context in which it occurred

The case study described in section 5 is a descriptive case study. The “intervention” to be investigated are the processes and changes when introducing ITIL service operation and the “real-life context” is the department “IT & organization” at Nordex in Germany.

The case study literature recommends developing and using propositions. Propositions usually increase the likelihood that the researcher will be able to place limits on the scope of the study and increase the feasibility of completing the project. The more a study contains specific propositions, the more it will stay within feasible limits. Thus, we follow the advice and develop propositions expressed as research questions (see section 3.2).

Although used in many research contexts, case study methodology often is discussed quite controversy, which partly might be due to the fact that several kinds of case study can be distinguished. Criticism often reflects the opinion that findings from just one case cannot be generalized [14], p. 20] and thus are of low value for research. Flyvbjerg states that conclusions based on single case studies can be fully valid [15] [S. 220] which contradicts Abercrombie [16] opinion who is an opponent of using case study methodology. However, the primary objective of our case study is not to come up with general findings, but to offer insights how an ITIL implementation happens and why certain adjustments or interpretations of ITIL are made. With these findings we offer new experiences, which are a contribution to further research in the field.

3.2 Research Questions

The research questions presented in this section were investigated by combining the results from the literature analysis (section 4) and the case study discussed in section 5:

Research Question 1: What is the industrial practice of introducing ITIL in a medium-sized enterprise?

The literature analysis will look for approaches, methods and success factors of an ITIL introduction with specific focus on SME. Furthermore, the case study from Nordex will document and analyze the ITIL introduction with focus on the service operation part at Nordex. By combining results from literature and from the case study, we expect to contribute to an understanding of the industrial practice of ITIL implementation.

Research Question 2: How can the ITIL framework be adjusted to the demands of a medium-sized enterprise??

As shown in section 2, the ITIL reference model consists of five core areas with numerous processes, functions and roles. Due to this complex structure the question arises whether the framework as such is too big to be efficiently implemented in medium-sized enterprises. Implementation if ITIL usually has to happen in parallel to daily business processes in IT departments, i.e. the changes of processes and structures does not only challenge the medium-sized enterprise with respect to the resources needed but also from a coordination and transition perspective. This aspect will be investigated in the case study at Nordex.

Research Question 3: What experiences gathered in the case study at Nordex could possibly be relevant for other medium-sized enterprises?

The third research question focuses on lessons learned and experiences. We aim at identifying practices from Nordex relevant for other enterprises by recapitulating and analyzing the process of introducing ITIL service operation at Nordex.

4 Selected Results of the Literature Review

In this section answers to the research questions are given using state of the art literature review. The review of the literature is directed to the following questions:

- What does industrial practice of implementing ITIL look like?
- Does a uniform step-by-step implementation guidance exist which shows the necessary procedure of an ITSM implementation according to ITIL?
- What are the key points that must be considered in an ITIL implementation?

Subject of the literature review is both the official ITIL core publication of the OGC and other accompanying secondary literature. For successful ITIL implementations recommendations for SMEs to implement a scaled down ITSM landscape have been published. [17] presents adaptations of the ITIL reference model specifically for SMEs.

A huge chance to reduce the complexity of the ITIL framework lays in the combination of user roles that act in service operation processes and their single activities. The incident, request and problem management process – to take an example - could be merged to a hybrid process. Nevertheless it is important to keep in mind, that the problem management process becomes not only an extension of the incident management process.

To keep things short and simple in SMEs thus lowering administrative effort daily activities of IT operations regarding monitoring and control, event management, security management and access management could be combined into a single process of activities rather then forming single processes. Another approach for IT organizations is to offer self-help functions for customers. This can be achieved by providing FAQ-material, lists of known errors and by presenting the customer possibilities to work around those known errors. These measures will reduce the operating pressure oft he ITIL service desk and results in direct benefits to the IT organization.

Regarding the event management process and the recording of events an alarms of the IT systems it is proposed to prioritize what data becomes monitored. The more limited the resources available, the more the IT organization should focus on relevant and useful data. According to [17] it is useful to introduce the key user concept when there exists a variety of company locations. This reduces the workload for the IT service desk and provides users with a familiar single point of contact.

Procedures should also never be designed more complicated than necessary. You should present the views of the customers on incidents and not the views of the IT organization.

The research report by Stoller [18] distributed by the Info-Tech Research Group deals with the adaptability of the ITIL reference model to the special needs of SMEs. The report provides an approach to ITIL implementation in SMEs. For this purpose, a sequence of steps is proposed that leads a project within the necessary direction of flow. The sequence of steps [18] [S. 19f] consists of:

- Assess your needs,
- Train key people at the ITIL Foundation level,
- Create a tactical plan,
- Create a configuration management database (CMDB),
- Get Incident Management under control,
- Implement Problem and Change Management and
- Implement processes to address specific business requirements.

The results of the literature analyzed are in many cases presented in the form of critical success factors (CSFs). This common formulation of the results from the investigated area of research is in accordance to Rockhart [19] [S. 9] defined as an overview of those key points of a project, which must be met to achieve the desired objectives and results successfully.

In German literature Zarnekow presents the following CSFs of an ITIL implementation project [20] [S. 274ff]:

- The human aspect of reorganization projects,
- Top-management support,
- Project design aspects,
- Separation of operational and project work,
- Change management internally rather than by external consultants,
- Short project lifecycles,
- Process design aspects,
- Sustainability,
- Service-orientation,
- Quality management,
- Separation of operational and process design tasks and
- Tool support.

Combined with the results of Tiemeyer [21] [S. 104ff] Bick [7] [S. 31] developed his list of CSFs when implementing ITSM in organizations:

- Management support for the project together with strategically integration,
- Project management; Short subprojects to keep momentum; Communication of results and usefulness; Clear and realistic objectives; Project planning and controlling,
- Supporting organization culture; Establishing a service culture; Change in the attitude of employees; Alignment towards systems has to change to a customer-oriented service alignment,
- Change management,
- Expectation management,
- Qualified employees,
- Quality management and
- Quick wins to establish and support the ITIL initiative.

These listings are each detailed in the respective literature sources.

The literature review has shown that there exists no widely known standard process model for the implementation of ITSM according to ITIL. Many results from research and literature approach the topic of ITIL implementation from the direction of the CSFs. But they do give little detail on how the implementation should be done exactly.

An exemplary approach to ITIL implementation is presented in [22] [S. 10-32]. The following sequence of steps presents a guide to ITIL implementation to be carried out in the following order:

- Preparation of the project,
- Definition of the service structure,
- Selection and occupation of the ITIL roles,
- Analysis of current process situation (as is),
- Definition of the future process situation (to be),
- Definition of the process interfaces,
- Establishment of the process controlling,
- Detailed elaboration of the processes,
- Selection and implementation of application systems and
- Process implementation followed intensive training.

This approach has been analyzed in the case study in practice. Section 5 presents the results of the case study.

5 Case Study: ITIL Implementation at Nordex

Subject of the presented case study is the IT and organization (ITO) department of the wind turbine manufacturer Nordex. In the context of this case study the term “product-related IT” of Nordex is used. This includes every aspect of hard- and software regarding the single wind turbines and the wind farm network as found in the field.

The product-related IT differs from the so-called office-related IT in terms of the potential user. While an employee of the Nordex group is an internal customer of the office IT, the buyers or operators of the wind turbines form the specific customers of the product-related IT. This technology is separated from the office IT and is supported by the Nordex-internal Park Operations (POP) department.

The introduction of IT service management does apply to all Nordex ITO departments, not only to the POP department. Due to scaling reasons this case study focuses on the objective to introduce ITIL service operation processes to the product-related IT. Nevertheless the holistic approach of IT service management should not be forgotten. That is why in most parts of the case study it is necessary to include broader aspects of the ITO department and its processes.

The case study has been conducted in form of a descriptive case study for the period from the second quarter of 2011 to the beginning of 2012. Following a first ITIL training for selected employees on ITIL Foundation level began the delimitation of the project scope. It was developed, which objectives in implementing ITIL should be achieved in a first project stage. The deadline of this first stage consisted of an ISO certification date in February 2012. Special attention was given to measures regarding process quality and maturity that have been worked out in an internal IT audit. With the service operation processes the first implementation activities have been determined due to the fact that they have strong operation focus and build the basis for further implementation activities. It was determined to focus in the first project stage on the processes request fulfillment, incident management and problem management. Access management was integrated into the request fulfillment process.

After the acquisition of process flow diagrams in BPMN 2.0 format these diagrams were made accessible throughout the company-wide intranet system. In a next step key team members analyzed the process flow diagrams. They were customized according to special requirements from Nordex captured in workshops and modeling sessions. Step by step the processes request fulfillment, incident management and problem management were analyzed and customization needs were instantly applied to the process flow diagrams.

For documenting the service operation processes Nordex decided to only accept a shape-based Microsoft Visio format. To comply to all process flow diagrams in the Nordex-world all service operation processes had to be translated into this format. This step should ensure, that every employee could instantly understand the diagrams and no further training effort had to be made. Due to the fundamental differences of the internal notation specification and the format of the acquired process flow diagrams a complete remodeling in hour-lasting sessions followed as the next implementation step. The reference process model was used as template to save time. Special attention was paid to the assignment of the individual process activities to specific ITIL roles.

Parallel to the modeling and documentation of the ITIL processes in the ITO department a list of all IT services of the ITO area was created. A review process started once the processes were modeled in a first version and ITIL process trainings as well as ITIL service descriptions raised awareness among all ITO employees. Therefor all employees were consulted to estimate the processes and to suggest improvements.

The overall solution implemented at Nordex guides service user and service operation staff through the specified processes using a 3-layered approach, which is depicted in figure 2. The first level (left hand side) shows the organizational function of the help desk with the process implemented. From the customer perspective, the actual process of working on a request or an inquiry is shown as a “black box” only in order to reduce complexity. For the actual work in service operations, the “content” of the black box is specified in much detail for the various types of inquiries and requests possible. This is illustrated by two examples in the middle part of figure 2. Finally, there are instructions and checklists for all activities comprising the process. The target groups for these instructions are the organizational roles performing the activities.

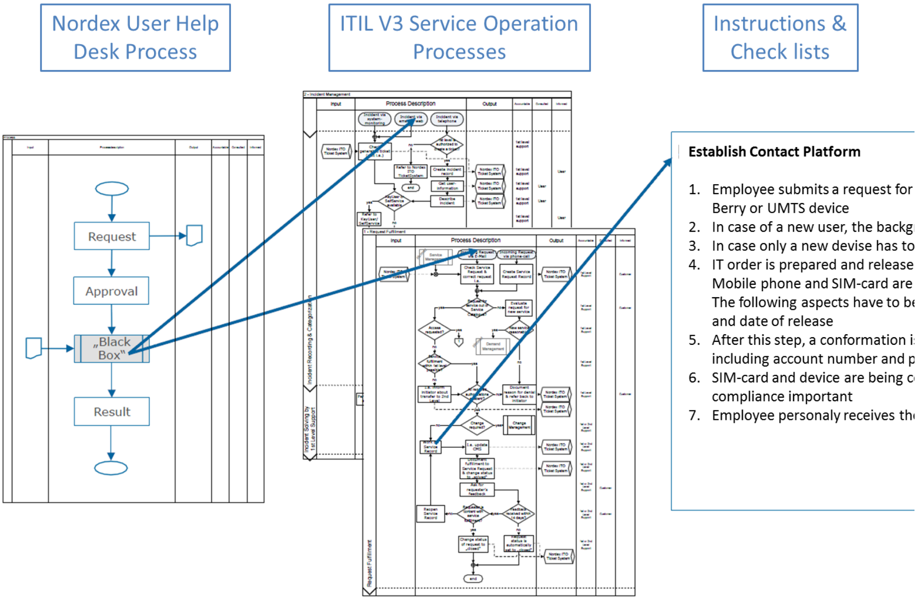


Fig. 2. Levels of the process landscape for IT service operations at Nordex

6 Conclusion

IT service management based on the Information Technology Infrastructure Library (ITIL) is increasingly used in industry and public authorities. Service operation is one of the five ITIL core processes encompassing service delivery and management of service quality. In this context, the objective of this paper is to contribute to an understanding of the industrial practice of ITIL implementation with a focus on medium-sized enterprises. Although ITIL offers a best practice reference model, each ITIL implementation usually is one-of-a-kind project. Existing literature in the field offers a lot of general advice for ITIL implementation with focus on large enterprises, but there are hardly any publications focusing on SME needs in this sector. Based on a state-of-the-art analysis and using case study methodology as main research approach the paper showed how an IT service management implementation in a medium-sized enterprise works. More concrete, the case study focused on the product-related IT of wind park operations at wind turbine manufacturer Nordex.

Research Question 1: What is the industrial practice of introducing ITIL in a medium-sized enterprise?

The literature analysis showed that there is no standardized and generally accepted procedure for implementing ITIL. Neither the official ITIL publications nor the complementary ITIL books include a step-by-step or cookbook-like procedure for implementing ITIL. In some of the proprietary materials from consultancy companies, general activities for project-based ITIL-implementation are defined, which are not very precise and detailed. The ITIL implementation within the case study showed that these general project activities are valid and useful, but have to be interpreted for the case at hand.

Research Question 2: How can the ITIL framework be adjusted to the demands of a medium-sized enterprise? Based on the lessons learned from the Nordex case study, implementation of service operation processes based on ITIL is fully feasible. ITIL offers many options to adjust and tailor IT service management functions and processes to the existing IT-organization and to the resources available in a medium-sized enterprise.

Research Question 3: What experiences gathered in the case study at Nordex could possibly be relevant for other medium-sized enterprises? The case study carefully documented the processes and organization structures created during ITIL implementation at Nordex and show a real-world example how ITIL can be adapted for medium-sized enterprises. What experiences from Nordex can be transferred to other enterprises very much depends on the needs and context of the enterprise in question. We propose to consider the experiences collected in the case study as reservoir and source of inspiration.

References

[1] Carr, N.G.: IT doesn't matter. IEEE Engineering Management Review 32(1) (2004)

[2] Kuschke, M.: IT-Servicemanagement - ein Überblick (April 27, 2005), `http://www.hitforum.de/itil-itsm.pdf`, `http://www.hitforum.de` (accessed January 2012)

[3] Schmidtbauer, P.: Industrielle Praxis der ITIL-Einführung am Beispiel von Service Operation Prozessen der Produkt-IT bei der Nordex SE. Universität Rostock, Rostock (2012)

[4] Johannsen, W., Goeken, M.: Referenzmodelle für IT-Governance, 2nd edn. Dpunkt. Verlag, Heidelberg (2011)

[5] Boehmann, T., Krcmar, H.: Grundlagen und Entwicklungstrends im IT-Servicemanagement. In: IT-Service Management. Dpunkt. Verlag, Heidelberg (2004)

[6] Cannon, D., Wheeldon, D.: ITIL Service Operation. The Stationary Office, London (2007)

[7] Bick, M., Boergmann, K.: IT Service Management Support System (IT-S3) - wertorientierte Einführung und Etablierung von IT-Servicemanagement. In: Wertorientiertes IT-Servicemanagement. Dpunkt. Verlag, Heidelberg (2008)

[8] Addy, R.: Effective IT Service Management. Springer, New York (2007)

[9] Krcmar, H.: Informationsmanagement, 4th edn. Springer, Heidelberg (2005)
[10] Office of Government Commerce, The Official Introduction to the ITIL Service Lifecycle, 2007th edn. The Stationery Office, London (2007)
[11] Baxter, P., Jack, S.: Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. The Qualitative Report 13(4), 544–559 (2008)
[12] Stake, R.E.: The Art of Case Study Research. Sage Publications, Inc., Thousand Oaks (1995)
[13] Yin, R.K.: Case Study Research: Design and Methods, 4th edn., vol. 5, p. 240. Sage Publications, Inc., Thousand Oaks (2009)
[14] Gerring, J.: Case Study Research - Principles and Practices. Cambridge University Press, New York (2007)
[15] Flyvbjerg, B.: Five Misunderstandings About Case-Study Research. Qualitative Inquiry 12(2), 219–245 (2006)
[16] Abercrombie, N., Hill, S., Turner, B.S.: Dictionary of sociology. Penguin, Harmondsworth (1984)
[17] Taylor, S., Macfarlane, I.: ITIL V3 Small-scale Implementation, 2009th edn. The Stationary Office, London (2009)
[18] Stoller, J.: Adapting ITIL to Small- and Mid-Sized Enterprises. Info-Tech Research Group, London (2006)
[19] Rockhart, J.F.: A Primer on Critical Success Factors. In: Bullen, C.V. (ed.) The Rise of Managerial Computing: The Best of the Center for Information Systems Research. Center for Information Systems Research, Massachusetts Institute of Technology, Cambridge (1981)
[20] Zarnekow, R., Hochstein, A., Brenner, W.: Serviceorientiertes IT-Management, Erstauflage. Springer, Heidelberg (2005)
[21] Tiemeyer, E.: IT-Servicemanagement Kompakt, 1st edn. Spektrum Akademischer Verlag, Heidelberg (2005)
[22] IT Process Maps GbR, “ITIL-Implementierung mit der ITIL-Prozesslandkarte V3 - Ein Projektleitfaden in 10 Schritten.” IT Process Maps GbR, Am Hörnle 7, 87459 Pfronten (2011)

Mobile Cloudable Applications: A New Way of Distributing Mobile Tasks into the Cloud

Sebastian Presecan[1] and Nicolae Tomai[2]

[1] iQuest, Cluj-Napoca, Romania
Sebastian.presecan@gmail.com
[2] Babes Bolyai University of Cluj-Napoca, Romania
tomai@econ.ubbcluj.ro

Abstract. Smartphones are nowadays ubiquitous. Smartphone owners are using mobile applications as a replacer for desktop applications. Due to the limited resources and due to the challenge of being a real replacer for personal computers, mobile devices should be able to run applications that are distributing their tasks for execution into the cloud. The present paper proposes a pervasive framework that combines the mobility of smartphone applications and the execution capabilities of the cloud. Cloud distribution reduces the mobile phone power consumption and also improves the execution of long running tasks. By using the AOP (Aspect Oriented Programming), the distribution into the cloud is transparent to the application developer. The proposed framework proves itself as a viable solution for implementing mobile applications that could distribute themselves into the cloud in order to improve users' experience.

Keywords: mobile-cloud computing, mobile computing, cloud computing, mobile cloud, pervasive computing.

1 Introduction

Pervasive computing aims at helping applications run in mobile environment by adapting them to the environment changes. All pervasive applications should be able to run on mobile devices that have a limited computational power and lack of good Internet connectivity.

Cloud Computing is one of the latest additions to the family of distributed computing paradigms. Cloud computing is a paradigm which shifts the location of computing infrastructure to the network in order to reduce the costs associated with the management of hardware and software resources. [1]

Mobile devices have limited resources but nowadays they are used in different situations as replacements for personal computers. There is an urge of resolving the lack of resources in the mobile space by using services that reside on the cloud. From these perspectives, a new paradigm has developed through mixing the mobile and cloud computing: mobile cloud computing. It was defined on 5 March 2010, in an entry made to the Open Gardens blog, as *"the availability of cloud computing services*

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 136–147, 2013.

in a mobile ecosystem. This incorporates many elements, including consumer, enterprise, femtocells, transcoding, end-to-end security, home gateways, and mobile broadband-enabled services." [2]

The purpose of the present work is to build a middleware that can be used for distributing user tasks that are performed by mobile application, into the cloud.

The current study starts by identifying the challenges that need to be fulfilled by frameworks which aim at distributing mobile applications into the cloud. We have analysed a series of existing frameworks which were built in order to solve various types of problems. The system architecture is subsequently presented in details, thus showing how the framework fulfils the challenges. The results of implementing such a framework in order to complete a concrete long running mobile tasks are presented in the final section of this study.

2 Challenges

In order to make a clear distinction between normal mobile applications (which run on mobile devices and use resources from those devices or from the Internet) and new mobile applications that are performing user requests by splitting the execution into small tasks, we propose to define the last type of mobile applications as*: "mobile cloudable application". A mobile cloudable application is a mobile application that is able to migrate into the cloud in order to perform the user tasks in a better way, by using the vast computational power available into the cloud.*

Any framework used in developing mobile cloudable applications which are used within the pervasive environment should aim at fulfilling the following requirements:

- *enable development of context-aware applications* – mobile users belong to different contexts: local, social, and behavioural. The pervasive applications use these contexts in order to understand users' habits and to adapt the task execution to users' preferences. In mobile cloudable applications, users' requests are split into tasks that are being executed on the cloud. But still, user context should be available through the development framework, enabling the application to know about users' preferences.
- *react to environmental changes* – during their occurrence, the application should react to the environmental changes. Mobile cloudable applications are those applications that sense the modifications within the environment and find the best distribution strategy to fulfil users' tasks with the lowest possible consumption of the mobile device resources. The mobile cloudable applications embrace the "react-and-adapt" strategy in the scope of adapting their execution to the environment changes.
- *transparent distribution of the execution into the cloud* – the application developer should not be constrained by using custom libraries in order to manage the service distribution. The tasks distribution into the cloud should be achieved by the usage of the framework and should be transparent to the application developer. Developing a mobile cloudable application should not differ from the development of normal mobile application. The distribution into the cloud should not concern the application developers, they should only be focused on implementing the

application functionality, the cloud distribution should be ensured transparently by the development framework.

- *application and data mobility* – the code written for the mobile application should be migrated transparently into the cloud. The code and data should migrate dynamically at runtime and should be requested in order to use optimally the resources of the mobile device.
- *application elasticity* - cloud services will be used on demand only when the local resource could not meet the expected QoS. The development framework should measure the QoS and should react on the lack of resource by distributing the execution over the cloud. The application distribution over the cloud should be decided at runtime and should not be done statically. The development framework should support the distribution and the aggregation of results. The usage of the cloud is not a must for running the application. Not all tasks need to be distributed into the cloud. The network usage and latency could increase the time for performing certain tasks requested by the user. [3] Mobile cloudable applications could run without Internet connection, but by distributing the tasks execution over the cloud on demand, they are using the system resources more efficiently.

Mobile cloudable applications are meant to use optimally the mobile device resources and when the device is lacking in computational power or when the execution of tasks on cloud would be more optimal, they are distributing the tasks execution transparently into the cloud.

3 Related Work

From the perspective of application providers, cloud computing is perceived as a vast alternative to assure service delivery. Still, the execution on the cloud could not benefit of device specific context (location, social). From this perspective, the cloud could be considered as an execution extension for limited resources devices.

There are several ways of achieving this particular goal: running clones of mobile applications into the cloud, using cloudlets as substitution for the real device [4]. Zhang and his colleagues have proposed a new approach for performing mobile tasks on the cloud. They have split tasks into so-called “weblets”. A “weblet” is a task that could be run either on a mobile device or on the cloud. The “weblet” could be designed in order to perform complex tasks. The location they are executed in is decided at runtime based on the device available resources. The solution has the advantage of adapting the execution distribution to the resources availability. Still, there are several disadvantages. The user tasks that could be executed are limited, due to the requirement of using certain tasks that could be modelled as weblets. These weblets are web applications which are able to run either on a device or on the cloud. Thus, there is a need to have webcontainer deployed on the device. Due to the enforcing of the weblets composition in order to perform user tasks, application developers have to use a custom SDK for implementing different applications. This could be a major drawback in developing applications that could run on different devices.

The mobile application needs to be written for different platforms, using different programming languages. All these features make the mobile-cloud environment a very heterogeneous one and increase the costs of porting the application from one mobile platform to another, as well as the moving of cloud applications from one IaaS to another. Manjunatha and his colleagues have proposed a more abstract approach towards the implementing of mobile-cloud application [5]. The application should be developed using DSL for modelling the user interaction. By using some code generation scripts, this application could be deployed in a hybrid environment: mobile and cloud. The generation scripts could be updated in order to generate code for different mobile and cloud platforms. The focus of any application developer should be the solving of users' problems by using DSL and not the solving of technical issues caused by different mobile platforms. Still, the solution is limited to the implementing of simple users' activities and it could not be used for implementing large-scale applications.

The migration of execution from the local machine into the cloud has been a major research challenge for many years. There are, nevertheless, several solutions such as agents' migration, VM cloning, thread migration and address space migration. They are all offering particular solutions to specific problems, but they also present a lot of drawbacks. Ricky and his colleagues have proposed an innovative solution called Stack-On-Demand (SOD) [6]. SOD proposes the migration of the execution stack. The solution relies on the JVM capabilities. The solution is applicable for applications running on similar JVM. In case of mobile devices, the solution has small chances of being applied. The JVM for mobile devices differs from the JVM implemented for servers. Thus, stack migration will not be possible. The solution is error prone since it is based on a low implementation level of the JVM. Still, it has the advantage of transparency – application developers are not aware of code migration during the implementation and it also provides elasticity – the execution stack is migrated at once and if there are dependencies to an external code, the code is loaded on demand.

The migration of the execution into the cloud has the advantage of lowering the computing power used to perform user tasks. Still, the capabilities of mobile devices could not be used into the cloud: context-awareness is missing in cloud environments. Hyun Jung and his colleagues have proposed a solution for designing context-aware mobile architecture integrated with the cloud [7]. The application consists of 3 different layers: user layer (the user application), agent layer (defines all mobile services used to monitor the QoS and to perform calls to the cloud) and the service layer (cloud services able to substitute calls to the mobile platform). The agent layer is responsible with the monitoring of the QoS of the mobile device and with decision-taking features in what the distributing of user tasks on the cloud is concerned. User requests are performed by decomposing the requests into tasks that could be performed either on local device or into the cloud. Still, application developers should know which services could be replicated on the cloud. The code replication is static. Thus, each time a new service is added, the services have to be implemented both on the cloud and on the mobile device.

Combining mobile application with webservice has become a de-facto standard in the past few years. The usage of SOA for integrating mobile application with the cloud has been proposed by Wang and his team [8]. Meanwhile, a lot of mobile applications are implementing this approach. In order to implement this solution, application developers must wisely decompose their application responsibility and use certain cloud services to fulfil user tasks. The solution lacks in deployment transparency and does not provide elasticity. In order to implement a new functionality that needs to be executed on the cloud, a new webservice has to be created and deployed on the cloud. The application not being able to migrate from mobile to cloud, forces developers to adopt a static configuration of the services deployment.

4 Mobile Cloudable Applications

4.1 High Level Architecture

In order to achieve the above mobile cloudable application challenges, the framework architecture described in Figure 1 could be used for developing a new generation of mobile applications.

The main components of the proposed architecture are:

- *Mobile application* – the application is developed using the mobile platform specific framework and programming languages: Java for Android, Objective C for iPhone/iPad, C# for WP7/8. The mobile application interacts with users by using a user interface that is specific to the target mobile device. In order to fulfil user requests, the mobile application uses different services located on mobile devices. The services are either system services or custom services. The services perform specific tasks and they are combined by the application so that they implement complex requests received from the mobile user.
- *Local services* – are implementing granular task used to fulfil user requests. The services could run on local mobile devices, but they could also be run on cloud. The services should implement the pure business logic and should not depend on the user interface. The services could depend on other custom services or on the system services. The application could combine different services in order to implement the operations requested by the end user.
- *Distribution service* – is one of the core components of the architecture. It monitors the QoS of the system and provides the appropriate implementation for a given local service at runtime. The Distribution Service is not being used directly by the application. The Distribution Service intercepts calls to the Local Service, analyzes the QoS and if needed, distributes the call into the cloud. If the Local Service can not perform the user requested operation due to the lack of resources or due to the complexity of the operation, the Distribution Service searches for a “Rider” that connects the local service with the cloud counterpart. If the Rider does not exist, it creates one and it uses the Distribution Service from the cloud to install the Local Service on the cloud. The Distribution Service communicates with the Distribution Service from the cloud in order to manage the lifecycle of the Riders.

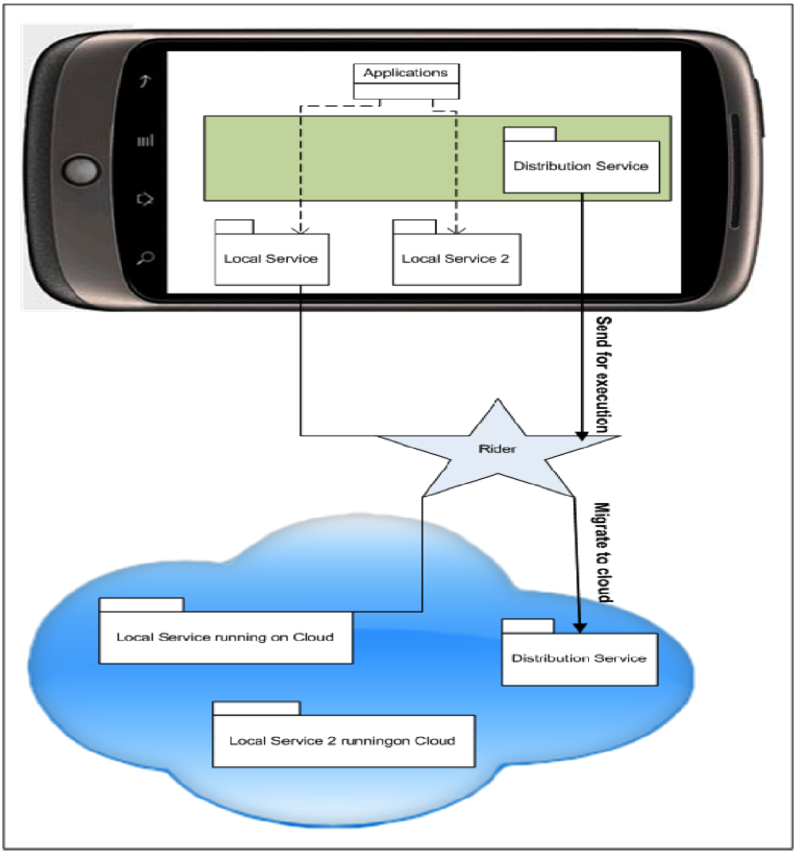


Fig. 1. - Mobile cloudable framework architecture

- *Rider* – the Distribution Service for migrating Local Service into the cloud is using it. The Rider contains the running data needed for the execution and the implementation of the local service. The Rider once installed on the cloud will return to the Distributed Service a smart-proxy to the corresponding service running into the cloud. The smart proxy is going to be used by the application as a replacer for the local service. The Rider serves as a data and code migration container and is used to support application mobility.

The Distribution Service and the Rider support the application mobility and application elasticity. The Distribution Service is constantly monitoring the system QoS and delegates, if be the case, the calls to the Local Service located on the cloud.

The Distribution Service intercepts the calls to the local service. From this perspective, the proposed architecture treats the application distribution through the cloud as a crosscutting concern. The cloud distribution is separated from the development of the mobile application. The best way of implementing such architecture is to use the innovative Aspect-Oriented Programming Paradigm.

4.2 Use AOP for Distribution

"Aspect-Oriented Programming (AOP) complements Object-Oriented Programming (OOP) by providing another way of thinking about program structure. In addition to classes, AOP gives you aspects. Aspects enable modularization of concerns such as transaction management that cut across multiple types and objects. (Such concerns are often termed crosscutting concerns.)" [9]

As any other programming paradigm, AOP promotes the encapsulation of concerns into separate, independent entities by providing abstractions (e.g. procedures, modules, classes, methods) that can be used for implementing and composing these concerns.

The primary goal of the AOP paradigm is to provide a mechanism for modularizing crosscutting concerns. Through the modularization of crosscutting concerns, AOP makes it possible to avoid code scattering and code tangling [10].

The distribution into the cloud that is intended to enhance the mobile application lack of computational power, could be considering a crosscutting concern. The distribution of the execution over the cloud should not be the main focus of developing a cloudable mobile application. The distribution could still be applied to multiple services without a direct dependency between the services implementation and the distribution mechanism. The distribution is thus a crosscutting concern that could be easily implemented by using an AOP framework.

In terms of AOP, the crosscutting concern is the distribution of calls to the business services into the cloud. The Advice is the implementation of the application mobility which enables the migration of the business logic into the cloud. The Advice should be implemented independently of the business logic that needs to be migrated. The Advice implementation should also take care of the code replication and deployment into the cloud. The PointCut could be the execution of certain methods of local services. The PointCut could be defined statically by the usage of metadata or dynamically, during the execution.

4.3 Implementation Details

An implementation overview for achieving cloudable features is described in Figure 2. The main classes for such framework architecture are:

- *Activity* – the class implementing user interface. It is used to collect user inputs and display information about the execution to the users. The Activities are part of the mobile development framework and they are specific to different mobile platforms.
- *LocalService* – a service that should contain the business logic needed to implement the required user tasks. In order to ensure the cloud distribution, it is mandatory that the implementation of the service is not dependent on user interface. The LocalService should depend on another system or local service or on other portable components. The LocalService should be used from Activities in order to perform user tasks. The LocalService will send back a notification to the Activity about the change in the status of tasks execution.

- *DistributionService* – manages the distribution interceptors. The implementation of the DistributionService depends on the mobile platform and it should be done on the system level, leveraging the possibility to perform the interception of all calls to the local services. The DistributionServiceInterceptors are registered to the DistributionService, either statically or dynamically. The way these interceptors are executed and the manner in which the Activities are injected with the proxies of the real LocalServices depend on the AOP framework which is used to provide method-intercepting support. The distribution service is also managing migration of the rider. The DistributionService contains all the logic needed to deploy the Rider remotely into the cloud.
- *DistributionInterceptor* – the core class of the cloudable framework. It intercepts the calls to the LocalService and based on the QoSStrategy, it decides whether the calls should be executed by the LocalService from the mobile device or they have to be distributed into the cloud. It creates the Rider in order to prepare the task execution into the cloud, if needed. If based on the QoSStrategy, it decides that the user tasks have to be executed into the cloud and there is no Rider already created, a new Rider is going to be set up and the DistributionService will be in charge with moving the Rider into the cloud.
- CloudDistributionService – responsible for loading the rider into the cloud. It installs the LocalService into the cloud and also loads all the needed classes from the Rider container. Moreover, it creates the proxy to the cloud LocalService that passes back to the Rider in order to be used by the Activities running on the mobile device.
- RiderLoader – a special classloader which is able to load the LocalService contained into the Rider into the cloud. The implementing of RiderLoader depends on the technology that is used for implementing the cloud.
- UserContext – defines user context data. It is used by the LocalService to get the user information. The user context is filled-in by the DistributionServiceInterceptor before sending the request to the cloud LocalService. The UserContext contains all the information that could be retrieved on the mobile device, with regard to the current user.
- QoSStrategy is used for monitoring the parameters of mobile environment. Based on different values, the framework could decide whether the tasks need to be executed locally into the cloud. The QoSStrategy could implement the pervasive reaction framework [11] in order to monitor and react to the environment changes. The framework chooses the best way of performing user tasks in order to maximize the usage of the mobile device and offer best mobile experience to the user.

Having a framework which implements the architecture described above will ease the implementation of the mobile cloudable application. The Rider container ensures both the application and data mobility. The LocalService has to be implemented once and could be run anywhere on the cloud. The migration to the cloud is transparent for the application developer. Thus, the application developer is focused on implementing user tasks and not on solving distribution issues. The distribution to the cloud is a crosscutting concern that is implemented independently of the tasks that need to be performed either locally or on the cloud.

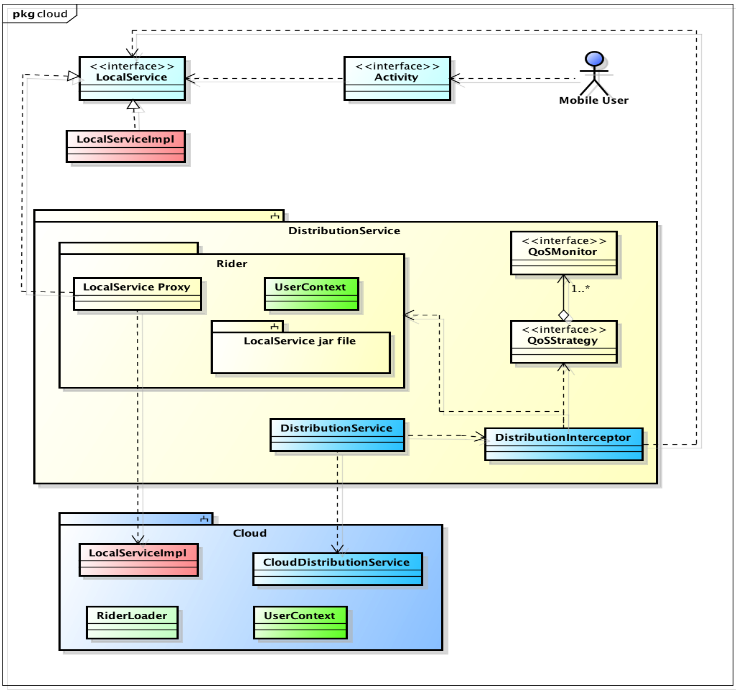


Fig. 2. – Detailed design

5 Results

In order to check the feasibility of the architecture described above, the following scenario has been considered:

Many people use different websites to store photos. There is also a tremendous request towards enabling users to view these photos on their mobile phones.

Loading photos of several megapixels on a mobile phone is a problematic aspect, due to the limited resources that mobile phones have access to. An 18 MP photo has an average file size of about 4.5 MB. For a mobile phone, displaying an 18 MP picture could be a big challenge in terms of memory and computational power used for such a task. In order to be displayed on mobile phones, photos first need to be scaled to a resolution that is closer to the screen resolution of the mobile device. This specific functionality could be used to check the way a cloudable application is performing its specific tasks.

This section of the current work displays the execution results of running the above-described application both locally and into the cloud. The application has been developed using Android 4.1 and was installed on a Nexus 4 device. RoboGuice [12] was used as an AOP framework. The server is a shared VM with 4 GB of memory, running into the Amazon EC2 cloud [13]. The mobile devices have been connected to the Internet by using a WIFI hotspot (100 Mbit).

In order to run the photo scaling process, the following steps have been performed:

- download a high resolution picture(18 MP) from web site – E.g: panoramio
- scale the picture to match screen resolution (1280 x 768)
- display photo on the mobile phone screen.

In case of the usage of cloud environment, the execution steps are the following:

- send the RIDER into the cloud
- download a high resolution picture (9 MB) from website into the cloud app
- scale the picture into the cloud
- download the small resolution picture into the mobile phone
- display the picture on the mobile phone screen.

Table 1 displays the execution times for performing the scaling of a 18 MP picture downloaded from website. As one can infer, the first call for the cloudable application is longer than the average call for the local version. This is due to the need of deployment and the installation of the Rider into the cloud.

Table 1. - Execution results for 1 scaling call

Execution time(ms)					
Run	1	2	3	4	5
Local	9324	9005	8700	9004	8700
Cloudable	12435	6004	7004	6500	7040

Once the Rider is installed into the cloud, the average performance of the cloudable application is better than the performance of the local application. In order to have a better overview upon the results, the following scenario has been executed several times: record the average execution time for scaling 20 photos.

Table 2. - Average execution time

Average Execution time for bunches of 20 scale calls (ms)					
Run	1	2	3	4	5
Local	8900	8100	8700	9400	9020
Cloudable	7200	7400	7300	8001	7023

As shown in Table 2, the average execution time in the case of the cloudable application is shorter with 2 seconds compared with the local application. A detailed analysis of causes determining this difference revealed the following factors:

- computational power difference: the cloud shared VM is much more powerful than the one displayed by the mobile device.
- available memory: the photo scaling process is a large memory consumer. During the scaling process, the garbage collector has been intensively used to free unused memory, with negative impact on performance.
- size of the downloaded file: downloading a 4.5 MB instead 1 MB over the WIFI connection decreased the performance of the local application.

The evaluation results have proven that the cloudable application is a viable solution for distributing user tasks that consume device resources into the cloud. The latency of distributing and installing the Rider into the cloud is overcome by a better performance of the cloud environment.

6 Conclusions and Future Work

Mobility is one of the key features of any pervasive application. The pervasive application must be able to run in different environments and even more, it must be able to be deployed seamlessly on different environments without user intervention. The current research proposes a solution for the migrating of mobile applications into the cloud without user intervention and in a way that is transparent to the application developer.

In what performance is concerned, resources available within the cloud cannot be compared with the resources available on mobile devices. Since mobile devices are more and more used as replacers for desktop computers, the tasks that need to be executed on such devices are becoming more and more complex. The architecture described within the current study proposes a seamless solution for executing user tasks within the cloud, when local mobile resources are not sufficient in offering a good user experience. The proposed solution envisages the running on the cloud of the application that has been developed especially for mobile devices. The proposed framework comprises the implementation for seamless integration with the mobile application, the distribution and the deployment into the cloud of mobile applications.

The proposed solution has been evaluated from a performance point of view and it proves to be a viable alternative for moving the mobile applications into the cloud. Moving the applications into the cloud also brings the advantage of better performance within the cloud environment. The framework is based on the innovative AOP paradigm that helps a seamless integration of the cloud distribution into mobile applications. Application developers are focused on implementing the application logic and they should not care about the application distribution into the cloud. The AOP advices are used to make mobile applications cloudable. The logical services are thus replaced with proxies to the cloud counterpart at runtime. By following some design guidelines for defining the components dependencies, logical services implemented for Android could be easily deployed into the cloud environments.

The current research is going to be extended to a more abstract level in order to have an architecture that could be virtually applied cross-mobile platforms.

References

1. Vaquero, L.M., Rodero-Merino, L., Caceres, J., Lindner: A break in the clouds: towards a cloud definition. SIGCOMM Comput. Commun., 50–55 (2008)
2. Cox, P.A.: developer Works (2011), `http://www.ibm.com/developerworks/cloud/library/cl-mobilecloudcomputing`

3. Satyanarayanan, M., Bahl, P., Caceres, R., Davies, N.: The Case for VM-Based Cloudlets in Mobile Computing. In: Pervasive Computing (October-December 2009)
4. Zhang, X., Kunjithapatham, A., Jeong, S., Gibbs, S.: Towards an Elastic Application Model for Augmenting the Computing Capabilities of Mobile Devices with Cloud Computing. In: Mobile Networks and Applications. Springer Link (2004)
5. Manjunatha, A., et al.: Power of Clouds in Your Pocket: An Efficient Approach for Cloud Mobile Hybrid Application Development. In: Cloud Computing Technology and Science. IEEE (2010)
6. Ricky, K., Ma King, K., Lam, T., Wang, C.-L., Zhang, C.: A Stack-on-Demand Execution Model for Elastic Computing. In: 39th International Conference on Parallel Processing. IEEE, San Diego (2010)
7. La, H.J., Kim, S.D.: A Conceptual Framework for Provisioning Context-aware Mobile Cloud Services. In: 3rd International Conference on Cloud Computing. IEEE, Miami (2010)
8. Wang, Q., Deters, R.: SOA's Last Mile-Connecting Smartphones to the Service Cloud. In: Cloud Computing. IEEE, Bangalore (2009)
9. Interface 21, Spring Framework. Aspect Oriented Programming with Spring, `http://static.springsource.org/spring/docs/2.0.x/reference/aop.html`
10. Monnox, A.: Rapid J2EE™ Development: An Adaptive Foundation for Enterprise Applications. Prentice Hall (2005)
11. Presecan, S., Tomai, N.: Distributed Context Provisioning and Reaction Middleware. In: Intelligent Computer Communication and Processing. IEEE Computer Society Press (2009)
12. Google. RoboGuice, `http://code.google.com/p/roboguice/`
13. Amazon EC2, `http://aws.amazon.com/ec2/`

The System Conception of Investigation of the Communication Security Level in Networks

Henryk Piech and Grzegorz Grodzki

Czestochowa University of Technology,
Dabrowskiego 73, 42201 Czestochowa, Poland
h.piech@adm.pcz.czest.pl, grzegorz.grodzki@icis.pcz.pl

Abstract. Communication processes have to be observed because of the possibility that different kinds of threats occur in the processes of exchanging information in a network. Those threats are connected with: decryption possibility, losing jurisdiction, believing and freshness of information, message interception by intruders, etc. We monitor the communication protocol during its operation. Standard security attributes (as proposed by Barrows, Abadi, Needham and others) have been introduced to analyze the chosen aspects of security. We also employed a standard set of rules which interrelate parts of communication operations with security aspects. Our previous research introduced a system that investigates security related issues. It could be employed for auditing purposes and/or to predict failures given different kinds of communication threats. In this paper the security analysis is continued in the direction of building the model of dynamic modification of chosen factors (adequate to security aspect) with prognosis possibility.

Keywords: protocol logic, probabilistic timed automata, communication security.

1 Introduction

Information is sent in the form of a message according to protocol systems that should guarantee: encryption safety, sufficient belief level, protection against intruders, and the freshness of information elements [1],[2]. Usually, many mutually interleaved protocols are used in networks [3],[4]. Obviously, information refers to a different group of users (usually, they are grouped in a pair). Therefore, security analysis will be referred to those groups and they will form the creation basis of the so called main security factor [1],[5]. Another main factor can include a set of messages, the public key, secret, nonce, etc. Security attributes may consist of: the degree of encryption, key and secret sharing, believing in sender or receiver, believing in user honesty, and the level of a message or nonce freshness [1]. M. Kwiatkowska indicates that security attributes are presented in the figure of probability parameters [6],[7]. This form is smart and very convenient. Therefore, we base our proposition on the transformation possibility

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 148–159, 2013.

of time attributes into probability characteristics. Apart from rule and time influences, we also regard intruder threats. The influence on security attributes is realized with the help of correction coefficients that also have a probabilistic form (according to the adopted approach). The above-mentioned rules are set on the basis of conditions that form actions which really appear in communication operations. In addition, one may observe that many works consider the division of protocols into operations and operations into actions [1],[4],[5],[8]. In our model we also exploit the so called tokens with binary character. A token directly appoints the secure or threat attribute level depending on the relation to a given security threshold. This type of approach improves the assessment reaction with respect to security state changing and helps in the estimation of distribution probabilities that lead to next stages and thereby to one of the forms of prognosis creation (presented in the further section)[10],[11],[12]. The proposed system, pertaining to the investigation of communication runs, can be easily realized with mutually converted probabilistic timed automata (PTA) and colored Petri nets (proven and shown in the works of M. Kwiatkowska [6],[7],[9]). These characteristics guarantee the effective realization of a parallel model.

2 Communication Protocol Actions and Attribute Grammars

The problem consists in the definition and recognition of actions. Rule conditions should be directly connected with actions, whereas their conclusion ought to be associated with attributes. The transformation of the run operation into an action is the first stage of action recognition. Each operation is divided into actions which are adequate to their function. The action definition is as follows:

Definition 1. A tuple $\{S, R, K, M, N, Ch, Ad\}$ is an action ac_v which may contain information about the sender (S), receiver (R), message (M), the character of dealing (Ch), additional information - e.g. secrets etc.(Ad).

- The sender is represented by one user or a set of users $S = \{s(1), s(2), ..., s(ls)\}$,
- The receiver is represented by one user or a set of users $R = \{r(1), r(2), ..., r(lr)\}$,
- Sender and receiver create a group of users that can be limited for the excluding possibilities of intruder activity.

Actions are both a part of protocol operation and influences on security attributes. The set of security attributes is defined by rules (their arguments and conclusions). In order to present the same example of rules, we should define the set of predicates of the communication BAN logic [4]:

$A \leftrightarrow^K B$ - users A and B communicate via shared key K,
$\rightarrow^K A$ - user A has K as its public key,
$A \Leftrightarrow^Y B$ - users A and B share Y as a secret,
$\{X\}_K$ - the message X encrypted by key K,
$< X >_Y$ - the message X with a secret Y attached,
$A| \equiv X$ - user A believes the message X,

$A \triangleright X$ - user A sees the message X,
$A \triangleleft X$ - user A sends the message X once,
$A| \Rightarrow X$ - user A has jurisdiction over X,
$\#(X)$ - the message is fresh.

Let us try to define the set of actions and attributes. In order to achieve this aim, we exploit the rules based on BAN logic:

1. Authentication rule – type I:
   if $(A| \equiv ((A \leftrightarrow^K B),\ A \triangleright X_K)$ then $(A| \equiv (B \triangleleft X)$.
   The rules can be interpreted as follows: if A and B shared key K and A sees message, then A believes that this message is from B.
2. Authentication rule – type II:
   if $(A| \equiv (\rightarrow^K B), A \triangleright X^C_{K-1}, C \neq A)$ then $A| \equiv (B \triangleleft X)$.
   The rules can be interpreted as follows: if A knows B's key, then A recognizes signed message from B.
3. Authentication rule – type III:
   if $(A| \equiv (A \Leftrightarrow^Y B), A \triangleright X^Y)$ then $(A| \equiv (B \triangleleft X)$.
   The rules can be interpreted as follows: if A and B shared secret Y and A sees message with attached secret Y, then A believes that this message is from B.
4. Nonce rule:
   if $(A| \equiv \#(X),\ A| \equiv (B \triangleleft X))$ then $A| \equiv (B| \equiv X)$.
   The rules can be interpreted as follows : if A believes that X is "current" and that B said X, then A believes that B believes X.
5. Jurisdiction rule:
   if $(A| \equiv (B| \Rightarrow X),\ A| \equiv (B| \equiv X)$ then $A| \equiv X$
   The rule can be interpreted as follows: if A believes that B has jurisdiction over X and A believes that B confirms X then A believes X.
6. Vision rule – type I:
   if $(A| \equiv (A \leftrightarrow^K B),\ A \triangleright \{X\}^C_K, C \neq A)$ then $A \triangleright X$.
   The rules can be interpreted as follows: A can see through encryption on a shared symmetric key, provided that the encryption was done by a user other than A itself.
7. Vision rule – type II:
   if $(A| \equiv (\rightarrow^K A), A \triangleright X^C_K, C \neq A)$ then $A \triangleright X$.
   The rules can be interpreted as follows: A knows its secret key, so it can decrypt message encrypted with public key.
8. Freshness rule:
   if $\#(X)$ then $\#(X, Y)$.
   The rule can be interpreted as follows: if X is fresh then $X \wedge Y$ is also fresh.

The idea of operation decomposition into actions may be presented on the basis of a simplified example. Obviously, there is a possibility to describe it as a set of actions by selecting a single operation; for example: operation $A \rightarrow B$: $NaK(a, b)$ (from ASF Handshake protocol) consists of actions:
$A \leftrightarrow K(a, b)B$ adequate description: $A, B, K(a, b), *, *, \mathit{Shared\ key}, *$,
$\rightarrow KA$ adequate description: $\{A, *, K(a, b), *, *, \mathit{has\ key}, *\}$,

$\#(Na)$ adequate description: $\{A, *, *, *, Na, nonce\ is\ fresh, *\}$,
where $\{*\}$ - irrelevant parameter in the described action.

In addition, we can exploit a system of coding to identify and recognize actions. This convenient approach consists of the coding weight system, e.g. binary, decimal, etc. The binary system is more extensive but it helps in describing particular actions more precisely. Generally, a simple coding system is proposed:

$$cta = \sum_{i=1}^{le} w_i * pos_el(i) \ - \ the\ code\ of\ an\ action\ type, \tag{1}$$

where $w_i = 2^{i-1} \ - \ (or\ 10^{i-1})$ - position weight.

The mutual cooperation among recognition and attribute correction procedures is illustrated in Fig. 1. Chosen sets of attributes create security modules

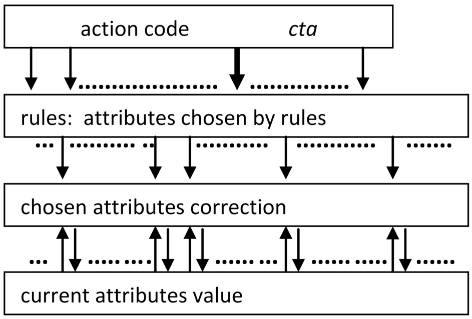


Fig. 1. Information flow in a result of action (cta) activity. Corrected attributes will become current attributes for the next action.

that concentrate around main factors like: communication protocols, keys, message service, and users.

3 Communication Security State and Node Structure

The security state is represented by an adequate security node structure.

Definition 2. A tuple (At, Th, Tk, na), where At - security attribute set, Th - the vector of a low level of feasible attribute values (thresholds), Tk - security tokens, na - the number of attributes, is a communication security state described as follows:

1. $At = \{at_1, at_2, ..., at_{na}\} \in [0,1]^n$ - the vector of attribute activation probabilities,
2. $Th = \{th_1, th_2, ..., th_{na}\} \in [0,1]^n$ - the vector of threshold attribute activation,

3. $Tk = \{tk_1, tk_2, ..., tkna\} \in \{0,1\}n$ - the binary vector (token) of attribute activation: if $at_i \geq th_i$ then $tk_i = 1$ else $tk_i = 0$.

Attributes may express assertion about user honesty belief, belief about message freshness, assertion about attestation, assertion about the shared key, belief that the receiver has jurisdiction over the message, etc. The attributes are corrected with the help of modification functions in cases when the current action appointed the attribute by the rule [1]. The sequence of actions influences a given set of attributes. The attribute set defines communication security level. The security value is estimated on three levels:

- global,
- in reference to the main security factor (security module),
- in reference to particular attributes.

According to the proposed development approach, we propose several structures of the security main factor (security modules), (Fig. 2,3).

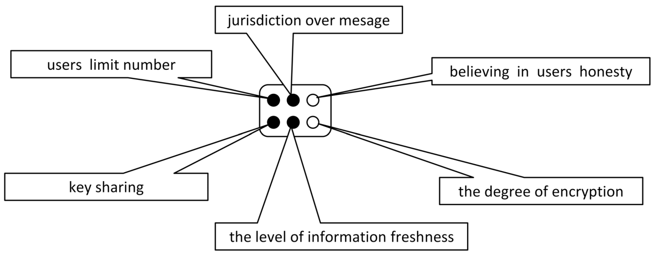


Fig. 2. The structure of the protocol security module, black - attribute activation, white - attribute has loosed activity

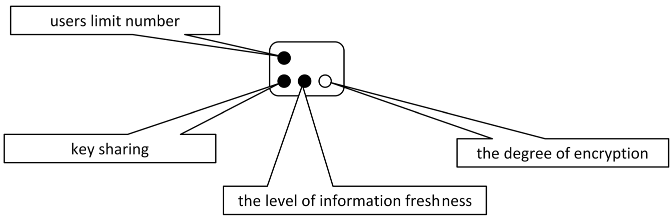


Fig. 3. The structure of the key security module

The action influences on attributes are conveniently presented and realized with the usage of equivalent tables that regard the above-mentioned rules. Therefore, the first table refers to action identifications and their characteristics, and

the second to attributes which will be corrected. In order to describe such situation, we consider two handshake operations as in Fig. 4,5 (column descriptions are adequate to action and attribute definitions):
Example:

1. $A \rightarrow B : \{N_a\}_{K(a,b)}$
2. $B \rightarrow C : \{N_b\}_{K(b,c)}$

These operations, belonging to two protocols, are decomposed into actions:

1. $A \leftrightarrow^{K(a,b)} B$ adequate description: $\{A, B, K(a,b), *, *, Shared\ key, *\}$,
2. $\rightarrow^{K} A$ adequate description: $A, *, K(a,b), *, *, has\ key, *$,
3. $\#(N_a)$ adequate description: $\{A, *, *, *, Na, nonce\ is\ fresh, *\}$,
4. $B \leftrightarrow^{K(b,c)} C$ adequate description: $\{B, C, K(b,c), *, *, Shared\ key, *\}$,
5. $\rightarrow^{K'} B$ adequate description: $\{B, *, K(b,c), *, *, has\ key, *\}$,
6. $\#(N_b)$ adequate description: $\{B, *, *, *, Nb, nonce\ is\ fresh, *\}$.

S1	*R1*	*S2*	*R2*	*M1*	*M2*	*N1*	*N2*	*K1*	*K2*	*Ch1*	*Ch2*	*Ch3*	*Ad1*	*Ad2*	*ac*
1	1	0	0	0	0	0	0	1	0	1	0	0	0	0	1
1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	2
1	0	0	0	0	0	1	0	1	0	0	0	1	0	0	3
0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	4
0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	5
0	0	1	0	0	0	0	1	0	1	0	0	1	0	0	6

Fig. 4. Action descriptions in the binary convention - example, where column descriptions respect the action structure (definition 1), *ac* - action numbers

JM1	*JM2*	*Bh1*	*Bh2*	*Nf1*	*Nf2*	*De1*	*De2*	*Kr1*	*Kr2*	*Ul1*	*Ul2*	*ac*
1	0	1	0	0	0	0	0	1	0	0	0	1
1	0	0	0	0	0	0	0	1	0	0	0	2
1	0	1	0	1	0	0	0	1	0	0	0	3
0	1	0	1	0	0	0	0	0	1	0	0	4
0	1	0	0	0	0	0	0	0	1	0	0	5
0	1	0	1	0	1	0	0	0	1	0	0	6

Fig. 5. The appointment of attribute corrections according to BAN rules, where JM_i - jurisdiction over i-th message, Bh_i -believing in i-th user honesty, NF_i - the freshness of i-th nonce, De_i - the degree (over one) of i-th message encryption, Ks_i - i-th key sharing, Ul_i - the exceeding limit number of users seeing i-th message in an encrypted form

The connection between tables (through action numbers) allows us to realize appointed security attribute corrections. The set of attributes is chosen on the basis of rules as well as time and heuristic functions [9]. In order to correct attributes we will use correction coefficients $CC(at(i))$ that were, previously predetermined for each attribute (Fig. 6). We present initial values of attributes by continuing the description of the example. Let us assume that the initial values of all attributes (obviously despite De) will be equal to 1(as a maximum value of trust probability). After 6 above-described actions (adequate 2 run operations), we may observe the following levels of attributes (Fig. 7). Attributes Nf and Ks are treated as timed attributes. Hence, the following formula is used for their correction: $at(k)(i) = 1 - ek - lf(at(i))$, where k - operation number [10].

correction coefficient of											
JM1	*JM2*	*Bh1*	*Bh2*	*Nf1*	*Nf2*	*De1*	*De2*	*Ks1*	*Ks2*	*Ul1*	*Ul2*
0,95	0,95	0,8	0,8	3	3	0	0	4	4	0,75	0,75

Fig. 6. Correction coefficient - example values. Attributes Nf and Ks are timed attributes, therefore their lifetimes $lf(N)$ and $lf(K)$ are given. Let us pay attention to De coefficient which is used for the *den*-th times blocking the correction of the adequate Ks attribute (obviously, only in the case when $De > 0$), where *den* - the degree of encryption.

attribute values after 2 communication eperations											
JM1	*JM2*	*Bh1*	*Bh2*	*Nf1*	*Nf2*	*De1*	*De2*	*Ks1*	*Ks2*	*Ul1*	*Ul2*
0,857	0,857	0,64	0,64	0,632	0,865	0	0	0,865	0,95	1	1

Fig. 7. The states of security attributes after two example operations

By considering the structure of tokens (binary structure) and the established threshold for all attributes on the level equal to $0,7$, it is possible to depict a security situation regarding different security modules (main factors) (Fig. 8). By treating all attributes with the same validity, we can estimate the level of security for all modules. This problem can be realized by the multiplication of specified component attribute probability values:
$SL(protocolA, B) = 0,875 * 0,64 * 0,632 * 1 * 0,865 * 1 = 0,306$,
$SL(protocolB, C) = 0,875 * 0,64 * 0,865 * 1 * 0,95 * 1 = 0,46$,
$SL(messageA, B) = 0,875 * 0,64 * 0,632 = 0,354$,
$SL(messageB, C) = 0,875 * 0,64 * 0,862 = 0,483$,
$SL(keyA, B) = 0,632 * 1 * 0,865 * 1 = 0,547$,
$SL(keyB, C) = 0,865 * 1 * 0,95 * 1 = 0,821$,
$SL(usersA, B) = 0,64 * 0,632 * 0,865 * 1 = 0,350$,
$SL(usersB, C) = 0,64 * 0,865 * 0,955 * 1 = 0,404$.
Similarly, token variant can be calculated on the basis of the estimation percent of active token participation in the full token set for a given security module (the main security factor). We may graphically present the security spectrum on the basis of the results estimated above.

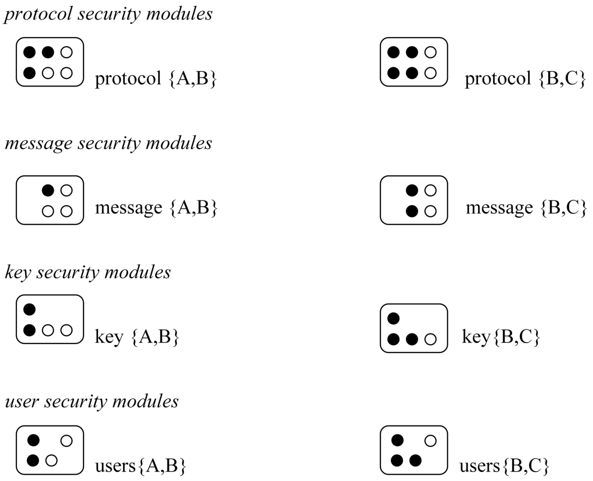


Fig. 8. The states of security modules, where A,B,C users

4 Probabilistic - Time Automata as Communication Security Investigation Model

We propose to use probabilistic - time automata (PTA) and converted to them colored Petri nets as a main tool for the investigation of communication security according to selected main factors, like: protocols, users, keys, messages, etc. The nodes (presented in Fig. 3,4 as examples) will be fundamental part of PTA. The global structure of proposed PTA is presented in Fig.9. If any attribute is decreased to an unacceptable level then there is not possibility to improve its value and security features cannot be increased. To regard the time parameter with intrinsic characteristic according to security aspect, we propose the following definition:

Definition 3. A probabilistic timed automaton PTA is a tuple in the form $(L, l', X, \sum, inv, p)$ where:

- L is a finite set of locations,
- $l' \in L$ is the initial location,
- X is a finite set of clocks (for each attribute),
- $\sum$ is a finite set of possible steps, of which $\sum_c \in \sum$ are declared as being current possible,
- the function $inv : L \to CC(X)$ is the invariant condition,
- the finite set $p \subseteq L \times CC(X) \times \sum \times Dist(2^X \times L)$ is the probabilistic edge relation.

A time state of a probabilistic timed automaton is a pair (l, v) where $l \in L$ and $v \in T^X$ are such that $v \in inv(l)$. Informally, the behavior of a probabilistic timed automaton can be understood as follows. The model starts in the initial

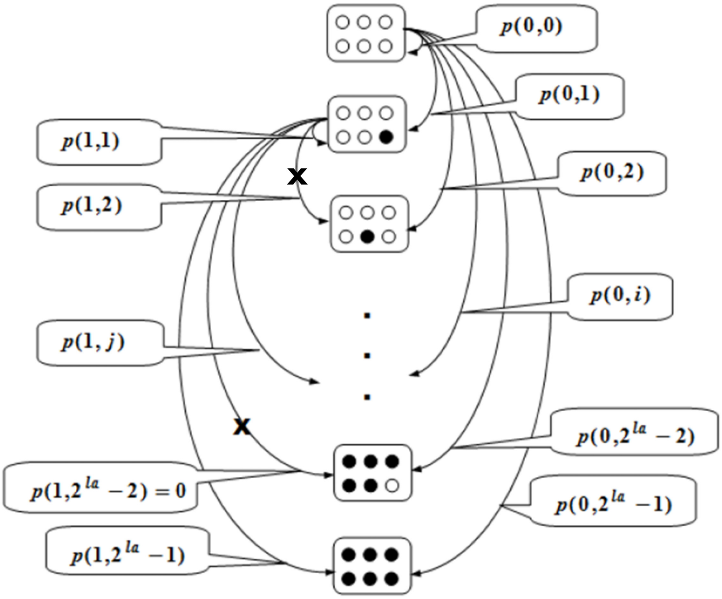


Fig. 9. Scheme of probabilistic - time automaton for communication security investigation, where $p(i, j)$ - the probability of state changing: from state i to j; $j \geq i$

location l' with all clocks set to 0, that is, in the state $(l', 0)$. In this, and any other state (l, v), there is a nondeterministic choice of either (1) making a discrete transition or (2) letting time pass. In case (1), a discrete transition can be made according to any probabilistic edge $(l, g, \sigma, p^*) \in p$ with source location l which is enabled; that is, the zone g is satisfied by the current clock valuation v. Then the probability of moving to the location l" and resetting all of the clocks in X to 0 is given by $p^*(X, l")$. In case (2), the option of letting time pass is available only if the invariant condition $inv(l)$ is satisfied while time elapses and an enabled probabilistic edge with a current step does not exist. Note that a timed automaton [2] is a probabilistic timed automaton for which every probabilistic edge $(l, g, \sum, p^*)$ is such that $p^* = \mu(X, l")$ (the point distribution assigning probability 1 to $(X, l"))$ for some $(X, l")\in 2^X \times L$.

Generally, the above formalisms are adequate to a single secure attribute (or action) connected with a single message. In order to adapt them to real communication runs, indexes are to be introduced to distinguish the threat zones. It is assumed that each message can be described by a set of security attributes. These attributes are involved with assertion, believing the sending, receiving messages, encryption, decryption by keys, nonce generation, attaching secrets, etc. Additionally, we should regard the number of users and their character (honest, intruder) [1],[2]. These considerations are based on time influences on chosen security attributes, strictly on their level (value).

The general system structure includes modules of security element definitions, input protocol operation descriptions, input table of dependence rules, lifetime checkers, and user set correctors. Each protocol operation includes strategic elements which will be chosen according to communication security, and hence, they will be named security elements. It should be said that security elements have independent or partly involved character. So, in relation to communication security, the mechanism gives us result information which consists of different aspects of security. One of the assumptions, which helps us to create and exploit probabilistic timed automata (PTA), is a finite set of the security state. The security element values are difficult to estimate. In this variant, according to known applications, we use their probability evaluation (security elements - tokens). Then, we introduce activation bounds (low and up) for each token. If the probability of the security attribute P_i (value of elements) is between $Plow_i$ and Pup_i then the adequate token is activated $tk_i = 1$, otherwise the token loses its activation $tk_i = 0$ (Fig. 10). This strategy allows us to define the finite set of states; their number will be equal 2^{le}, where le - the number of tokens (security elements). The best security situation is described by all activated tokens

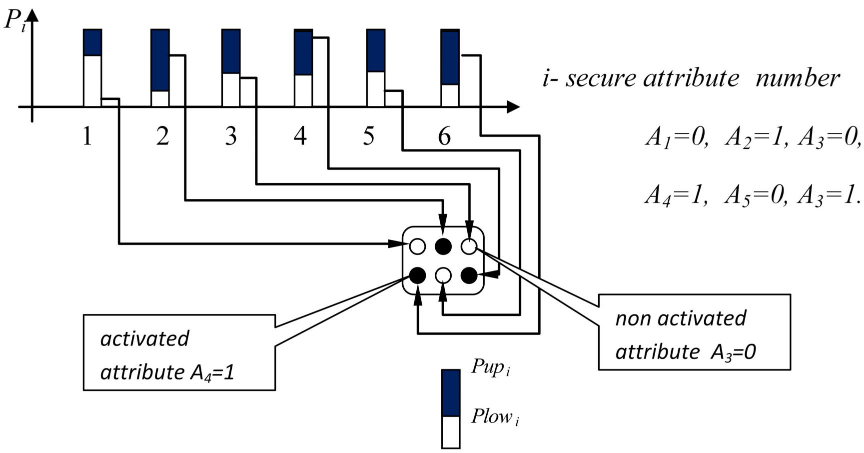


Fig. 10. Token binarization

$\forall_{i=1,2,\ldots,le}\ tk_i = 1$. The states can be changed during the realization of a protocol operation, but it is not necessary. The reasons for the non changing state are as follows:

- stable security attribute activation probability level $\forall_{i=1,2,\ldots,le}\ P_i^{(j+1)} = P_i^{(j)}$, where j - the number of the protocol operation,
- stable token activation $\forall_{i=1,2,\ldots,le}\{tk_i^{(j+1)} = tk_i^{(j)}\ |\exists_{k\in\{1,2,\ldots,le\}} P_k^{(j+1)} \neq P_k^{(j)}\}$.

The protocol operation description is a complex process because the following information should be provided:

- the scale of message encryption,
- the source and target of message transmition in pre-assumption,
- sharing keys (secrets) of users,
- belief (or assertion) parameters about security elements (attributes),
- lifetime parameter referring to a message, key, secret, nonce,
- the set of active users (honest and intruders).

It is assumed that in one operation only one message and one nonce will be sent, and that a set of keys and secrets will be exploited to respect standard rules (convention). However, also in this case, we have to use the multi-dimension table (due to the set character of the identical parameters). This situation can be illustrated by the tree structure depicted in Fig. 11. One part of operation

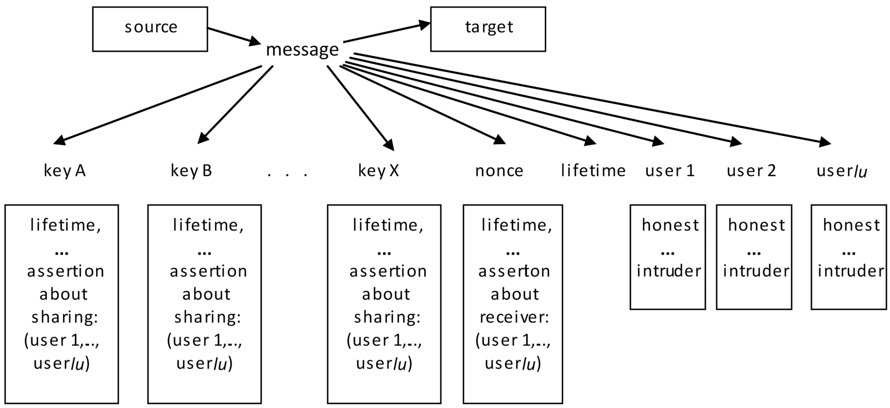


Fig. 11. The tree illustration of protocol operation description

parameter description has strictly deterministic character (real or integer) and the other part has probabilistic values. Time parameters are included in the first and belief of assert estimators in the second group. The assertion about honest or dishonest user behavior is obviously included in the assertion characteristic. If user honesty is certain then initial probability of this attribute is equal to 1: $P_i^{(0)} = 1$. The lifetime parameter is obviously constant but the time of adequate element activation increases according to real process realization (protocol operations). It is necessary to define and input the presented set of data for every message (message rules). They are the basis for using the chosen set of rules that change the level of attributes (security elements and next tokens). In order to present an identical example of message rules, we should define the set of communication logic predicates.

5 Conclusions

In the proposed approach it is possible to realize communication operation auditing and dynamically estimate the full spectrum of security aspects. The investigation is based on correction security attributes regarding rules, lifetimes and heuristic. Generally, the proposed algorithm is simple, but the preparation of dealing with the subject, which consists in the creation of security module structures and correction coefficient evaluation on the basis of experiences, can be more absorbing for communication security analytics. These structures and parameters should respect concrete situations and regard network information transfer and possible communication threats connected with different protocol realizations.

References

1. Burrows, M., Abadi, M., Needham, R.: A Logic of Authentication. Robert Harper., Logics and Languages for Security, pp. 15–819 (2007)
2. Focardi, R., Gorrieri, R.: A Classification of Security Properties. Journal of Computer Security 3, 5–33 (1995)
3. Beauquier, D.: On Probabilistic Timed Automata. Theoretical Computer Science 292, 65–84 (2003)
4. Gray III, J.W.: Toward a Mathematical Foundation for Information Flow Security. Journal of Computer Security 1, 255–294 (1992)
5. Evans, N., Schneider, S.: Analysing Time Dependent Security Properties in CSP Using PVS. In: Cuppens, F., Deswarte, Y., Gollmann, D., Waidner, M. (eds.) ESORICS 2000. LNCS, vol. 1895, pp. 222–237. Springer, Heidelberg (2000)
6. Kwiatkowska, M., Norman, G., Segala, R., Sproston, J.: Automatic Verification of Real-time Systems with Discrete Probability Distribution. Theoretical Computer Science 282, 101–150 (2002)
7. Kwiatkowska, M., Norman, R., Sproston, J.: Symbolic Model Checking of Probabilistic Timed Automata Using Backwards Reachability. Tech. rep. CSR-03-10, University of Birmingham (2003)
8. Alur, R., Dill, D.L.: A Theory of Timed Automata. Theoretical Computer Science 126, 183–235 (1994)
9. Szpyrka, M.: Fast and flexible modeling of real-time systems with RTCP- nets. Computer Science, 81–94 (2004)
10. Tadeusiewicz, R.: Introduction to Inteligent Systems. In: Wilamowski, B.M., Irvin, J.D. (eds.) Chapter No 1 in Book. The Industrial Electronic Handbook. CRC Press, Boca Raton (2011)
11. Tadeusiewicz, R.: Place and role of Intelligence Systems in Computer Science. Computer Methodsin Material Science 10(4), 193–206 (2010)
12. Alur, R., Courcoubetis, C., Dill, D.L.: Verifying Automata Specifications of Probabilistic Real- Time Systems. In: Huizing, C., de Bakker, J.W., Rozenberg, G., de Roever, W.-P. (eds.) REX 1991. LNCS, vol. 600, pp. 28–44. Springer, Heidelberg (1992)

Parent Plan Support System – Context, Functions and Knowledge Base

Michał Araszkiewicz[1], Agata Łopatkiewicz[2], and Adam Zienkiewicz[3]

[1] Department of Legal Theory, Faculty of Law and Administration, Jagiellonian University, Kraków, Poland
michal.araszkiewicz@uj.edu.pl
[2] Institute of Education, Jagiellonian University, Kraków, Poland
agata.lopatkiewicz@uj.edu.pl
[3] Department of Theory and Philosophy of Law and State, University of Warmia and Mazury, Olsztyn, Poland
adam.zienkiewicz@uwm.edu.pl

Abstract. This paper presents a knowledge base (and foundational issues of its operation) for a support system designed for divorcing parents and intending to decide on issues of exercise of parental authority and contacts with children. The legal and functional context of this system is also presented. However, the main focus is on ideas concerning representation of a very complex legal concept, i.e. well-being of the child.

Keywords: Child, dimensions, divorce, factors, negotiation, parent plan, well-being.

1 Introduction

Divorce cases in which minor children are present belong to the class of the most difficult and delicate issues under any particular legal system. The main claim in these cases (whether divorce should be granted or not) is connected with problems in the division of spouses' property and, when minor children are present, with decisions concerning contact with the children and child custody after the divorce is granted. Parties in divorce disputes are often seriously conflicted with each other, and emotions often impair their ability to cooperate on important issues concerning relations with their children after the divorce. What is particularly important here, under Polish law, which will serve as an illustrative material for the following discussion, it is possible for the parents to draft the so-called "Parent Plan" in which they can decide on issues concerning exercise of child custody and contacts with the child will be settled. Our contribution aims at supporting them in drafting such plan, therefore leading to amicable resolution of potential (and often actual) conflicts on this matter.

The order of the investigation is as follows: in the following Section, the legal framework for dealing with parental authority and contact with children is presented in the context of the laws applicable to Polish divorce procedure. In Section 3 we present a list of the assumed functions of the system and a structure of an exemplary

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 160–171, 2013.

Parent Plan. In Section 4 we present the structure of the knowledge base of the Parent Plan Support System (PPSS). The functioning of the PPSS is briefly presented in Section 5; we focus on interrelations between different layers of knowledge within this system. Section 6 is devoted to a brief discussion of relevant literature in AI and Law research, focusing on the research on CBR (Case Based Reasoning) with emphasis on factor-based and dimension-based systems. In the last section, we offer partial conclusions and present perspectives for further development of ideas discussed in this paper.

2 Legal Background

In this section, we outline the legal and conceptual context for the institution of Parent Plan. Our analysis is based on Polish law. Obviously, legal systems of UE countries create the most proximate ‘neighborhood’ for this analysis.

A decision on parental authority over a minor child is a part of a divorce judgment under Polish civil procedure. The judge should also decide on issues of parents’ contact with children. The court takes into account an agreement (Parent Plan) between the parties as regards the mentioned issues, however, such Parent Plan may be accepted only insofar it is in accordance with “well-being of the child”. The court may grant parental authority to one of the parents, thereby limiting parental authority of the other parent to specified duties and rights as regards the child. The court may grant parental authority to both parents provided that they have submitted the Parent Plan, and it is reasonable to assume that they will cooperate on matters regarding the child. The issue of parental authority is always decided on the basis of the circumstances of a concrete case.

At this point, it is necessary to briefly summarize the fundamental legal categories relevant to the Parent Plan.

According to Polish law, parental authority should be defined as the duty and right of the parents to have custody over the person and property of the child, with consideration of the child’s dignity and rights. Parental authority should be exercised in accordance with the child’s well-being and with public well-being. The parents should take care of the physical and spiritual development of the child and prepare him or her to work for the society in accordance with his or her abilities. The child is under parental authority until he or she reaches legal maturity. Parental authority, in principle, is vested to both parents.

Independently of parental authority, the parents and the child have both the right and duty to contact with each other. Contact with the child encompass, *inter alia*, spending time with the child, direct communication, and distance communication, for instance by means of electronic devices.

The child’s well-being should be considered as the central and supreme category as regards the exercise of parental authority and contact between the parties.

The other very important category is public interest, because the child’s upbringing is a realization of the parents’ social function. The parents, taking care of the child’s physical and spiritual development, prepare him or her to work for the benefit of the society, according to his or her capacities.

Consequently, the court, while deciding on the issue of parental authority, should be guided first and foremost by the child's well-being. The interest of the parents is an important premise, but it cannot be decisive. The court should, if possible, take the position and interest of the parents into account. This approach is consistent with the Convention on the Rights of the Child.

Moreover, it is necessary to emphasize the principle of rational cooperation between the parents and the child as regards exercise of parental authority.

The legal context for the system presented in this paper is given by rules of law that do not prescribe any concrete answers to questions important to the context of the drafting of a Parent Plan. The legal domain presented above can certainly be classified as ill-defined [1]. The knowledge base presented in this paper aims to re-characterize the domain, giving the interested parties a tool that may be useful in solving the problems arising in connection with drafting the Parent Plan.

3 Functions and Structure of the Parent Plan

The discussion of legal domain in question leads us to clear formulation of the task which is to be performed by the parents: it is to draft such Parent Plan that could possibly pass the test of judicial assessment against the criterion of (realization of) the value of well-being of the child.

The Parent Plan should perform a set of more specific functions:

1. Facilitation of regulation of relations between the parents and the child after divorce;
2. Securing of the child's interest;
3. Elimination of unacceptable behavior of the parents;
4. Education of the parents as regards pedagogical, legal, psychological, or social rules concerning their relation with the child; and
5. Prevention of future judicial disputes concerning the child's well-being.

The tool we outline in this paper should perform, in particular, the following functions:

1. Provide the parties with a complex template of a Parent Plan.
2. Enable the users to enter rich context-sensitive data concerning their present situation.
3. Provide the parents with a rich variety of options, to make them aware of the broadness of the decision-making space and of its limits.
4. Gather information about the nature of the conflict (if any) between the parties.
5. Enable the negotiating parents to communicate, leaving an option to directly (traditionally) deliberate some choices on the one hand and to opt for automated generation of parts of the plan on the other hand.
6. Establish the space of agreement for the parties concerning the contextual data provided by the parents, in order to negotiate options that should be mutually acceptable, mutually advantageous, and realistic.
7. Provide the parents with suggestions concerning their choices of options and justifying these suggestions in an interdisciplinary knowledge base encompassing psychological, pedagogical, and, very importantly, legal data.

8. Generate a final electronic version of the Parent Plan to be submitted before the court; however, the system should be able to update the plan if circumstances change in the future.

The presentation in this paper is concerned with architecture of the PPSS. The system has not been implemented yet; however, the results presented below are not purely speculative, because they are based on actual legal sources. A crucial part of the system's knowledge base has been extracted from 30 actual judgments enacted by the Polish Supreme Court. Also, case analysis in connection with an existing sample form (unofficial) of the Parent Plan made it possible to systematize the content of the Parent Plan into ten parts, which are referred to as "Categories". Exemplary "Subcategories" are listed below the name of each Category. We present the structure of the plan's content in more detail in order to make it more easy to follow the discussion of the system's functioning as described further in the paper. Additionally, in each Parent Plan, data referred to as Essential Information should be entered. For the sake of simplicity, we assume here that the parents have only one child. The structure of the Parent Plan is as follows:

Essential Information 1 (formal) – this part encompass the most basic formal data enabling the court to identify the parties to the agreement and the child to whom the Parent Plan is dedicated: (1) The personal data of the child and parents, including the child's family name; (2) The addresses of the child and parents; (3) Health insurance information; (4) Method of conduct in case of the decease of one or both parents.

Essential Information 2 (emotional) – this part gathers important declarations of intention of the parties thereby signaling them that their aim is to cooperate for the sake of well-being of the child and not their particular interests, and encompasses declarations of: (1) Unconditional care for the child's security; (2) Unconditional lack of violence; (3) Love towards the child.

Essential Information 3 (realization of the plan) – this part is more technical and encompasses meta-rules concerning method of conduct as regards 'management' of the plan: (1) Methods and forms for solving conflicts concerning issues not stated explicitly in the plan; (2) Methods applicable in case of infringement of the plan; (3) Rules concerning modification of the plan.

As for the content of the specific parts of the plan, it is structured as follows (only exemplary subcategories are listed):

Category 1 – **Contact between the parents and the child**

- Planning meeting schedule between the parents and the child;
- Establishing forms of contact between the parents and the child.

Category 2 – **Contact between the child, family members, and other people**

- Establishing forms and range of contact between child, family members, and other people.

Category 3 – **Education**

- Selection of educational institutions;
- Establishing forms and range of contact between the parents and the representatives of educational institutions;
- Choice of extracurricular activities and career paths.

Category 4 – **Livelihood**
- Estimating permanent needs of the child related to livelihood issues;
- Estimating occasional needs of the child;
- Planning financial investments for the child.

Category 5 – **Leisure, holidays, and ceremonies**
- Leisure:
    - Planning forms and ways of spending the child's spare time;
    - Planning forms and ways of spending free time that are unacceptable.
- Holidays:
    - Planning places of departure;
    - Planning duration of holidays;
    - Establishing rules for going abroad;
    - Planning forms and ways of spending holidays that are unacceptable.
- Ceremonies:
    - Establishing a list of celebrated ceremonies;
    - Establishing rules for celebrating ceremonies.

Category 6 **– Worldview and raising rules**
- Confession issues:
    - Establishing issues related to the parent's and the child's participation in religious ceremonies.
- Issues of upbringing:
    - Establishing crucial rules of the child's upbringing;
    - Planning a typical day's schedule.

Category 7 – **Child health**
- Choosing a primary care physician;
- Planning related to periodic tests and vaccinations;
- Planning how the child will be cared for during illnesses.

Category 8 – **Informing and documentation**
- Establishing rules of contact between the parents in daily affairs;
- Establishing rules of contact between the parents in emergencies.

Category 9 – **Proceeding in emergency cases**
- Establishing custodian in case of the illness or death of the parent(s);
- Establishing custodian in case of emergency departure of the parent(s);
- Establishing custodian in case of committing an offence by the parent(s).

Category 10 – **Remaining issues**
- Planning related to child functioning that are not categorized in any of the categories described above.

4 The Knowledge Base Structure

In this section we outline the structure of knowledge base for the PPSS. It is worth emphasizing that it is a multi-layered structure based mainly on CBR elements (reasoning with factors and values); however, defeasible rules also play an important role in it.

An important point of departure for further analysis is that the court formulates a value judgment on any Parent Plan submitted to it. In order to handle the complexity concerning formulation of this value judgment [2, 3] the system's knowledge base contains different types of knowledge represented by different structures.

1. Options. Options are basic propositional elements that have a simple logical structure (no logical connectives involved if possible). Let O be the set of all Options in the system. Therefore, each concrete Parent Plan (PP) generated by the system is a subset of O. Options are first grouped into sets called Questions (Q). An answer to each Question in the system consists of choosing one Option from the set provided with a given Question. For each PP and a given Question_i ($Q_i \subset O$), there exists only one such Option that belongs to this Question_i. In consequence, in each Parent Plan, exactly one Option should be chosen for each Question. We state that a given Parent Plan is *complete* if and only if it contains an Option for each Question. We also define a relation of Incompatibility (Inc) over a set of Options, and we state that if two Options from one and the same Parent Plan, o_i and o_j, $\in$ Inc, then this Parent Plant is *inconsistent*. A set of Questions forms a Subcategory (S). A set of Subcategories forms a Category. The set O is equivalent to the sum of the Categories.

2. Dimensions. The system contains a library of ten basic Dimensions that are scalable representations of Categories of the Parent Plan. For each Category outlined above, a Dimension is created. Unlike in the classic dimensional approach to CBR, the range of a Dimension does not go from the extreme pro-defendant to the extreme pro-plaintiff point, but instead from the unacceptable point of view of the child's well-being to the optimal point as regards realization of this value. Conventionally, we adopt a 10-step ordinary scale (from 0 to 9) for each Dimension, ranging from the Unsatisfactory point, through Sufficient (1-3), Good (4-6), and Excellent (7-9) realization of the child's well-being. Ultimately, the scale is reducible to four steps, but we adopt a 10-step scale to account for nuances stemming from contextual information.

3. Defeasible Rules (DR). A crucial part of the system is a huge (but finite) set of defeasible rules extracted from pedagogy and developmental psychology research and from the decisions of the Polish Supreme Court, as well as common-sense-based rules. For each Option $\in$ O, there exists a default defeasible rule of the following form:

$r_i \in (DR)\text{: } o_i \Rightarrow [0 \leq n \leq 9]$

which means that rules accept Options as their antecedents and one of the values on the scale of a given Dimension of the system. Consequently, defeasible rules provide for a default dimension value for any Option chosen by the user of the system. It is important to note that the rules should be seen as default rules only, because the system offers multitudinous possibilities to modify their default outcomes. Due to the fact that certain Options may belong to Incompatibility relation, the standard relations of attack and default priorities [4] between the Defeasible Rules are also present in the system's knowledge base.

4. Environmental Factors (EF). This set encompasses factual statements that are given by the parents in the system's Questionnaire before they deliberate the Options. This set encompasses data such as distance from the residence of each parent to the schools in the surrounding area, the parents' monthly salaries, as well as relevant data concerning the parents' and the child's personal characteristics. Like Options, there are Environmental Factors that are mutually incompatible (as for an obvious example, one child cannot have different ages at the same time). There are also Environmental Factors that are incompatible with certain Options $\in$ O. In certain environments, some Options are not relevant. If a given Option o_i is incompatible with an Environmental Factor ef_i, then this option does not belong to the feasible set of Options.

Environmental Factors are related to Defeasible Rules by means of Influence Relation. We state that a factor $ef_i \in EF$ and a rule $r_i \in DR$ belong to the Influence Relation ($(ef_i, r_i) \in Inf$) if and only if the presence of this factor in the Parent Plan changes the function expressed by the rule. A given Option being an antecedent of this rule is qualified differently as regards the value of a given Dimension, in contrast to how it would be qualified if the factor was not present in the Parent Plan. Consequently, Environmental Factors change the consequents of Defeasible Rules (they can make certain Options more or less valuable).

5. Cases and Legal Factors. The system contains a knowledge base of judgments of the Polish Supreme Court in which the value of the child's well-being was discussed. This base of cases performs a twofold function in the system:

First, these cases are sources of Legal Factors that are used to characterize extreme points of the Dimensions. Consequently, an important part of Defeasible Rules stored in the DR set of the system which yields a 0-point on the scale of a given Dimension or which yields an Excellent valuation is grounded in the explicit wording of the Polish Supreme Court. Thus, construction of Dimensions of the system is rooted in actual authoritative judicial statements concerning legal understanding of this value. Hence, the presence of this kind of knowledge in the system enhances the likeliness of acceptance of a given Parent Plan by the divorce court.

Second, the set of Legal Factors, if applicable to a given Question, provides stronger justification for default valuation of certain Options by the Defeasible Rules. Like Environmental Factors, they are also related to Defeasible Rules by means of Influence relation, and they can change the default valuation provided by these rules.

5 Functioning of the Parent Plan Support System

This section is devoted to presenting the functioning of the PPSS from the point of view of the user. Purely technical steps are described briefly, while steps that are relevant from a theoretical point of view are discussed in more detail. This exposition is illustrated by examples. The step-by-step picture of the PPSS is as follows:

1. A Tutorial.

2. Terms and Conditions Acceptance Step.

3. Basic Questions Step. These questions aim at establishing the intentions of the parties to negotiate for the sake of the child's well-being.

4. Logging In. The system provides for both an option for a single parent to enter the system (for instance for the sake of testing its options) and for both parties to log in separately but work simultaneously on the Parent Plan.

5. Questionnaire. In this step, multifarious information concerning the parents and the child(ren) is gathered from the parties, comprising the data concerning age, education, financial status, and current family situation. As a result of this, a set of Environmental Factors is obtained.

Filling in the Questionnaire and accepting it yields two important computational consequences in the system:

First, the system maps the set of Environmental Factors to the set of all Options in the system and filters out the Options that are incompatible with the former set of factors. Consequently, the feasible sets of Options are generated for each parent.

Second, the system maps the set of Environmental Factors to the set of Defeasible Rules and retrieves such EFs that Influence the Defeasible Rules, changing the valuations prescribed by consequents of the Defeasible Rules elements accordingly.

6. The Option Choice Phase. In this step, both parents simultaneously go through the Categories of the Parent Plan and choose options in response to Questions. Each Question provides for a feasible set of Options, one of which should be chosen by the parent.

The choice of an Option by the parent yields three important computational consequences in the system.

First, the choice of the parent is stored in the set referred to as Parent_Options_[x], where x may be substituted by "mother" or "father" depending on which parent chose an Option.

Second, the choice of an Option activates the indexation of options available to the other parent such as Matching, Neutral, and Conflicting.

Third, as far as the Options chosen by the parent are antecedents of Defeasible Rules stored in the system's knowledge base, the Defeasible Inference Engine accepts the chosen Option as input and yields a valuation of this Option on the scale of relevant Dimension as output. Note that this valuation may not be the original default valuation, but a modified valuation resulting from the information provided by the parent in the Questionnaire.

Consequently, each chosen Option is valued, where values range from Unsatisfactory to Satisfactory, and Good to Excellent (as regards realization of the child's well-being in respect to the given Dimension). The exact assigned number (from 0 to 9) is not presented to the parent, although it may play a role when a more nuanced bargaining process between the parents begins.

If a parent is not able to find an acceptable Option in the list provided by the system, he or she may type it in natural language

7. The Option Deliberation Phase. After the parents have chosen their Options to a given Question, their choices are classified as Matching, Neutral, or Incompatible. Let us assume that parent (1) chose Option x to Question A and that parent (2) chose Option y to the same Question. Valuation of these chosen Options may be identical, or Option x may be valued higher as regards realization of the child's well-being, or vice versa.

If Options are classified as Matching, then they possess identical valuation (not necessarily the highest one on the scale). The system informs the parents about this result; however, if in the space of agreement there is a pair of Matching Options that are valued higher, the parents are informed about the suboptimal character of their choice and the superior Options are suggested by the system.

If Options are classified as Neutral, it means that they are in the space of agreement between the parents; however, they can possess different valuation and choosing these

Options would leave some important matters encoded in a given Question undecided by the parents. A suggestion is given to the parent whose Option is valued lower (hereafter referred to as Parent #2) to change his chosen Option. If the valuation of both Options is equal, the system gives suggestions to the parents in a random sequence.

If Options are classified as Conflicting, then they cannot be accepted simultaneously. The parents are informed that they are not within the space of agreement. If the valuation of the chosen Option is unequal, the parent who chose the lower valued Option (Parent #2) is asked to change it. An equal valuation of Options leads to a random sequence of suggestions.

The system offers the parents several possibilities for reaction to the information provided by the system as regards the Option Deliberation Phase.

7.1. If the chosen Options are Matching, the possibilities are as follows:

1) To ACCEPT the chosen Options

2) To IMPROVE the chosen Options as regards the valuation, according to the system's suggestion, or

7.2. If the chosen Options are Neutral, it is possible for the addressee of the system's suggestion (Parent #2):

1a) To IMPROVE the chosen Option according to the system's suggestion, or

2a) To REJECT the suggestion and to PROVIDE A REASON for this decision (the reason should be typed in by the parent, whereby his or her decision becomes visible to the other parent). The responding parent is given a possibility:

1b) To DENY the reason provided by the other party by stating that it is untrue; obviously, such differences concerning factual matters can create serious obstacles to the process of seeking agreement;

2b) To ACCEPT the reason provided by the other parent as a matter of fact, but at the same time to assess it as INSUFFICIENT in providing a good ground for rejection of the suggestion;

3c) To ACCEPT the reason and to acknowledge it as SUFFICIENT. In such a situation, the Option chosen by Parent #2 remains intact while the Option initially chosen by Parent #1 is automatically modified in order to establish a Matching relation between the Options. Parent #1, if his or her Option was initially valued higher, is awarded a Concession Point, which is visible to both parents, relaying that this Parent changed his or her original choice due to a suggestion from the other party and not to the system.

7.3. If the chosen Options are Conflicting, Parent #2 obtains a suggestion from the system to modify his or her Option. In consequence he or she may:

1a) MODIFY the chosen Option, in particular by withdrawing it, or

2a) REJECT the suggestion and PROVIDE A REASON for this rejection. The rest of the possibilities, also as regards the responding parent, are analogous to those discusses above as regards Neutral Options. An addition here is that if Parent #1 makes a concession, he or she is awarded a Strong Concession Point.

As it turns out, in some circumstances the parents may not be able to reach an agreement within the frame of the aforementioned possibilities. Such Questions are marked as NOT SETTLED. The parents may continue until an Option is chosen by each parent for each Question, even if some of the Questions are not settled.

8. Compatibility Testing Phase. At the conclusion of the previous case, each parent receives a complete Parent Plan. In this phase, the internal compatibility of the chosen Options is tested. In the case of detection of Incompatibility between the chosen Options, the parent is asked to modify the Options belonging to this relation.

9. The Final Bargaining Phase. At the beginning of this phase, both parents have complete and consistent Parent Plans at their disposal; however, many questions may still be labeled as NOT SETTLED due to differing opinions of the parents. In order to overcome these difficulties, the system offers the parents the following four possibilities:

9.1. Returning to phase 7 and further deliberating the controversial Options with supportive suggestions from the system.

9.2. Choosing an option for automated bargaining. The output of the maximizing function will be constrained by the data provided by the parents in the Questionnaire, as well as by the choices they accept. Consequently, the computation of maximizing results will concern only those Questions that are labeled as NOT SETTLED. In the case of a huge number of different well-being maximizing plans being retrieved, the system takes into account Concessions and Strong Concessions made by the parents in the Option Deliberation phase, according to the following simple algorithm:

(1) Compute all Concessions and Strong Concessions made by the parents in the Option Deliberation Phase, assigning two points for each Strong Concession and one point for each Concession.

(2) Decide the NOT SETTLED Questions in favor of the parent possessing a larger amount of points by paying two points for each Question, provided that the choice of the Option favoring this parent is not assessed lower on the scale of Dimension relevant to the Question than the Option chosen by the other parent, and until the amount of points of the parents is unequal. Then, continue applying the algorithm alternating between a Question in favor of Parent #1 and a Question in favor of Parent #2.

Note that there is no significant threat that application of this algorithm could demote the realization of the child's well-being because this algorithm is applicable only if the system retrieves more than one well-being maximizing Parent Plan.

9.3. Deciding to settle the controversial issues by means of informal discussion in the chat room.

10. Acceptance phase. In this phase, the parents may ultimately accept the generated Parent Plan or choose to return to one of the earlier phases.

6 Related Work and Discussion

The creation of a Parent Plan can be characterized as multi-criteria optimization process. As for the general theory of multi-criteria optimization, the present contribution presents an idea concerning the transformation of an ongoing problem with a potentially infinite number of alternatives into a discrete problem with a huge yet finite number of alternatives [5]. The system uses a simple ordinal scale concerning realization of the value in question, ranging from 0 to 9. This kind of scale is often applied in literature on AI and Law, for instance by Grabmair and Ashley [2, 3].

The system described here is an instance of teleological reasoning, performed in a specific, non-adversarial context. It encompasses the general idea of Pareto-optimality [6, p. 184] because the system signalizes to the parents the choice of better Options, if they are available. In this paper, the issue of marginal analysis and the weighing of values [6] is not discussed, however it may play a role when the parents' preferences are discrepant as regards different aspects of the value of the child's well-being.

The dimensions employed in PPSS are scalable knowledge representation structures, although there are some differences between our account and the classical account of dimensions in CBR [7]. The first difference is that the extreme points of the dimensions do not support the opposite outcomes of the case (this difference is not vital from mathematical point of view, although it influences argumentation moves). The second difference concerns the structure of the dimension. Each point on the scale of the dimension may be satisfied by different Options (or sets of Options). Third, each Parent Plan is indexed by all of the Dimensions (in HYPO for instance, a case could be characterized by one or more dimensions).

CATO-style factors [8] also play their role in the PPSS, and they serve as antecedents of Defeasible Rules as well as factors that modify the default valuation of Defeasible Rules. These computational procedures can be described in terms of attack, defeat, and priority relations, which is a standard account in AI and Law literature [4].

The framework presented here can be seen as a domain-concretized instance of comparative case-based reasoning. For instance, PPSS' default ordering of Options satisfies the weak (defeasible) transitivity condition discussed by Hage [9, p. 107]. There are also important connections between the PPSS framework and the model of a two-phase policy deliberation as proposed by Bench-Capon and Prakken [10] and developed in further work [11] regarding its phases. In PPSS, both the Criteria Deliberation Phase and the Proposal Deliberation Phase are present, although the former is somewhat limited due to the fact that lists of Options and Default Valuations are presented to the system's user, and not created by him or her; however, the user may modify these default data in different manners.

Due to the fact that negotiations between the parents take place in a special context (determined by legal premise to a far greater extent than by the interest of the negotiating parties), the CBR context seems to be more important for a discussion of PPSS than classical research on negotiation support systems. However, there are also important and obvious connections with this line of AI and Law research. It must be emphasized that PPSS represents the structure of a dialogue system (in the Option Deliberation Phase) and tradeoff and compensation module of a negotiation support system (The Final Bargaining Phase) [12]. Although we use symbolic representation of knowledge only in this conceptualization, the possibility of training an artificial network on a huge basis of cases to provide for a higher degree of automatization of the system's reasoning is not excluded (this technique has been used in a well-known Split-Up system designed by Zeleznikow, Stranieri and Gawler [13]). This option seems promising, as there are more than three thousand published judicial decisions that mention the expression "well-being of the child" in different contexts. However, this is a topic for further investigations.

7 Conclusions and Further Research

The paper presents an outline of a multi-layered knowledge base (and its logical operation) for creation and assessment (in the sense of compliance) of an agreement between the parties. The knowledge base is grounded mainly in CBR knowledge representation structures.

The most important prospect for future research is to implement the system and to test it empirically (where one of the most important issues is to test the plausibility of entered sets of Options).

The research project entitled "Mediation and other alternative dispute resolution methods in Polish legal system - theory and practice" has been financed by the National Center for Science, agreement no. 1307/B/H03/2011/40. This paper has been prepared in the frame of this research project.

References

1. Lynch, C., Ashley, K.D., Pinkwart, N., Aleven, V.: Concepts, structures and goals: Redefining ill-definedness. International Journal of Artificial Intelligence in Education 19, 253–266 (2009)
2. Grabmair, M., Ashley, K.D.: Argumentation with value judgments – An example of hypothetical reasoning. In: Winkels, R. (ed.) Legal Knowledge and Information Systems JURIX 2010: The 23rd Annual Conference, pp. 67–76. IOS Press, Amsterdam (2010)
3. Grabmair, M., Ashley, K.D.: Facilitating case comparison using value judgments and intermediate legal concepts. In: Ashley, K., van Engers, T. (eds.) Proceedings of ICAIL 2011, pp. 161–170. ACM Press, New York (2011)
4. Prakken, H., Sartor, G.: Modeling reasoning with precedents in a formal dialogue game. Artificial Intelligence and Law 6, 231–287 (1998)
5. Ehrgott, M.: Multicriteria optimization. Springer, Heidelberg (2005)
6. Sartor, G.: Doing justice to rights and values: Teleological reasoning and proportionality. Artificial Intelligence and Law 18, 175–215 (2010)
7. Ashley, K.D.: Modeling legal argument: Reasoning with cases and hypotheticals. MIT Press, Cambridge (1990)
8. Aleven, V.: Teaching case-based argumentation through a model and examples (Unpublished doctoral dissertation). University of Pittsburgh Graduate Program in Intelligent Systems (1997)
9. Hage, J.C.: Studies in legal logic. Springer, Berlin (2005)
10. Bench-Capon, T.J.M., Prakken, H.: A lightweight formal model of two-phase democratic deliberation. In: Winkels, R. (ed.) Legal Knowledge and Information Systems JURIX 2010: The 23rd Annual Conference, pp. 27–36. IOS Press, Amsterdam (2010)
11. Bench-Capon, T.J.M., Prakken, H., Visser, W.: Argument schemes for two-phase democratic deliberation. In: Ashley, K., van Engers, T. (eds.) Proceedings of ICAIL 2011, pp. 21–30. ACM Press, New York (2011)
12. Lodder, A., Zeleznikow, J.: Developing an online dispute resolution environment: Dialogue tools and negotiation support systems in a three-step model. Harvard Negotiation Law Review 10, 287–337 (2005)
13. Zeleznikow, J., Stranieri, A., Gawler, M.: Project report: Split-Up: A legal expert system which determines property division after divorce. Artificial Intelligence and Law 3, 267–275 (1996)

Linked Open Data for Legislative Domain – Ontology and Experimental Data*

Martin Nečaský, Tomáš Knap, Jakub Klímek, Irena Holubová, and Barbora Vidová-Hladká

Charles University in Prague
Faculty of Mathematics and Physics
necasky@ksi.mff.cuni.cz
http://xrg.cz/

Abstract. In this paper, we present an approach of publishing legislative content as Linked Open Data (LOD). LOD is a set of principles of publishing data on the Web in a machine-readable way so that links between different data sets, possibly published by different publishers, can be created. Therefore, LOD enable not only to publish data but also enrich with other existing data published according to the principles. We present what is the motivation for publishing legislation as LOD and what benefits can be gained. We then introduce a legislative ontology which builds on existing commonly used ontologies. We also show how we converted existing sources of legislation in Czech Republic to LOD.

Keywords: Linked Data, legislation, ontologies, Semantic Web.

1 Introduction

Linked Open Data (LOD) [3] is a set of principles of publishing data on the Web in a machine readable way which enables linking data of various publishers at different places on the Web. Linking means that publishers enrich their data with other data published on the Web. A data consumer can then work not only with the data of a particular publisher but can also work with linked enriching data of other publishers. The links can be created by one of the data publishers of the linked data sets or by any third party publisher.

In this paper, we show how legislation can be published as Linked Open Data. We work with legislation of the Czech Republic, namely acts issued by the National Parliament of Czech Republic, decrees and regulations issued by the National Government of the Czech Republic, and court decisions published by the Supreme and Constitutional Court of Czech Republic. These documents are generally called *sources of law*. They are published by the authorities at various places on the Web in a distributed way in different formats (usually HTML or PDF). At some places only metadata about legislative documents are published. At other places documents themselves are published. Those different places are

* This work was supported by the TAČR grant no. TA02010182.

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 172–183, 2013.

not linked in any way. Therefore, it is hard for citizens to find legislative documents they need and to search for related content. We propose how legislative documents in Czech Republic can be published as LOD and demonstrate advantages of such style of publication. We propose an ontology for legislative documents which is based on existing ontologies already exploited for publishing legislative documents. We enrich them with new specific parts. Then we show how current sources of legislative documents can be transformed to their LOD representation and how the created LOD representation can be exploited by application developers.

There are currently several projects which aim at publishing legislative documents as Linked Data. The project CEN MetaLex [4] provides a standard of how sources of law and references to sources law are to be represented in XML. It also provides a LOD variant of their publication standard. An ontology for publishing legislation as LOD has been provided by the project. MetaLex publication server[1] publishes Dutch legislation according to the standard. A similar project is Legislation.Gov.UK[2] which also publishes legislation as LOD according to MetaLex format. In our work we reuse the rules given by CEN MetaLex and apply them to publishing legislation in Czech Republic. We extend the publication format so that other kinds of relationships not only between acts (as supported by MetaLex) but also other sources of law (e.g., court decisions) can be recorded.

2 Motivation

Because of the complexity of various sources of law, it is very hard for users, i.e. legal professionals as well as common citizens, to search for the required documents. In this section, we demonstrate most common use cases of how professionals as well as citizens work with sources of law.

Sources of law are usually structured to sections which may contain further subsections. Moreover, a source of law may contain references to other sources. A reference may target not only a whole source but also its particular section. Therefore, the structure encoded in sources of law and references between them form a complex network. Moreover, related sources of law are often published by different public authorities. It would be useful to enable to browse this distributed network among different data sources and search for relationships between sources of law and/or their parts. Examples of common use cases are listed in the following list. For the demonstration, we use Figure 1. It shows a part of the network comprising several related acts and court decisions.

1. A user is reading a particular section of an act (e.g., Act Section § `10` of Act `§99/1963`). He would like to see what court decisions have been made in the last decade related to this particular section (i.e., decisions `20 Cdo 1691/2005` and `20 Cdo 1691/2005`).

[1] `http://doc.metalex.eu/`
[2] `http://www.legislation.gov.uk/`

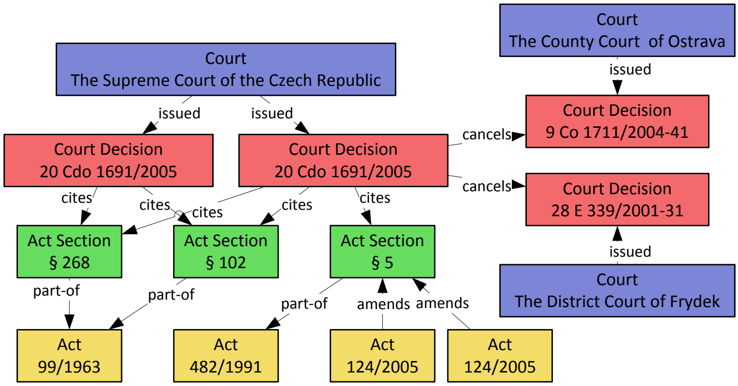


Fig. 1. Sample Links between Sources of Law

2. A user is reading a particular section of an act (e.g., Act Section § 5 of Act §482/1991). He would like to find out what amendments correcting the act the chosen section belongs to came to force in 2005 (i.e., amendment of Act 482/1991 defined by Act 124/2005).
3. A user received a court decision (e.g., 20 Cdo 389/2004). There are various references to other court decisions and also sections of acts and amendments encoded in the text. He would like to see the reading of each of the referenced decisions and sections.

All these use cases are problematic because sources of law are published as textual documents by various authorities at different places of the Web. Moreover, the sources are not interlinked at all. Their logical structure (sections and their subsections) and links between them are encoded in the text in a way which can only be interpreted by a human. Therefore, the user can only read the sources and has to search for relationships manually. This is very time consuming, cumbersome and the user omits important relationships very easily.

3 Linked Open Data Representation of Czech Legislative Documents

Linked Open Data (LOD) is a set of principles of publishing data on the Web in a machine-readable form which enables to link related data. The links are recorded in a machine readable form and published on the Web as well as part of the data itself. LOD principles are simple:

– Use URIs to denote things.
– Use HTTP URIs so that things can be referred to and looked up by people and machines.
– Provide useful information about the things when its URI is looked up (use standards such as RDF [2] and SPARQL [6]).

– Include links to other related things (using their URIs) when published on the Web.

In our case, things are sources of law and their parts. Useful information about the things are metadata about the sources and their part (e.g., creation date, data from which a source is valid, author/publisher of the source, etc.). Links between the things are relationships between the sources (e.g., a section *is part of* an act, an act *amends* another act, court decision *cites* a section of an act, court decision *cancels* another court decision, etc.). Applying the principles to the legislative domain therefore means assigning HTTP URIs to sources of law and their parts, representing their metadata as RDF, extracting relationships among the sources of law from their original textual expressions and publishing all the metadata together with the links in RDF form so that the sources of law and their parts can be accessed via dereferencing their HTTP URIs or using SPARQL query language.

3.1 LEX Ontology

The first step in our effort to publish sources of law as LOD was designing an ontology. We call it *LEX ontology*. Its core classes and properties are depicted in Figure 2. Before we designed the ontology we analyzed the legislation domain and produced so called domain model. Then we identified what parts of the domain model had already been covered by existing ontologies. We found out that various parts were already covered by ontologies like FRBR[3] or Dublin Core. For these parts we therefore did not created new classes and properties but we reused existing classes and properties from these ontologies. On the other hand, various parts of the domain were not covered by existing ontologies and we had to develop new classes and properties. The resulting LEX ontology is, therefore, a mixture of existing ontologies and our new components.

There also exist various other approaches to describing the legal domain with ontologies. For example, in [5] on ontology for legal concepts was designed. The ontology is called LKIF. It is a detailed ontology of various legal concepts like roles (epistemic roles, functions, person roles, etc.), actions (processes which are performed by some actors in roles), etc. Paper [1] describes the results of a study consisting of two tasks: (i) how the "obligation" Fundamental Legal Concept is differently represented in the FrameNet[4] resource, in terms of Semantic Frames, and (ii) how the concept of "public function" stemmed from the "obligation" Fundamental Legal Concept can be ontologically characterized. However, such ontologies are too detailed for our purposes – our goal is to represent the structure of sources of law and relationships between them. These detailed ontologies aim at representing the semantics encoded in legislation which is a level beyond our current efforts.

Let us now describe the most important classes and properties of LEX ontology in a more detail. As we have already noted, legislation is defined by written

[3] Functional Requirements for Bibliographic Records (FRBR) `http://www.loc.gov/cds/downloads/FRBR.PDF`

[4] `https://framenet.icsi.berkeley.edu/fndrupal/`

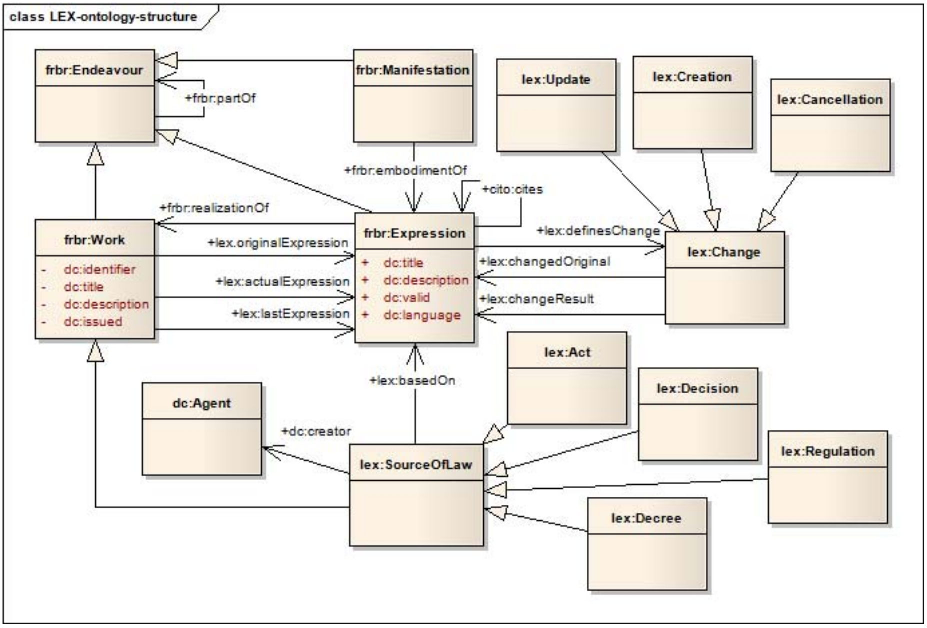


Fig. 2. LEX Ontology

documents called *sources of law*. There are different kinds of sources of law for which we introduce the following classes in our ontology: `lex:SourceOfLaw` for sources of law of all kinds (superclass of all other classes), `lex:Act` for acts, `lex:Decree` for decrees, `lex:Regulation` for regulations, and `lex:Decision` for court decisions.

Sources of law of most kinds (except of court decisions) exist in different versions. Some versions are outdated, at most one version is currently valid, and some versions are enacted but have not come to force yet. A particular version may have several physical embodiments, i.e. may be embodied in different physical documents (of different formats, e.g. HTML, PDF, XML, RDF, etc.). From this viewpoint, it is reasonable to represent a source of law as an abstract notion of intellectual creation which is independent of particular versions of the source. Moreover, each version of the source as well as its each physical embodiment should have representation on its own. This logic is built into the LEX ontology. However, we do not introduce own ontological constructs but reuse the Functional Requirements for Bibliographic Records (FRBR) ontology. We reuse the following three FRBR classes: `frbr:Work` for abstract notions of an intellectual creation which are sources of law, `frbr:Expression` for particular versions of sources of law, `frbr:Manifestation` for particular documents which are physical embodiments of particular versions of sources of law.

The usage of FRBR allows to distinguish a source of law itself, its particular versions and their physical embodiments. From the linked data point of view, it is therefore possible to link and query the source of law as an abstract entity

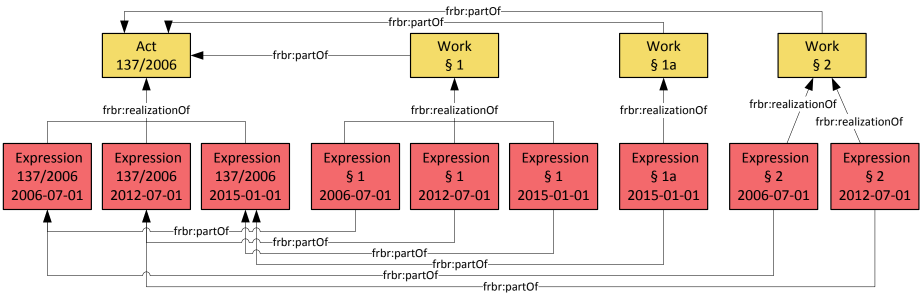


Fig. 3. Sample LEX Representation of an Act

which is independent of particular versions of the source. It is also possible to link and query its particular versions.

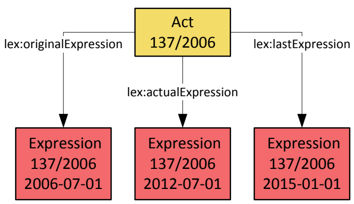
We also reuse two FRBR properties: `frbr:realizationOf` to link a version (member of `frbr:Expression`) to its source of law (member of `frbr:Work`), and `frbr:embodimentOf` to link a document (member of `frbr:Manifestation`) to a version of a source of law it is embodiment of (member of `frbr:Expression`).

Figure 3 shows a sample representation of a *Public Contracts Act* which is a part of the law of Czech Republic. The boxes in the upper part (yellow ones) are resources which are instances of `frbr:Work`. The left-most box is a resource representing the act itself. Therefore, it is not an instance of `frbr:Work` directly but an instance of its subclass `lex:Act` which is intended for acts. The other boxes at the upper levels represent parts of the act. These are represented simply as instances of `frbr:Work` linked to the resource representing the whole act. The fact that a given resource represents a part of an act cannot be recognized from its type. It is recognized from the fact that it is an instance of `frbr:Work` and that it is a part of a resource which is an instance of `lex:Act`. Next, there are various instances of `frbr:Expression` at the lower layer of the figure depicted (red boxes). Each such resource is linked to its corresponding instance of `frbr:Work` (or its subclass `lex:Act`, respectively) via the property `frbr:realizationOf` and models a particular version of that instance. In other words, the instances of `frbr:Expression` linked to the resource representing the act represent different versions of the act. Concretely, there is depicted a version of the act valid from `2006-07-01`, another version valid from `2012-07-01` and a version which will be valid from `2015-01-01`. The same is for the parts of the act. For part §`1`, we have three versions. In other words, the part has been present in all versions of the act. On the other hand, the part §`2` has only expressions valid from `2006-07-01` and `2012-07-01`. In other words, this part will be canceled in the version of the act valid from `2015-01-01`. Part §`1a` was added to the version of the act valid from `2012-07-01`. It was not present in the first version valid from `2006-07-01` as can be seen from the figure.

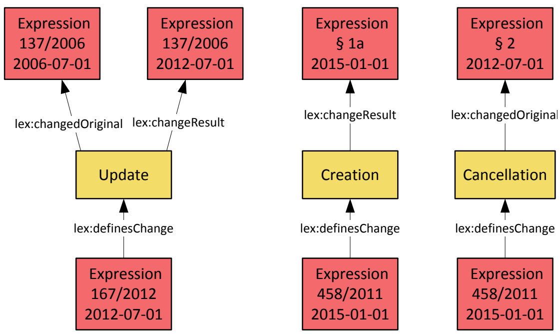
The example demonstrates the advantages of our proposed representation of sources of law and their parts. Anyone can link his or her data to as an abstract concept representing a source of law or any of its parts independently of a particular version. For example, it is possible to link §2 of *Public Contracts Act* independently of a particular version. But, it is also possible to link a particular version if necessary, e.g., §2 of *Public Contracts Act* valid from `2012-07-01`. This reflects real situations. In some situations we only know that, e.g., an act or its part is referenced. In that case we need to link an abstract concept (i.e. an instance of `frbr:Work`). In other situations we know exactly which version of the source of law is referenced. In that case we need to link a particular version (i.e. an instance of `frbr:Expression`).

For a given source of law we need to know its currently valid version, original version (i.e. the first version), and the last enacted version (which have not necessarily needed to come to force yet). For this, there is no corresponding property provided by FRBR. Therefore, we introduced three new properties in the LEX ontology: `lex:originalExpression` to link the original (first) version to the respective source of law, `lex:actualExpression` to link the currently valid version to the respective source of law, and `lex:lastExpression` to link the last enacted version to the respective source of law.

Figure 3.1 (a) shows *Public Contracts Act* as an instance of `lex:Act` and links to its original, actual and last expression. The links specifies that the first (original) version was the expression valid from `2006-07-01`. The actually valid expression is the one valid from `2012-07-01`. It will be later replaced by the last enacted expression in `2015-01-01`.



(a) Sample representation of original, actual and last expression of a work according to LEX ontology



(b) Sample representation of changes in an act according to LEX ontology

A source of law can change another source of law. For example, an act can be an amendment of one or more another acts. The same is true for other kinds of sources of law. A regulation or decree can change another regulation or decree, respectively, and a court decision can cancel another court decision (in case of court decisions, the situation is more specific – only cancelation is possible). For representing changes LEX ontology proposes a new class `lex:Change`. A change (instance of `lex:Change`) is defined by some source of law in some version. Therefore, the instance is referred from the respective expression (instance of `frbr:Expression`) which represents that version (via property `lex:definesChange`). Each change says that a given version of a given source of law (i.e. instance of `frbr:Expression`) is changed and a new version of the same source of law (i.e. another instance of `frbr:Expression` linked to the same instance of `frbr:Work` as the previous version) is created. Therefore, the resource representing the change is linked to the expression which is changed via property `lex:changedOriginal` and to the expression which is the result of the change via property `lex:changeResult`.

We distinguish three kinds of changes. LEX ontology proposes three respective subclasses of `lex:Change`: `lex:Creation` to express that an expression of a new source of law or new part of a source of law has been created by the defining expression; only `lex:changeResult` is present to link to the created expression, `lex:Cancellation` to express that an expression of an existing source of law or existing part of a source of law has been canceled by the defining expression; only `lex:changedOriginal` is present to link to the canceled expression, and `lex:Update` to express that an expression of an existing source of law or existing part of a source of law has been amended (updated) by the defining expression; both `lex:changedOriginal` and `lex:changeResult` are present to link the changed expression and the resulting expression.

Figure 3.1 (b) shows sample changes. The left-most update is an update. It is defined by an expression of act `167/2012` valid from `2012-07-01`. The update changes an expression of the act `137/2006` valid from `2006-07-01` to a new expression of the same act valid from `2012-07-01`. The second change is creation. It is defined by act `458/2011` valid from `2015-01-01`. The change creates a new expression which represents a new part §`1a` of the act `137/2006` which has not existed previously. The last change is cancelation. It is again defined by act `458/2011` valid from `2015-01-01`. The change cancels an expression which represents an existing part §`2` of the act `137/2006`.

3.2 Experimental Linked Data

On the base of LEX we transformed available sources of law available in Czech Republic to Linked Data representation according to the ontology. There are several web sites where data about sources of law can be accessed in Czech Republic. First, it is the web site of The National Parliament. Here, meta-data about each act, decree and regulation can be obtained in a form of HTML pages. It is also possible to get information that a given source of law changes another source of law (i.e. updates, creates or changes). However, this is only at the level of whole sources of law. The parliament does not provide information about parts of the sources of law. Another source of information is the National Governmental Portal. Here, it is possible to access original expressions of all acts, regulations and decrees. Also, consolidated versions of some of these sources of law can be accessed here (approx. 20 % of all of published sources of law). However, the national portal only provides texts, there are no information about changes published.

We have already processed both sources of data and translated them to LOD. The third place where sources of law can be obtained are web sites of courts in Czech Republic, namely the Supreme Court of the Czech Republic and The Constitutional Court of the Czech Republic. Both publish their decisions on their web sites. Other courts in the Czech Republic do not publish their decisions on their web sites. However, we have not processed this sources yet.

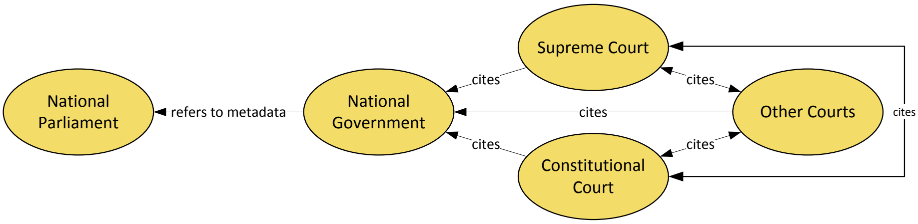


Fig. 4. Datasets converted to LOD according to LEX ontology

The goal of our work was to transform these available sources to Linked Data. For each data source mentioned above we created its RDF equivalent. In the following subsections, we describe translation of each data source in a more detail. Data sources and links between them are visualized in Figure 4. Currently, we have fully translated sources of law from the National Parliament of Czech Republic and the National Government of the Czech Republic. The resulting numbers of resources and RDF statements are depicted in Table 1 summarizes the amounts of resources and statements obtained from each data source. The National Parliament publishes metadata for each act, regulation, decree and decision of the Constitutional Court since 1945. It also contains information about updates of acts. The National Government publishes content of acts, regulations, decrees and decisions, some of them in current consolidated versions. It does not publish structured information about updates and metadata. As shown in Figure 4, there are links from the governmental data set to the parliamentary data set. Each instance of `lex:Act`, `lex:Regulation`, `lex:Decree` and `frbr:Expression` in the governmental data set is linked to its equivalent in the parliamentary data set where further meta data and information about changes in acts (amendments) are published.

Let us note that LOD representation of court decisions published at Supreme and Constitutional Court web sites is not listed in the table. The reason is that we still work on the translation and currently we have only several (approx. 500) experimental decisions translated. This is not a complete set of court decisions.

Parliament of the Czech Republic. From this data source we obtained by a proprietary web scraper meta data about each act, decree and regulation issued in our country since 1945. We represented each obtained source of law as an RDF resource. In other words, we had to create a HTTP URI for each such resource and create RDF triples recording metadata for each source. For constructing URIs we chose the following pattern:

Table 1. Number of resources and statements in the created LOD representation of sources of law obtained from various public authorities. The numbers at each row are presented in the following order: # `frbr:Work`, # `lex:Act`, # `lex:Regulation`, # `lex:Decree`, # `lex:Decision`, # `frbr:Expression`, `lex:Change`, # RDF statements.

Authority data set	# 1	2	3	# 4	# 5	# 6	7	8
National Parliament	0	6372	8238	3149	1069	33238	20640	710950
National Government	552690	6372	8238	3149	0	617445	0	4789725

$$\texttt{http}://\texttt{linked.opendata.cz/resource/legislation/cz/}$$
$$\{\texttt{sourceOfLawKind}\}/\{\texttt{validFromYear}\}/\{\texttt{number}\}$$

where `sourceOfLawKind` is `act`, `decree`, or `regulation`. `ValidFromYear` is the year in which the source of law became or will become valid. And, `number` is the number of the source of law issued by the parliament. All this information is present at the parliament web site and we were able to scrape it using our scrapers. Moreover, for each source of law we were able to scrape the list of other sources of law changed by this source. In the list there is the number of the changed source of law and the kind of change (i.e. creation, cancelation, or update).

As the result, we created an RDF data set where each act, decree and regulation issued in Czech Republic since 1945 is represented as an instance of the respective subclass of `lex:SourceOfLaw` with RDF statements specifying the number of the source of law, its title and validity date. From the list of changes for each source of law we were able also to reconstruct RDF resources representing all expressions of the sources. For constructing URIs of the expressions, we chose the following pattern:

$$\texttt{URIofSourceOfLaw/version/cz/}\{\texttt{validityDate}\}$$

where `URIofSourceOfLaw` is the HTTP URI of the source of law the expression realizes. `ValidityDate` is the date from which the expression is valid.

Figure 5 shows a part of the obtained RDF representation of the act `137/2006`. The representation is serialized in Turtle format.

National Portal of Czech Government. We built a proprietary scraper which downloads sources of law published by the portal (acts, regulations, decrees). We parsed this representation so that parts and sections of each source of law were obtained. However, the obtained textual representations were only original expressions of the sources of law. The other expressions (i.e. versions) of the sources could not be obtained in most cases because the Czech government does not produce consolidated texts of all acts (actual consolidated version of only 20 % of acts is published by the government). Moreover, no historical consolidated versions are published.

```
<http://linked.opendata.cz/resource/legislation/cz/act/2006/137-2006>
    rdf:type lex:Act ; dcterms:title "Zakon o verejnych zakazkach"@cs ;
    dcterms:identifier "137/2006" ;
    dcterms:issued "2006-04-19" ;
    lex:actualExpression <http://linked.opendata.cz/resource/
                            legislation/cz/act/2006/137-2006/
                                version/cz/2012-07-01> .

<http://linked.opendata.cz/resource/legislation/cz/act/2006/
        137-2006/version/cz/2012-07-01> rdf:type frbr:Expression ;
    dcterms:title "Zakon o verejnych zakazkach"@cs ;
    dcterms:issued "2012-07-01" ;
    frbr:realizationOf <http://linked.opendata.cz/resource/
                            legislation/cz/act/2006/137-2006> .
```

Fig. 5. RDF representation of a sample act (serialized in Turtle)

```
<http://linked.opendata.cz/resource/legislation/cz/act/2006/137-2006>
    dcterms:description [CONTENT-OF-ORIGINAL-VERSION] .

<http://linked.opendata.cz/resource/legislation/cz/act/2006/137-2006/
        section/1>  a frbr:Work ; dcterms:identifier "1" ;
    frbr:partOf <http://linked.opendata.cz/resource/legislation/cz/
        act/2006/137-2006> .

<http://linked.opendata.cz/resource/legislation/cz/act/2006/137-2006/
        section/2> a frbr:Work ; dcterms:identifier "2" ;
    frbr:partOf <http://linked.opendata.cz/resource/legislation/cz/
        act/2006/137-2006> .
```

Fig. 6. RDF representation of a sample act (serialized in Turtle)

Therefore, the result is that we gathered original expressions of all acts, regulations and decrees and obtained the RDF representation of all their parts and sections according to the LEX ontology. This could not be gathered from the parliament. On the other hand we do not have information about changes. These are present only at the web site of the parliament. The resulting RDF representation is linked to the representation obtained from the parliament based on the Linked Data principles.

Figure 6 shows a part of the obtained RDF representation of the act `137/2006`. The representation is serialized in Turtle format. It shows how the act is structured to its sections (only the first two main sections are depicted, the act contains 1481 sections in total). The figure also demonstrates how links to the previous data set are expressed – a given source of law has the same HTTP URI in both data sets. Therefore, the link is realized implicitly. This is possible because we store both data sets in the same database and publish them in the same HTTP domain. If we would publish them in two different domains, resources representing the same source of law in different data sets would have differnt HTTP URIs. These URIs could be linked on each other using the OWL predicate `owl:sameAs`.

4 Conclusions

We presented an experiment with publishing legislative documents as Linked Open Data. We demonstrated the benefits LOD principles bring to users of legislation (common citizens, lawyers, etc.). We proposed an ontology for publishing structured legislative documents and relationships between them. We also converted two existing data sources in Czech Republic to the proposed representation. In the future work we will concentrate on publishing court decisions in the Czech Republic and linking them to the already published LOD data sets.

References

1. Agnoloni, T., Fernández Barrera, M., Sagri, M.T., Tiscorni, D., Venturi, G.: When a FrameNet-Style Knowledge Description Meets an Ontological Characterization of Fundamental Legal Concepts. In: Casanovas, P., Pagallo, U., Sartor, G., Ajani, G. (eds.) AICOL-II/JURIX 2009. LNCS, vol. 6237, pp. 93–112. Springer, Heidelberg (2010), http://dx.doi.org/10.1007/978-3-642-16524-5_7
2. Beckett, D.: RDF/XML Syntax Specification (Revised). W3C (February 2004), http://www.w3.org/TR/rdf-syntax-grammar/
3. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data – The Story So Far. International Journal on Semantic Web and Information Systems 5(3), 1–22 (2009)
4. Hoekstra, R.: The metalex document server - legal documents as versioned linked data. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., Blomqvist, E. (eds.) ISWC 2011, Part II. LNCS, vol. 7032, pp. 128–143. Springer, Heidelberg (2011)
5. Rubino, R., Rotolo, A., Sartor, G.: An owl ontology of fundamental legal concepts. In: Proceedings of the 2006 Conference on Legal Knowledge and Information Systems: JURIX 2006: The Nineteenth Annual Conference, pp. 101–110. IOS Press, Amsterdam (2006)
6. Harris, A.S.S.: SPARQL 1.1 Query Language. W3C (March 2013), http://www.w3.org/TR/sparql11-query/

Causality and Other Aspects of Instrumental Reasoning

Tomasz Żurek

Institute of Computer Science, Maria Curie-Sklodowska University, Pl. Marii Curie Sklodowskiej 1, 20-031 Lublin Poland
zurek@kft.umcs.lublin.pl

Abstract. Instrumental inference is one of the most interesting ways of reasoning, commonly used in legal practice. One of its characteristic features is the causal necessity relation that must occur between the reference actions. The aim of this study is to present and discuss model of causal necessity specific to instrumental inference, in the context of its application in the legal advisory system.

Keywords: Legal Expert Systems, Instrumental reasoning, Causation.

Introduction

Instrumental inference is one of the most interesting ways of reasoning, commonly used in legal practice. It refers to deontic notions and takes three different forms, i.e. instrumental admission, instrumental prohibition and instrumental obligation. One of their characteristic features is the causal necessity relation that must occur between the reference actions (i.e. one action must be causally necessary to implement either the obligation or the admission), or alternatively the relation in which one action is causally sufficient not to implement the obligation. According to the author, the major problem related to formalising instrumental inference lies in the formal description of the aforementioned relations of necessity and sufficiency to implement, or not to implement, the obligations, bans or admissions in question. [22] features a preliminary model of instrumental inference. Its major drawback stems from a rather questionable formalisation of causal necessity, which definitely seems worth further analysis. In light of the above, the main aim of this study is to present and discuss model of causal necessity specific to instrumental inference, in the context of its application in the legal advisory system.

Further discussions should also focus on the issues related to the application of an instrumental reasoning model in argumentation, including especially the problem of solving potential conflicts between the arguments employing instrumental reasoning.

1 Instrumental Reasoning

The principles of instrumental obligation, prohibition, and permission are among clearly defined legal mechanisms employed to deal with legal cases which are not

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 184–195, 2013.

expressly regulated in the law. The principle of instrumental obligation assumes that the existence of a norm imposing an obligation to do X causes the law abiding entity to be also bound by the norm imposing an obligation to do Y, which constitutes a causal necessity for the obligation to do X [15]. The principle of instrumental prohibition assumes that that the existence of a norm imposing an obligation to do X causes the law abiding entity to be also bound by the norm imposing a prohibition on Y which is sufficient, in terms of causality, not to perform the obligation [15]. The principle of instrumental permission assumes that the law abiding entity bound by norm X permitting a certain situation, may consider norm Y as binding, provided that it permits a certain act which constitutes a causal necessity for X [15].

As required by the theory of law, all these principles should be reasonably applied, especially within the penal and tax laws where obligations must be defined expressis verbis.

2 Related Works

[22], which presents a model of instrumental inference, together with a related causal necessity model, serves as the principal point of reference for this study. To express causal necessity, reference was made to modal logic, and in particular to temporal logic, including related functor R. (release: p releases q if q is true until the first position in which p is true) [22]. Assuming A as a set of actions, and having two actions, x and y, which belong to action classes X and Y ($x \in X$, $y \in Y$, $X \in A$ and $Y \in A$), the action class X will be considered causally necessary to implement action class Y, in which case the action class X will always precede the action class Y, which can be expressed as follows:

X R ¬ Y

It is worth noting that the relation so modelled does not define the source of this causal necessity but only its outcome. The former is associated with the nature of this phenomenon, and with the relations occurring between specific knowledge actions.

Based on the above causal necessity model, the principle of instrumental admission was modelled as follows: A is a set of actions, X, Y and Z are action classes.

When we assume that: X R ¬ Z

and a rule of law:

$Y \Rightarrow \ perm(Z)$

then:

$(Y \Rightarrow perm(Z)) \wedge \sim forb(X) \Rightarrow (Y \Rightarrow perm(X))$

The principle of instrumental obligation is similar to that of instrumental permission, and it assumes that a certain rule of obligation N makes the law abiding entity recognise the validity of another rule of obligation M which is causally necessary to realise the obligation N. As in the case of instrumental permission, this principle is based on a causal necessity model:

When we assume that: X R ¬ Y

and a rule of law:
$Z \Rightarrow obl(Y)$
then:
$(Z \Rightarrow perm(Y)) \wedge \sim forb(X) \Rightarrow (Z \Rightarrow perm(X))$

The third principle, i.e. the principle of instrumental prohibition, assumes that the existing rule of obligation N makes the law abiding entity recognise the validity of a prohibiting rule M which is causally sufficient not to realise the obligation.

Action X is causally sufficient not to realise the obligation of action Y whenever the occurrence of action X is connected with the non-occurrence of action Y.

This principle is based on the assumption of a prohibiting rule:

Z ⇒ obl(Y) The action class X should be causally sufficient not to realise the prohibition, which can be modelled in a more simplified way than the causal necessity described on previous examples:

$X \rightarrow \neg Y$

In this case, the principle of instrumental prohibition can be presented as follows: $(Z \Rightarrow obl(Y)) \Rightarrow (Z \Rightarrow forb(X))$

3 Other Works

The issue of modelling and implementing instrumental inference is rather underrepresented in the literature concerning the use of artificial intelligence tools in law. A simplified and highly imperfect kind of such reasoning was described in the studies by [21]. Unfortunately, the example provided therein deals exclusively with a very specific instance where causal necessity arises directly from legal principles. An argumentation model, which slightly resembles the instrumental inference model discussed here, was briefly presented in [18]. It was referred to as practical reasoning, which entails arguing from a goal to an action required to realize the goal. However, the said study fails to provide any details regarding this model.

Instrumental inference requires knowledge on the relation of causal necessity (or sufficiency not to implement) which may occur between the actions or phenomena examined. In order to get a more accurate picture of the said relation, let us refer to the notion of causation as presented elsewhere in the literature. First of all, one should note a clear-cut distinction, introduced in certain publications The authors in [4] point out that even philosophy has certain problems with causation. According to one approach, it is treated as a secondary notion, arising from basic relations, whereas others claim that it is one of the non-defined underlying terms. The authors in [4] have decided not to investigate this theoretical issue, instead focusing on concrete examples. Nevertheless, this dispute translates itself into the issues analysed here, as it boils down to the most basic question of whether it is possible to create a causation model consistent enough to determine (e.g., by referring to logic and inference) which events cause others, based on the pre-defined principal notions included in the knowledge base.

Should this be possible, then a properly constructed database would allow for determining which events cause others. Otherwise, the cause and effect relations between various events would have to be expressly declared in the database. Several different approaches to causation modelling in the artificial intelligence systems are presented below. The issue of causation is presented in detail in [6]. In general, the model described focuses on causality in the context of transitions between various states and the causes behind them. At the beginning, the authors introduce a modal operator C, denoting the cause, together with the general causal principles ($A \supset CB$ implies that action A causes action B, which is then expressed as $B \Leftarrow A$). In the following sections of the reference study, following the model construction, the authors refer to the fact that the action precedes the effect (which also provided the reason for using temporal logic in the model described in [22]). The authors however fail to present the model of causal necessity, i.e. the situation in which no change of states is possible without performing a specific action A. J. Hage in [7], as part of the Reason Based Logic presented, defines the notion of reason as the fact which makes that other facts obtain or do not obtain. He makes a distinction between decisive reasons (which guarantee the presence of the fact for which it is a reason, or the absence of the fact against which it is a reason) and contributive reasons (which do not guarantee the facts for which they plead). The decisive reason (DR) predicate implies that a fact is a decisive reason for some other fact. The authors of [3] propose to include the knowledge on causation in the ontology describing the realities of a legal act. Valente in [3] [20] describes the Functional Ontology of Law, which involves a division into normative knowledge, world knowledge, responsibility knowledge, meta-legal knowledge and creative knowledge. The second type, i.e. world knowledge, comprises the knowledge on causation. Referring to the studies on the FOL, Lehmann et al. attempted at creating the ontology describing the nature of cause and effect relations. [10] presents the ontology of CAUSATIOnt which models the issues of causation and legal responsibility. Causation was defined as an occurrence of two events, the cause and the effect, and it was divided into physical causation and agent causation:

Physical causation is causation between an event E1, which is an occurrence of a physical process P1 (the occurrence of) involving an object O1 (the subject), and event E2, which is an occurrence of a physical process P2 (the occurrence of) involving an object O2 (the subject). A relation of physical causation holds between E1, the cause, and E2, the effect, if the specific conditions connected with the objects, time, energy transfer and directions of the transfer are met [10].

The notions of energy-form, direction, power and dimension characterise the CausatioNT ontology process, and they are described in further detail in [10]. One should also note that physical causation is a transitional relation, and the fact that the result precedes the cause is one of its specific features. In [11] the authors describe the modelling and representation of causation in their DOLCE ontology. The model is based on the analysis of quality changes in particular ontology objects. The reference changes, similar to other models, take into consideration the timing factor (the effect being preceded by the cause) as well as dependency (structural, causal and occasional), which allows for describing the impact of various events on each other and, at further stages, for assessing causal dependency between them. The authors present their theory, inter alia, against the logic based approach, in which:

C is a necessary condition of E (E $\rightarrow$ C); C is a sufficient condition of E (C $\rightarrow$ E); C is a necessary and sufficient condition of E (C $\leftrightarrow$ E) and probabilistic in which, the conditional probability of E given C is higher than the conditional probability of E given not C ($p(E|C) > p(E|\neg C)$).

Probability Theory, similarly to Logic, provides no principled way for distinguishing between genuine causes and mere co-occurrences.

The issues of cause and effect relations in the probabilistic context have been widely discussed in the studies by J. Pearl (i.e. [16]. Probabilistics is probably the most common tool used in causation modelling. However, the specificity of instrumental inference raises serious issues of application justifiability. The occurrence of causal necessity and sufficiency not to implement, given their categorical character, force the declaration of probability values, sufficient to be qualified as the aforementioned cases, which appears rather questionable and sometimes hard to realise.

A distinct approach to causation modelling attempts at defining the very nature of this phenomenon. Such models usually require implementation of very comprehensive knowledge as the scope of possible impacts of certain actions on others is extremely wide and multi-faceted. Analysing the impact of two events on one another, it can be noted that it may stem from various sources (a variety of physical phenomena). It can be strongly time-shifted, or it may arise from such relations which are hard to define, or it may be indirect (A causes B by affecting a different action), etc. With such a wide, complicated and hard to explain phenomenon, it may turn out that its entire model (reflecting its intrinsic nature) may be useless, considering the complexity level and the scope of required but unavailable knowledge. R. Wright, in his classic work [19] focuses on the problem of causation in tort law. He criticises, i.a, the so-called 'but-for test which states that an act (omission, condition, etc.) was a cause of an injury if and only if but for the act, the injury would not have occured. This definition fails in the cases of pre-emptive causation and duplicative causation. In his study, he also presents the solutions proposed by various authors with a view to evading the reference restrictions, i.a., by means of detailing the manner in which the injury occured, detailing the injury, excluding hypothetical facts and aggregating multiple potencial causes, unfortunately, as proven by the author, all those mechanisms turn out fallible as they do not fully overcome the problems arising from the 'but-for test' restrictions. According to the author, the so-called NESS test (Necessary element of sufficient set) is much more accurate.

The essence of the concept of causation is that a particular condition was a cause of (condition contributing to) a specific consequence if and only if it was a necessery element of a set of antecedent actual conditions that was sufficient for the occurence of the consequence [19]. Wright also writes, that NESS is based on conception of D. Hume: "Hume rejected earlier notion that we acquire causal knowledge through direct sensory perception of causal qualities or forces inherent in objects or events. Instead he insisted that we only observe certain succesions of events, more or less frequently repeated. From these observations we inductively derive the belief that certain antecedent events are not only always cojoined with, but also are sufficient for the occurence of, certain subsequent events".

The theory of causal singularity is distinguished as one of the most interesting philosophic theories of causation. It stipulates that: "Causation is a relation which holds essentially between single, individual events though it may of course be generalized, and propositions containing kinds of events then be formulated. (...) The cause of a particular event [is defined] in terms of but a single occurrence of it, and thus in no way involves the supposition that it, or any like it, ever has occurred before or ever will again. The supposition of recurrence is thus wholly irrelevant to the meaning of cause." [5].

Such an approach to causation excludes any general rules governing causation. In other words, the fact that C caused E does not lead to a general rule that C-type event causes E-type event. However, without further inquiring into whether causal singularity excludes the possibility of any universal causal rules to apply, it may seem justified to assume that there may be such rules (C causes E) which merely imply that C *could* causes E. This assumption could then be used in the instrumental reasoning model.

Many authors abstain from modelling the nature of causation itself, assuming that the cause and effect relationship is given in advance. They are more inclined to denote causation by formal means.

Kanger ([8] and Lindhal ([12]) make use of modal operator E_j, where E_jA means that agent j brings it about that A. Such a formalism is often connected with deontic logic and allows for reasoning on the responsibility of various agents for various actions and their effects. At this point, it is worth noting that it is the actor and not the act that is of crucial importance to formalism, as in the reference case the latter is not defined at all. As a result, this mechanism is not well-suited to modelling instrumental modelling.

A different attitude is presented by F. Bex and B. Verhij in [1]. Their model of evidential reasoning makes use of the causal rules in which $e_1 \Rightarrow_c e_2$ stands for e1 (presumably) caused e2, where e1 and e2 represent events. This causal rule is defeasible.

4 Causal Necessity in Instrumental Inference

In general, the aforementioned methods of describing and defining causation relations can be divided into two basic groups. The first group defines causation using the analysis of event consequences (logic and probabilistics based methods) while the second one is based on including commonsense knowledge in the model, on exploring the causation phenomenon, or on adding the knowledge regarding the clauses representing cause and effect relations to the model constructed. Referring to the afore-mentioned view expressed by D. Hume, it transpires that even the application of the most refined model in order to reflect and describe the influence of one event on another, which would make it possible to infer the causal relations between them, does not ensure a thorough understanding of all such relations. It should be borne in mind that this model would require a huge amount of knowledge of the features and characteristics of the phenomena analysed, which would be difficult to gather, and which would not ensure correct reasoning. As a result, at least for the time being, the author decides to assume that the causal relations in question may only be analysed inductively, based on the previous cases, and even then human commonsense knowledge (which is also not fool-proof) would be required. Once detected and examined, a causal relation will be expressed through predicate $p(X_i, a)$, where $X_i \in A$ is a set of causal actions whereas $a \in A$ is the effect. It will mean that the set of actions X_i previously led to effect a. In this way a set of instances is established, allowing for the inference that the set of actions X_i may cause a (because it has already done so at least once).

Such a presentation of the causation problem gives rise to two major problems. Firstly, we treat causation as primary knowledge one may proceed with an inductive analysis of previous cases, but it cannot be inferred in an unambiguous and non-defeasible way that event X will be the cause for Y. Secondly (which is partly based on the first argument), the pre-declared cause and effect relation must only be considered potential. While it tells us that X may be the cause for Y (as it has happened before at least once), there is no evidence that in a new case it has to be so. The declared cause and effect relation may therefore be only the outcome of a certain inductive reasoning, resulting from the analysis of previous instances (intuitive or intuitive-supported, e.g. probabilistic). This is indirectly related also to the singular causality theory, which also assumes the lack of universal causal relations. The inductive methods of analysing causation, which have been discussed in detail by many authors, are not the subject of this work.

Furthermore, it is worth noting that causal relations may be quite complex and they can be considered at various levels of detail. For example, if a child ran away from school and, while playing truant, kicked the ball which broke the window, then what is the cause for the broken window? Kicking the ball towards the window? The ball hitting the window? Teachers failing to take proper care of the child? Generally, in many cases, the inter-event causation chain may be easily extended or scattered by dividing the actions examined into more detailed events.

Obviously, it is impossible to decide in advance on a certain level of detail in the description of the phenomena analysed, which is why the model should make it possible to further divide individual causes into sub-causes.

As a result, it should be assumed that among the actions from the set of all possible actions A it may happen that action a_i actually consists of several actions, e.g. $(a_1 \wedge a_2) = a_i$ means that action a_i comprises actions a_1 and a_2. A situation may also be encountered in which $(a_1 \vee a_2) = a_j$, where $a_1 \in a_j$ and $a_2 \in a_j$.

Another issue worth discussing is the fact that a given causation relation may also be a generalisation of certain sub-relations, e.g. the child kicked the ball, the wind blew and slightly changed its trajectory, the ball hit the window and its force caused the latter to break. Such an event may either be considered as $p(child\_kicked\_ball, broken\_window)$, or it may be presented in more detail, e.g.: $p(child\_kicked\_ball, ball\_flies\_fast) \wedge p(ball\_fast\_flies, powerful\_hit\_in\_window)$ $\wedge p(powerful\_hit\_in\_window, broken\_window)$

Considering the possibility of presenting causal sequences at various levels of detail, the notion of direct causation can be defined, i.e. the cause which led to a certain effect without any intermediate events. It can be defined as follows::

Assuming that $C \in A$ and $R \in A$, a X represents a variable and if

$\sim \exists_X (p(C, R) \wedge p(C, X) \wedge p(X, R))$

then C is the direct cause.

Looking at this problem from a different angle, we may assume the transition of the causal relation, i.e.:

$p(C, M) \wedge p(M, E) \Rightarrow p(C, E)$

A mere declaration of which actions may serve, or actually serve, as the cause for other actions does not solve the problem of identifying the necessary and sufficient conditions, which is crucial to instrumental reasoning. A certain set of all causation relations, expressed through predicate $p(C, R)$., is therefore assumed.

Whether a given cause (or a set of causes) is necessary or sufficient should be examined by analysing the reference set in terms of satisfying the conditions of necessity or sufficiency.

Since the necessary condition should intrinsically appear in all instances of relation $p(C, as)$ (where $C \in A$ and $as \in A$ and as are actions caused by C), the decision on whether a given phenomenon reflects a condition necessary for a given effect as to arise will entail searching through the set of causes and finding this condition which has previously occurred in all instances. At the same time, as was also noted by D. Hume, it is essential distinguish between those phenomena on the causal side which served as the real cause and those that constituted a mere coincidence. Unfortunately, without an additional commonsense knowledge, such a distinction can hardly be made. Assuming however the primary character of causal relations, the problem with their identification somehow translates into the stage of declaring causal relationship.

The general assumptions:

Firstly, a set of actions A=a1, a2,..., an is assumed.

The set of instances C described by actions from set A.

The set of causal relations P described by predicate $p(Ci, as)$, where $Ci \in A$ is a set of conditional actions while $as \in A$ is the effect.

The necessary conditions:

Assumptions: $R \in P$ – a set of causal relations taking the form of $p(Xi, as)$, where Xi is a set of conditions i of this relation, X is a set of sets of conditions ($Xi \in X$), and as is the effect (the same for all relations from set R). If $\forall_{(Xi \in X)} \exists_C : (p(Xi, as) \wedge C \in Xi)$

then the set of conditions C is the set necessary to implement as.

The sufficient conditions:

Assumptions: $R \in P$ – a set of causal relations taking the form of $p(Xi, as)$, where Xi is a set of conditions i of this relation, X is a set of sets of conditions ($Xi \in X$), as is the effect (the same for all relations from set R). If:

$\exists_{(Xi \in X)} \neg \exists_{(Xj \in X)} : (p(Xi, as) \wedge p(Xj, as) \wedge Xj \in Xi)$

then the list of conditions Xi is sufficient to implement the objective of as

The instrumental ban requires the condition to be sufficient not to implement a given action, as a result of which, if

$\exists_{(Xi \in X)} \neg \exists_{(Xj \in X)} : (p(Xi, \neg as) \wedge p(Xj, \neg as) \wedge Xj \in Xi)$

then the list of conditions Xi will be sufficient not to implement as.

The above model is based, to a certain extent, on an attempt characteristic of the so-called logic models although it requires an additional pre-declared knowledge of the previous causal relations. While it does not allow for determining whether action X caused action Y (which as we should note is the task prone to errors also when done by a man), it makes it possible to describe causal relations in a simple way and, in particular, to model the conditions necessary for a certain event to occur, or sufficient for it not to occur.

Unfortunately, it should be borne in mind that such causal necessity (and the sufficiency not to implement) fails to guarantee a certainty of proper identification of the phenomena at issue (as a man can also be wrong here). Therefore, the principle of instrumental reasoning based on such necessary and sufficient conditions must be defeasible, so as to allow for it being defeated by another argument.

5 Instrumental Reasoning in Argumentation

Instrumental reasoning may serve not only as a formal base of the legal expert system, but also as a tool of argumentation modelling. This gives rise to problematic situations in which there is a conflict between the instrumental reasoning conclusion and the result obtained through other reasoning model. In the model described in [22], the result of instrumental reasoning may be defeated by a different reasoning method (equations No. 2 and 4 were complemented by the clauses verifying whether the instrumental reasoning result is not defeated by any other argument). For example, some time ago the Polish law allowed for the possession of a small amount of marihuana for private use while separate provisions banned its sales, import and cultivation (and all other ways of obtaining it). Referring to the principle of instrumental permission, at least one way of getting the drug should be allowed. Practically speaking, the sales, purchase, import or cultivation was subject to penalty whereas the possession of a small amount was not (which clearly contradicts the commonsense; in fact, later on the provisions were exacerbated and a ban on any possession of marihuana was introduced).

The model developed in [22] allows for presenting the above case: $T \Rightarrow forb(obtaining)$ $obtaining = (cultivation \vee import \vee purchase)$
$T \Rightarrow perm(possesing)$

We know than in order to possess marihuana, you may first obtain it from a certain source $p(obtaining, possessing)$

As there is no other cause for possessing apart from obtaining the drug (by way of as cultivation, import or purchase), *obtaining* is the necessary condition and by applying the principle of instrumental permission, the following can be inferred:

$perm(possessing) \wedge \neg forb(obtaining) \Rightarrow perm(obtaining)$

The above clauses therefore imply that *obtaining* (in view of the principle of law which directly regulates the case under analysis) should be banned even though instrumental reasoning would lead to a contradictory conclusion.

The issue of conflicting arguments was dealt with by many authors, i.e., in several studies by G. Sartor and H. Prakken (i.a. [17], [14]) and others (i.e. [13], or [2]). In the above case, the conflict arises because of the complementary nature of two conclusions drawn from distinct arguments. This situation seems interesting because each of these arguments employs a different reasoning method. One uses the traditional modus ponens based on the rule of statutory law whereas the other refers to instrumental reasoning (the issue discussed here makes reference to statutory law and to the argumentation based on the rules of statutory law). Collisions between various rules (e.g. in the Sartor and Prakkens logic [17], [14], or in other systems: [13], [2] [24]) are usually dealt with by assuming that the conflicting rules exhibit a certain order reflecting certain preferences, which may result from one of the rules eliminating collisions between legal rules (e.g. *lex specialis*, or *lex posteriori*). In the case of instrumental inference, the situation is slightly more complicated as there is a collision between the rule of statutory law and the result of instrumental reasoning which is also based on a certain standard, though it does not necessarily share its priority.

The model described in [22] can be modified so as to integrate in within any of the existing argumentation systems (e.g.:[17], [14] [13], [2]). The changes introduced may boil down to removing the clauses ensuring the coherence of reasoning (see [22] equations No. 2 and 4) and the assumption that the argument based on instrumental reasoning should be weaker than the argument based on the rule of law. As a result, the argument flowing from instrumental inference would only be applicable to the situation not directly regulated by the law.

6 Summary

Instrumental inference is a specific way of reasoning used in the law. It goes in line with several different mechanisms which make it possible to infer the deontic state of a given action, in the lack of any legal principles that would expressly regulate the case under analysis. The need to investigate whether the relation of causal necessity, or sufficiency not to implement, occurs between the actions analysed is one of the key elements of the instrumental inference. This study is an extension of the article [22], presenting approach to modelling the cause and effect relation, and the necessary and sufficient conditions. It especially provides a consistent tool to describe and assess the necessary and sufficient conditions, required in order to correctly present instrumental reasoning. This model also allows for operating at different levels of generality of analysed actions, as a result of which it may be adjusted to the descriptive level of detail, required in a given moment. This work also presents the problem of collision between the arguments constructed on the basis of classic and instrumental reasoning.

Further studies will relate to the development and improvement of the reference models, and to the integration of the instrumental inference model with other inference models, including non-formal ones (for example [23]). Another interesting task into which instrumental reasoning will be involved is realisation of system of evaluation of quality of protection of computer systems (the work will be basing on [9]) Utilisation of instrumental reasoning may help while dealing with systems which are not strictly described in knowledge base.

References

1. Bex, F., Verheij, B.: Legal stories and the process of proof. In: Artificial Intelligence and Law, pp. 1–26 (2012)
2. Brewka, G.: Reasoning about priorities in default logic. In: Proceedings AAAI 1994, pp. 940-945 (1995)
3. Breuker, J., Valente, A., Winkels, R.: Legal Ontologies in Knowledge Engineering and Information Management. Artificial Intelligence and Law 12, 241–277 (2004)
4. Breuker, J., Hoekstra, R.: Commonsense Causal Explanation in a Legal Domain. Artificial Intelligence and Law 15(3), 281–299 (2007)
5. Ducasse, C.J.: On the Nature and Observability of the Causal Relation. Journal of Philosophy XXIII(3), 125–136 (1926); Reprinted in Sosa and Tooley (1993)
6. Giunchiglia, E., Lee, J., Lifschitz, V., Mccain, N., Turner, H.: Nonmonotonic Causal Theories. Artificiai Intelligence 153, 49–104 (2004)
7. Hage, J.: Studies in Legal Logic. Springer (2005)
8. Kanger, S.: Law and logic. Theoria 38, 105–132 (1972)
9. Ksiezopolski, B.: QoP-ML: Quality of protection modelling language for cryptographic protocols. Computers & Security 31(4), 569–596 (2012)
10. Lehmann, J., Breuker, J., Brouwer, B.: $CAUSATIO^{NT}$: Modeling Causation in AI&Law. In: Benjamins, V.R., Casanovas, P., Breuker, J., Gangemi, A. (eds.) Law and the Semantic Web. LNCS (LNAI), vol. 3369, pp. 77–96. Springer, Heidelberg (2005)
11. Lehmann, J., Borgo, S., Masolo, C., Gangemi, A.: Causality and causation in DOLCE. In: Formal Ontology in Information Systems, Proceedings of the International Conference FOIS, pp. 273–284 (2004)
12. Lindahl, L.: Position and change: a study in law and logic. Synthese library, vol. 112. D. Reidel, Dordrecht (1977)
13. Nute, D.: Defeasible logic. In: Bartenstein, O., Geske, U., Hannebauer, M., Yoshie, O. (eds.) INAP 2001. LNCS (LNAI), vol. 2543, pp. 151–169. Springer, Heidelberg (2003)
14. Prakken, H., Sartor, G.: A system for defeasible argumentation, with defeasible priorities. In: Gabbay, D.M., Ohlbach, H.J. (eds.) FAPR 1996. LNCS, vol. 1085, pp. 510–524. Springer, Heidelberg (1996)
15. Leszczynski, Zagadnienia teorii stosowania prawa: Issues of theory of application of law, Zakamycze, Krakow (2001)

16. Pearl, J.: Embracing Causality in al Reasoning. Artificial Intelligence 35(2), 259–271 (1988)
17. Sartor, G., Prakken, H.: A system for defeasible argumentation, with defeasible priorities. In: Gabbay, D.M., Ohlbach, H.J. (eds.) FAPR 1996. LNCS, vol. 1085, pp. 510–524. Springer, Heidelberg (1996)
18. Walton, D.N.: Argumentation Methods for Artificial Intelligence in Law. Springer (2005)
19. Wright, R.W.: Causation in tort law. California Law Review 73, 1735–1828 (1985)
20. Valente, A.: Legal Knowledge Engineering A Modelling Approach. IOS Press, Amsterdam (1995)
21. Zurek, T., Kruk, E.: Supporting of Legal Reasoning for Cases Which Are not Strictly Regulated by Law. In: ICAIL 2009: Proceedings of the 12th International Conference on Artificial Intelligence and Law, pp. 220–221. ACM, New York (2009)
22. Zurek, T.: Instrumental Inference in Legal Expert System. In: Atkinson, K. (ed.) Legal Knowledge and Information Systems, pp. 155–159. IOS Press (2011)
23. Zurek, T.: Modelling of afortiori reasoning. Expert Systems with Applications 39(12), 10772–10779 (2012)
24. Zurek, T.: Conflicts in legal knowledge base. Foundations of Computing and Decision Sciences 37(2), 129–145

Crawling Data-Intensive Web Sources Using Structure Information

Dawid Grzegorz Węckowski

Department of Information Systems, Poznań University of Economics,
al. Niepodległości 10, 61-875 Poznań
dawid.weckowski@kie.ue.poznan.pl

Abstract. As more and more information resources are being published on the Web, companies, in order to fulfil their information needs, run a monitoring process of data-intensive Web sources. Although current Web monitoring solutions (Web crawlers) address many challenges of the information resources retrieval, they seem to miss an important business-related factor – the utility of the resulting resources' collection. This article presents a concept of utility-aware Web crawler that can use Web site structure information to build a collection of information resources of high utility for a company and maintain the utility level when the resources change.

Keywords: Web crawling, information resource utility, data-intensive Web sources, Web monitoring.

1 Introduction and Motivation

Every day, more and more information resources are being published on the Internet. At the same time, the availability of the most recent and relevant information becomes crucial for different areas of economic activity. As the Web information resources change fast, especially in data-intensive Web sources, using the standard search engines in the process of information acquisition from the Web can be insufficient in terms of the information resources freshness and relevance for economic entities, e.g. companies. To keep up with the dynamic nature of Web information resources, companies can employ a custom Web crawler for Web document retrieval. Although, taking into consideration the ever increasing information overload phenomena, the crawler is supposed to build a collection of information resources, which is not only fresh, but in general has a high utility for economic entities. Only then the information resources collection can be effectively used in various business processes of a company.

Current solutions in this area mainly rely on publicly available crawling engines or some well-established, general-purpose crawling strategies for Web documents retrieval [1]. As the preliminary study shows, these approaches appear to be insufficient in terms of responding to a high variability of the data-intensive Web sources. Building a collection of high-utility information resources from such dynamic Web sources, requires defining a proper crawling strategy, based

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 196–207, 2013.

on some utility measures. The resources' collection utility can be related to its freshness, relevance or acquisition cost. Defining proper utility measures is prerequisite for building an utility-aware Web crawler that can efficiently harvest Web information sources.

In this article we propose a new approach for custom Web crawling, that can fulfil specific information needs of a company, contrary to the unfocused, general-purpose search services available on the Internet, which strive for wide coverage of the whole Web. The presented idea indicate the possibility of exploiting information about the structure of Web sites for improving crawling performance in terms of utility measures.

2 Related Work

Several approaches to the Web resources crawling can be found in literature. The first group of them are the general-purpose crawlers, designed for a wide-range, unbounded Web page retrieval, used mostly in online search engines. Another group are focused crawlers that are guided and restricted to retrieve pages of a predefined topic. Then there are application-specific crawlers that strive for harvesting some types of Web applications, e.g. Web forums, social networks, online store databases. The applications can consist of both statically addressed pages ("shallow Web") and form-based generated pages ("deep Web").

2.1 General-Purpose Crawling

In the most common approach a Web crawler uses a list of URL addresses, defining the documents that are to be retrieved. Initially, the list contains some predefined starting addresses. The crawler periodically executes a sequence of actions aimed at building, expanding and refreshing a local collection of retrieved resources, as well as broadening the list of URL addresses. Firstly, the crawler selects addresses from the list according to some specified criteria. Then the documents at the addresses are downloaded and analysed. If a document is retrieved for the first time, it is added to the local collection, otherwise the crawler checks whether the new version of the document differs from the one that is already in the local collection and updates it accordingly. From the retrieved documents the crawler extracts hyperlinks to other Web pages and adds them to the URL list, then the process goes back to the address selection phase.

General-purpose crawlers are designed for gathering as many resources as they can to build comprehensive local collections. They are used mainly for search engine purposes. An example of such crawlers can be Googlebot [2] used for the Google search engine, Nutch[1], developed by Apache Foundation, or Heritrix[2] build by Internet Archive library.

[1] http://nutch.apache.org/
[2] https://webarchive.jira.com/wiki/display/Heritrix/Heritrix

2.2 Focused Crawling

Focused crawlers are designed to selectively search for Web pages that correspond to a defined topic [3]. During the Web page retrieval, a focused crawler conducts content analysis to select only those hyperlinks that are most likely relevant for specific topic, and it ignores other ones, to determine the topic-based monitoring scope. The selection process is preformed according to the purpose of the crawler's activity — instead of preparing a collection to address any adhoc query, focused crawler builds a collection with regard to stable information needs [4]. Focused crawlers can serve better for filtering purposes than for generic search applications, and since they limit the number of monitoring Web pages, they can reduce the resources usage to some extent or increase the collection freshness.

One of the first implementations of a full-featured focused crawler was presented by Chakrabarti et al. [3], showing that it can gather relevant documents, identify a quality content in the crawl, and then use this information for self-guidance in terms of further selection. Rennie and McCallum [5] proposed a focused crawler that utilises a reinforcement learning method for improving subsequent fetches, based on previous results. The crawler designed by Diligenti et al. [6] with the same purpose uses context graphs for credit assignments to different Web pages on the navigation path.

A group of focused crawlers relays on PageRank-based [2] algorithms. Haveliwala [7] proposed a topic-sensitive PageRank algorithm, designed to improve the Web page ranking accuracy by computing PageRank vectors that are biased towards a set of predefined, representative topics. Navigational Rank by Feng et al. [8] tries to effectively determine a link capability of leading to a target pages not directly but further on the navigation path, by taking into account both relevant and irrelevant pages the link points to.

Some other issues of focused crawling were also discussed in literature. Pirkola and Talvensaari [9] described a method to deal with the low recall of the focused crawlers, caused by the nature of common crawling strategies and the locality of breadth-first search algorithms. Another issue, researched by Barbosa and Bangalore [10], is the improvement on novelty search in a predefined topic.

2.3 Web Application Crawling

There are several Web crawlers that are designed for document retrieval from specific types of Web applications, including blogs, Web forum sites or social network sites. A method for crawling the "blogosphere" was proposed by Kolari et al. [11]. While fetching Web pages it identifies blog sites for monitoring, and additionally reduces the disturbance caused by spam blogs.

As forum Web sites became more and more popular, some customised crawling solutions were developed to harvest the content of Web forums. Cai et al. [12] built an "intelligent forum crawler", called iRobot, which uses initial sampling to find traversal paths for visiting various types of pages within Web forum. Yang et al. [13] researched on incremental crawling of web forums with a use of list-wise strategy, taking into consideration the relations between Web pages in the same forum thread. FoCUS (Forum Crawler Under Supervision) developed by Jiang et al. [14] is aimed at visiting only relevant content in Web forum, keeping minimal overhead thanks to the exploration of, so called, implicit navigational paths.

Some examples of social network crawlers can be the solution by Catanese et al. [15], introducing methods of collecting and analysing extensive sets of Web pages form Facebook[3], as well as the TwitterEcho [16], an open source crawler, designed for continuous data collection from particular communities in the Twitter Web site.

2.4 Deep Web Crawling

Deep Web crawling is a process of retrieving Web pages generated with response to a query on an underlying database. Queries are usually formulated with the use of HTML form controls, and the process of selecting appropriate query terms is additional challenge in the crawling process.

One of the initial studies in the area of Deep Web crawling was conducted by Raghavan and Garcia-Molina [17]. They worked on design principles for Web crawler aimed at retrieving documents from the Deep Web. Later studies mainly concentrate on automatic methods for Deep Web crawling. Ntolaus et al. [18] studied effective crawler design for autonomous discovery and download of the content from the Deep Web. The issue of automatic formulation of meaningful queries to Deep Web sources was analysed by Wu et al. [19]. Barbosa and Freire [20,21] investigated another issue of crawling this type of pages, namely the automatic location of Deep Web sources entry points. Madhavan et al. [22] worked on solution for crawling wide range of Deep Web sites by, so called, "surfacing" process. An analysis of incremental Deep Web crawling was made by Chen et al. [23], who targeted at new documents retrieval from the Web databases and minimising the communication costs.

3 Research Problem

While current solutions address various aspects of Web document retrieval, they seem to miss an important indicator of the retrieval process success, namely the utility of the gathered documents' collection. Companies, or other economic entities, run the crawling process within different Web sources to gather information resources that will allow them to make better decisions, gain some economic advantage or to keep up with competitors. The retrieved document collection needs to meet the information needs of the companies and to be useful in different business processes.

Thus, the defined problem is how to retrieve a collection of Web information resources of high utility for economic entities and maintain the utility level while the resources are changing.

[3] `http://www.facebook.com/`

4 Proposed Approach

Improving and maintaining utility level of information resources' collection can be ensured by a Web crawler, which runs the document retrieval process. The proposed approach is to guide the crawler with an additional information concerning a Web source structure, as well as change patterns of information resources and economic entities information needs with regard to the source structure and time. This information, gathered in a form of profile, can be used to instruct the Web crawler how to run the monitoring of particular data-intensive Web source in order to collect information resources of high utility.

4.1 Data-Intensive Web Sources

Flejter [24] defines the data-intensive Web sources as follows:

> "Web sites [that] aim at publicising structured or semi-structured information in a structured or semi-structured way. (. . .) [They] are typically focused on a few classes of objects and [they] posses stable schema of records presented on-line."

This form of content publishing on the Web is popular among various sites, e.g. social networking sites (e.g. forums, blogs), news portals, e-commerce Web sites, Web applications or Deep Web sites. Different data-intensive Web sites can serve as a valuable source of information that can be utilised by companies in different business contexts.

The key features of data-intensive Web sources is a fast growing volume of new resources and a high variability of the resources. This mean that two challenges need to be addressed while building the local collection of information resources: the selection of relevant documents and refreshing them according to their variability, company information needs and cost constrains.

4.2 Web Source Structure Information

The information about the structure of a data-intensive Web source can support a Web crawler in several ways. First of all, a navigational model of the Web source, defining types of Web pages the crawler should retrieve, as well as the navigational patterns between different types of Web pages, can limit the monitoring scope, and indicate potentially relevant resources. Given the navigational structure model, Web crawler can visit only those pages, which are likely to contain useful information or to be otherwise useful in the crawling process. Guided by the model, crawler can ignore Web pages that are assumed irrelevant, e.g. pages outside the Web source or pages containing advertisements.

Furthermore, different areas in the navigational model or different Web page types can be assigned with specific change patterns. If the crawler strive for ensuring freshness of the local collection, then the areas of Web source with higher change frequency should be revisited more often. This information can be gathered during the consecutive crawls and can be supported by an expert information. By adjusting the crawling strategy to the Web pages change frequency with respect to the Web source structure, Web crawler can retrieve resources in a more timely manner.

Finally, the structure model can also serve for defining Web source areas of different preference by economic entities according to their information needs. Based on expert or collaborative assessment, companies can assign information concerning importance level to specific areas in the navigational model and the time relevance for each of them. Using that information, crawler can constantly improve the local collection of information resources, by adjusting crawls to the company's information needs.

Using the structure information in the crawling process influence the cost of information resources acquisition. With the navigational model of a Web source, the crawler does not waste time, computational power or bandwidth for retrieving unimportant pages. Moreover, proper estimation of resources' changes limits the number of fetches, since only the Web pages that are likely to be changed are revisited. Those improvements can be considered in two perspectives: the same information resources can be acquired at lower cost, or at the same cost the crawler can build better collection of information resources by improving coverage of a Web source – crawler can find more important Web pages since it disregard other ones.

Figure 1 shows a simplified example of aforementioned navigational model for a Web forum. Web crawler visits only the pages within the forum, i.e. the main page, category pages and pages with single posts. Pages are grouped by their content, showing areas of different change frequency patterns, e.g. category pages change with different frequency than post pages. The intensity of colour indicates the preferences of economic entities, e.g. new posts are more preferred than the old ones.

4.3 Incremental Crawling

Incremental Web crawling is an iterative process of retrieving documents available on the Web in order to build a local collection of the documents [1]. In this process new Web pages are being discovered and retrieved and then each page in the collection is revisited at custom frequency to update its content. The revisiting frequency can be adjusted at run-time based on the history of retrieval, as well as an additional information, like the Web source profile information.

While operating, an incremental Web crawler has to perform several tasks, including Web page fetching, link analysis, revisitation scheduling, content change detection, bandwidth load balancing, conforming politeness policy. At each step it should reach a compromise between the monitoring scope and the local collection freshness, deciding whether to visit a new Web page or revisit one of those already crawled. The revisiting strategy is highly dependent on the estimation of Web pages change frequency – the better this estimation is, the easier the crawler can schedule Web pages retrieval short after their change, assuring the local collection freshness [25].

Providing additional information in the form of structure-based profiles of data-intensive Web source can further improve the performance of an incremental Web crawler in terms of utility of the local collection of resources. The navigational model of a Web source allows for clear definition of retrieval scope, eliminating not useful resources, thus limiting the number of fetches required to maintain the local collection. In some cases, especially when the network bandwidth and the computing power are limited, the use of Web source navigational model increase the local collection coverage level – the Web crawler can retrieve more relevant documents from the source because it is not busy with crawling irrelevant resources.

Moreover, the freshness of the information resources can be improved by iterative process of adjusting revisitation frequency to the frequency of resource changes, but it can also be supported by an expert information about change patterns in different areas of the Web source. Guided by the information, crawler can presume that some of the resources within the Web source will change at higher or lower frequency than it can be estimated based on the history of changes. This is important especially for resources with a high or fluctuating change frequency as the crawler can have only partial history of resources' changes since it cannot reliably detect all the changes. This is due to the fact that a resource change can be detected only upon the resource fetch, and multiple changes between consecutive fetches of the resource are perceived as one.

The most important factor in terms of the utility value of the gathered collection of information resources are the information needs of the company or other economic entity that runs the monitoring. Considering the Web source structure basis, the crawler can be instructed where are the most important resources in the context of the information needs, and how that needs change in time, e.g. if a company needs some information resources constantly or once in a time. While scheduling subsequent fetches and managing the queue of addresses to retrieve, the crawler can prioritise more important resources to remain fresh by most of the time and can schedule some resources to be fresh at some defined point in time when they will be needed.

Using the Web source profile information, Web crawler can improve different aspects of information resources retrieval process and maintain high utility level for the resources' collection.

5 Methodology

The first part of the presented work consists of a literature review on monitoring Web sources and the utility theory in the context of information resources. Over a hundred publications concerning information science and economics has been analysed. A significant part of them were articles from online publication databases Springer Link and ACM Digital Library.

The analysis resulted in defining a set of possible improvements for the current Web monitoring solutions, in order to increase the utility level of gathered information resources. It was discovered that there is a need for defining a model of Web source profiles as well as a method of monitoring data-intensive Web sources with the use of those profiles. Over 30 data-intensive Web sources were analysed to indicate the elements that need to be included in the profile's model

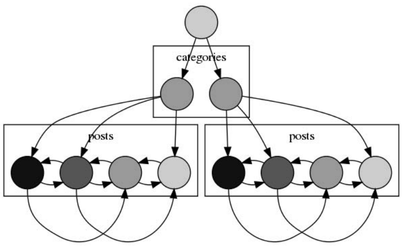


Fig. 1. An example of Web source graph model for a Web forum with categories and posts' list paging. Stronger intensity of the colour indicates higher relevance of a resource for an organisation's information needs

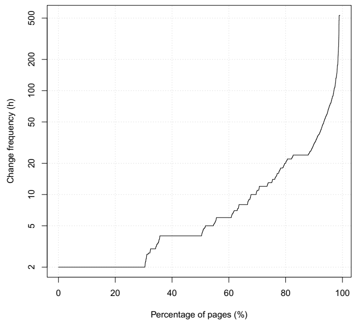


Fig. 2. Percent of pages with a given change frequency rate (log scale)

to enhance Web crawlers in working towards increase in resources' utility. Upon defining the final form of the model, a method of monitoring data-intensive Web sources will be designed and a prototype application will be built.

Empirical studies will be organised in two separate environments – a controlled environment and a real-life one. As the first one there will be a Web source built for the purpose of the study, with known parameters, including the size of the source, detailed structure, the time of publishing and modifying resources within the source. Thanks to this arrangement, the monitoring mechanism will be precisely assessed, as for every resource the times between publication/modification and retrieval will be known, as well as the coverage of the monitoring process. The research in a real-life environment, using one of the Web source from the Internet, requires a set of estimates for the monitoring mechanism assessment. In this case the resources' changes and the Web source size need to be estimated.

In both environments several monitoring mechanisms will be put to the test, including a prototype employing the proposed method and some state-of-the-art Web crawlers, designed for general-purpose, focused and application crawling. Their performance will be evaluated with the use of measures considered to be determinants of the gathered resources' utility. The research is designed based on the spiral model of Boehm [26], which describes subsequent research tasks concerning definition of objectives, needs and risks, review of current methods for reaching the detailed objectives, design of new solutions, running experiments, results analysis and applying progressive enhancements to the solutions.

6 Preliminary Experiment

In order to gain an insight in the nature of data-intensive Web information sources, a preliminary experiment has been run [27]. The aim was to analyse the change patterns of the information resources. Sixteen different Polish forum sites were chosen for the experiment, and they were crawled periodically every

2 hours over 23 days. In that period 266 crawls were managed to run, each time creating a snapshot of files from every Web site. The monitoring scope covered nearly 28,000 unique URL addresses that were discovered during the experiment, and that made for nearly 650,000 successful fetches of Web documents.

The experiment was based on a seminal paper by Cho and Garcia-Molina [28], and similarly it employed a crawling technique with a page window. The page window defines the policy of limiting the monitoring scope based on a volume of retrieved resources or the structure of a Web source. In the experiment, for every Web forum the crawler each time started from a predefined main page and continues to fetch hyperlinked pages according to the bread-first search algorithm. The crawl stopped when the fifth level within the Web site was reached. The retrieved set of pages for particular site was perceived as the page window, which could change during consecutive crawls as new pages and the ones moved to a higher level could be added to the page window, other pages could be removed from the window as they appeared deleted or moved to lower levels in the site structure.

As it was shown in the experiment results, the data-intensive Web sites have relatively high frequency of changes. Figure 2 presents the fractions of pages with different change frequencies in the experimental set. It can be seen that over 30% pages are worth visiting more often than every 2 hours, and over 50% of pages are worth visiting more often than every 4 hours. The lifespan and the turnover of Web resources were also measured. It appeared that only no more than 25% of pages live longer than one day, and during a single day around one-third of the monitored pages becomes unavailable and is replaced by new resources.

The high variability of Web information resources that was showed in the preliminary experiment, poses a strong challenge for the Web monitoring solutions. In case of limited bandwidth and computing power wide range monitoring of data-intensive Web sources that strives for reacting on every Web resource change is a difficult if not impossible task. That is why new approaches are needed to focus the crawling process on resources of high importance and to optimise the scheduling process so that the results would be valuable for the company. A crawler focused on building an information resources' collection of high utility can significantly improve the performance of monitoring data-intensive Web sources.

7 Future Work

The future work in the area of utility-aware Web information sources monitoring includes defining concrete structures and methods regarding the process of Web source profiling, construction of crawler prototype, running comparative experiments and results' analysis.

Employing the structure-based information that can support a Web crawler in building resources' collection of high utility, a profile model needs to be defined that will gather different pieces of the information. This includes information about:

– the Web source's navigational model in a form of a graph,
– the change frequency estimation modifiers,
– the information needs with regard to the Web source structure and the time.

For every element there is a need to determine an appropriate form of representation, as well as a set of methods for exploring the information and integrating it with a profile. Furthermore, the standard crawling process needs to be redesigned to be able to gain from the Web source profile information and deliver results of high utility value. Also an aggregate measure is to be defined in order to indicate the utility level of information resources' collection gathered by a Web crawler. It is important in terms of assessing the performance of the prototype crawler, its tuning and providing progressive enhancements. The utility measure will also serve as the basis for the prototype comparison with other monitoring solutions during the planned experiments.

8 Summary

This article presents a concept of utility-aware Web crawler that can build a collection of information resources featuring high utility value for companies or other economic entities. The crawler can monitor data-intensive Web information sources employing a Web source profiles with an additional, expert-provided information about the navigational model of a source, change frequency patterns in different parts of the source, as well as the importance of particular parts of the source for fulfilling the information needs of companies. As the preliminary experiment showed, data-intensive Web sites change frequently, which poses a challenge for general-purpose Web crawlers. The proposed approach based on utility measure may improve the performance of crawling data-intensive Web sources.

References

1. Levene, M.: An Introduction to Search Engines and Web Navigation. John Wiley & Sons (2010)
2. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. In: Proceedings of the Seventh International Conference on World Wide Web 7, WWW7, pp. 107–117. Elsevier Science Publishers B. V., Amsterdam (1998)
3. Chakrabarti, S., van den Berg, M., Dom, B.: Focused crawling: a new approach to topic-specific web resource discovery. Comput. Netw. 31(11-16), 1623–1640 (1999)
4. Abramowicz, W.: Filtrowanie informacji. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu (2008)
5. Rennie, J., McCallum, A.: Using reinforcement learning to spider the web efficiently. In: Proceedings of the Sixteenth International Conference on Machine Learning, ICML 1999, pp. 335–343. Morgan Kaufmann Publishers Inc., San Francisco (1999)
6. Diligenti, M., Coetzee, F., Lawrence, S., Giles, C.L., Gori, M.: Focused crawling using context graphs. In: Focused crawling using context graphs. In: Proceedings of the 26th International Conference on Very Large Data Bases, VLDB 2000, pp. 527–534. Morgan Kaufmann Publishers Inc., San Francisco (2000)

7. Haveliwala, T.H.: Topic-sensitive pagerank. In: Proceedings of the 11th International Conference on World Wide Web, WWW 2002, pp. 517–526. ACM, New York (2002)
8. Feng, S., Zhang, L., Xiong, Y., Yao, C.: Focused crawling using navigational rank. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM 2010, pp. 1513–1516. ACM, New York (2010)
9. Pirkola, A., Talvensaari, T.: Addressing the limited scope problem of focused crawling using a result merging approach. In: Proceedings of the 2010 ACM Symposium on Applied Computing, SAC 2010, pp. 1735–1740. ACM, New York (2010)
10. Barbosa, L., Bangalore, S.: Focusing on novelty: a crawling strategy to build diverse language models. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM 2011, pp. 755–764. ACM, New York (2011)
11. Kolari, P., Finin, T., Joshi, A.: SVMs for the blogosphere: Blog identification and splog detection. In: AAAI Spring Symposium on Computational Approaches to Analysing Weblogs (2006)
12. Cai, R., Yang, J.M., Lai, W., Wang, Y., Zhang, L.: iRobot: an intelligent crawler for web forums. In: Proceedings of the 17th International Conference on World Wide Web, WWW 2008, pp. 447–456. ACM, New York (2008)
13. Yang, J.M., Cai, R., Wang, C., Huang, H., Zhang, L., Ma, W.Y.: Incorporating site-level knowledge for incremental crawling of web forums: a list-wise strategy. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2009, pp. 1375–1384. ACM, New York (2009)
14. Jiang, J., Yu, N., Lin, C.Y.: Focus: learning to crawl web forums. In: Proceedings of the 21st International Conference Companion on World Wide Web, WWW 2012 Companion, pp. 33–42. ACM, New York (2012)
15. Catanese, S.A., De Meo, P., Ferrara, E., Fiumara, G., Provetti, A.: Crawling facebook for social network analysis purposes. In: Proceedings of the International Conference on Web Intelligence, Mining and Semantics, WIMS 2011, pp. 52:1–52:8. ACM, New York (2011)
16. Boanjak, M., Oliveira, E., Martins, J., Mendes Rodrigues, E., Sarmento, L.: Twitterecho: a distributed focused crawler to support open research with twitter data. In: Proceedings of the 21st International Conference Companion on World Wide Web, WWW 2012 Companion, pp. 1233–1240. ACM, New York (2012)
17. Raghavan, S., Garcia-Molina, H.: Crawling the hidden web. In: Proceedings of the 27th International Conference on Very Large Data Bases, VLDB 2001, pp. 129–138. Morgan Kaufmann Publishers Inc., San Francisco (2001)
18. Ntoulas, A., Zerfos, P., Cho, J.: Downloading textual hidden web content through keyword queries. In: Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital libraries, JCDL 2005, pp. 100–109. ACM, New York (2005)
19. Wu, P., Wen, J.R., Liu, H., Ma, W.Y.: Query selection techniques for efficient crawling of structured web sources. In: Proceedings of the 22nd International Conference on Data Engineering, ICDE 2006, p. 47. IEEE Computer Society, Washington, DC (2006)
20. Barbosa, L., Freire, J.: Searching for hidden-web databases. In: Doan, A., Neven, F., McCann, R., Bex, G.J. (eds.) WebDB, pp. 1–6 (2005)
21. Barbosa, L., Freire, J.: An adaptive crawler for locating hidden-web entry points. In: Proceedings of the 16th International Conference on World Wide Web, WWW 2007, pp. 441–450. ACM, New York (2007)
22. Madhavan, J., Ko, D., Kot, Ł., Ganapathy, V., Rasmussen, A., Halevy, A.: Google's deep web crawl. Proc. VLDB Endow. 1(2), 1241–1252 (2008)

23. Liu, W., Xiao, J.: Incremental structured web database crawling via history versions. In: Chen, L., Triantafillou, P., Suel, T. (eds.) WISE 2010. LNCS, vol. 6488, pp. 524–533. Springer, Heidelberg (2010)
24. Flejter, D.: Semi-Automatic Web Information Extraction. PhD thesis, Poznań University of Economics (2011)
25. Cho, J., Garcia-Molina, H.: Estimating frequency of change. ACM Trans. Internet Technol. 3(3), 256–290 (2003)
26. Boehm, B.: A spiral model of software development and enhancement. SIGSOFT Softw. Eng. Notes 11(4), 14–24 (1986)
27. Kaczmarek, T., Węckowski, D.G.: Web forums change analysis. In: Proceedings of the 9th International Conference on Web Information Systems and Technologies, Aachen, Germany (2013)
28. Cho, J., Garcia-Molina, H.: The evolution of the web and implications for an incremental crawler. In: Proceedings of the 26th International Conference on Very Large Data Bases, VLDB 2000, pp. 200–209. Morgan Kaufmann Publishers Inc., San Francisco (2000)

Aspect-Oriented Ontology Development

Ralph Schäfermeier*

Freie Universität Berlin, Corporate Semantic Web Group,
Königin-Luise-Str. 24/26, 14195 Berlin, Germany
{schaef,paschke}@inf.fu-berlin.de
http://www.corporate-semantic-web.de/

Abstract. In this paper, we describe our PhD research on an aspect-oriented approach for multi-dimensional specification of ontology modules driven by cross-cutting modularization requirements. The problem with existing modularization techniques lies in their respective focus on a single dimension along which a problem space can be segmented. We believe, however, that the majority of real-world ontology development scenarios is more complex, and that most applications have the potential for providing multiple cross-cutting modularization use cases. We examine commonalities between ontology modules and software aspects and propose an approach to applying the latter to the problem of a priori construction of modular ontologies and a posteriori ontology modularization.

Keywords: ontology modularization, aspect-oriented development, cross-cutting concerns.

1 Introduction

Ontologies are first order logic constructs enabling humans to share knowledge about a domain of interest in a formalized and machine readable way, thereby providing the metaphysical grounding of the semantic web. From an engineer's point of view, an ontology is an artifact resulting from an engineering process. Like software products, ontologies can grow arbitrarily large in terms of concept size, interrelatedness, and domain complexity. In order to facilitate development, selection, and re-use of ontologies, ontology modularization approaches aim to break up large monolithic ontologies into smaller, more tangible fragments.

Building modular ontologies from scratch is a non-trivial endeavor which requires foresightful planning and a thorough understanding of the problem domain. Even under these premises, it often fails because the imposed structure is not optimal [1] or not understood by users [2], or because the stipulated criteria for the structure cease fitting once the ontology evolves. Possible criteria for a priori and a posteriori modularization are manifold, and the choice is driven by the task the ontology is supposed to fulfill [3].

The problem with existing modularization techniques lies in their respective focus on a single dimension along which a problem space can be segmented.

* Supervisor: Adrian Paschke.

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 208–219, 2013.

We believe, however, that the majority of real-world ontology development scenarios is more complex, and that an application may contain multiple cross-cutting modularization use cases. For example, an application might require different (possibly overlapping) modules depending on specific functional requirements, such as subdomain extraction and, for example, temporal attribution. Maintainers of the ontology might specify additional non-functional requirements that comprise project specific customization/personalization with modules based on customer specific extensions. Application developers could additionally require a module of the entire ontology with tractable reasoning complexity (cf. [4,5,6]). See figure 1 for an illustration.

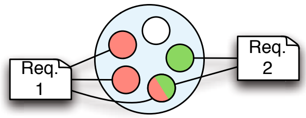


Fig. 1. Overlapping modules as a result of cross-cutting requirements. The blue circle is an ontology; the red and green small circles symbolize concepts which are selected according to requirement 1 and 2, respectively.

The problem of cross-cutting requirements for the segmentation of a problem space is not unique to the field of ontology modularization. In software engineering, the problem is known under the term cross-cutting concerns. As defined by the IEEE standard 1471 of software architecture [7], *"concerns are those interests which pertain to the systems development, its operation or any other aspects that are critical or otherwise important to one or more stakeholders"*.

In this doctoral research, we propose an aspect-oriented approach for multi-dimensional specification of ontology modules driven by cross-cutting modularization requirements.

Aspect-oriented software programming (AOP) languages provide syntactic means for the definition of join points, points in the execution flow of a program where code fragments that represent cross-cutting concerns, so called advices, can be interwoven with the main execution flow. While single join points can be defined in an extensional fashion, for example by means of code annotation tags, so called pointcuts allow for an intensional definition of join points by using quantification. More recent research combines the specification of aspects in software systems with formal requirements analysis, making aspects first-class citizens of software models. By transforming the execution semantics of software systems into a first order logic representation, we can apply the AOP paradigm to the DL based ontology model. Applying formal requirements analysis using formal competency questions, we intend to integrate the identification of aspects early in the development process.

2 Related Work

Ontology modularization is an active research field, and there exists a rich body of related work. D'Aquin et al. [3] distinguish between different perspectives on the problem of which two different subfields have emerged.

First, there exist approaches to *ontology partitioning*, where a monolithic ontology is decomposed into smaller fractions. The motivation for ontology partitioning comes from requirements concerning maintenance and reuse, thus constituting requirements rooted in an engineering point of view. The second class of approaches is referred to as *ontology module extraction*. The motivation for module extraction is mainly selective use and reuse [3].

2.1 Ontology Partitioning

In [8], the authors present a partitioning approach using so called $\mathcal{E}$-Connections [9]. The criterion for the partitioning process is semantic relatedness. This is determined by checking the $\mathcal{E}$-safe property a structural constraint which avoids the separation of semantically dependent axioms in order to achieve semantically consistent modules. The relatedness of the different modules is retained by the $\mathcal{E}$-Connections. One drawback of this approach, however, is that concept subsumption or the use of roles across different modules is not possible.

Schlicht et al. propose a semi-automated approach to ontology partitioning based on application-imposed requirements [10]. The method constructs a dependency graph of strongly interrelated ontology features, such as sub/super concept hierarchy, concepts using the same relations, or similarly labelled concepts. Then, the ontology is partitioned retaining strongly related groups in the same module. The method is parametrizable by specifying the features taken into account for constructing the dependency graph and the size a module should have in terms of the number of axioms.

Another class of partitioning approaches uses graph-based and social network metrics in order to determine central concepts and interrelated features which should be part of the same module [11].

2.2 Ontology Module Extraction

Approaches to ontology module extraction comprise logic-based extraction methods, for example [12], [13], [14] and [15]. These approaches are automatic and aim at producing self-contained, consistent ontology modules. They make use of logical properties such as semantic locality and inseparability.

A more unified approach to the problem of ontology module extraction and thereby similar to our work is the work of Doran et al. [16]. The authors propose using SPARQL queries in order to define ontology modules. They show that the more specialized approaches can be replicated in the form of SPARQL queries. The work of Doran et al. conforms with ours in what concerns the intensional module definition, in such a fashion that what we define as a pointcut is defined by Doran et al. in terms of SPARQL queries. It does not, however, include the possibility of extensional module definitions. In addition, SPARQL is an RDF graph based query language and neglects the semantics of DL-based ontology languages. Altogether, our vision behind this doctoral research is driven by the aim to enhance reusability of ontology modules. From our point of view, this does not only comprise the modules themselves, but also reproducibility of the modularization task. In order to achieve this, our approach makes the external module definition part of the model itself, by placing it at the same language level (as aspect-oriented languages do). In this way, ontologies can be shipped together with their aspect, i. e., module definitions. We argue that this approach improves traceability, comprehensibility, and thus reusability of ontologies and their modules.

A similar approach, using graph transformations and relying on user-defined graph-based extraction rules, has been proposed in [17].

2.3 A Priori Design of Modular Ontologies

While the latter two classes comprise approaches for *a posteriori* modularization of existing ontologies, a third arising class of methodological approaches aim at modular construction of ontologies in an *a priori* manner.

Related work in this area has been accomplished by [18], proposing a methodological framework for constructing modular ontologies driven by knowledge granularity. The proposed approach involves a separation into three levels: an upper ontology, modeling the theoretical framework, domain ontologies for reusable domain knowledge, and domain ontologies for application specific knowledge.

The shortcoming of existing modularization approaches is, as already mentioned in the introduction, their one-dimensionality, which is also acknowledged by [19] and [17]. The latter propose more unified approaches to the problem, however, they are restricted to the (graph-based) RDF model. Moreover, they lack formalisms of mapping modularizations to requirements, hindering relaying and re-use of module specifications.

3 Research Question and Hypothesis

The artifact produced in this research will be a unified method and a formalism for the extensional or intensional, a priori or a posteriori specification of aspects which allow for modularization of ontologies based on cross-cutting requirements. We will model the formalism using the common logic framework which, besides allowing for the definition of custom syntax, has the advantage of including sequence markers, since quantification over open sets of predicates is not allowed in first-order logic.

The research question addressed by this research is how the application of the aspect-oriented paradigm affects the quality of ontology modularizations. Our hypothesis is that aspect-oriented ontology development yields better ontology modules with respect to cross-cutting modularization requirements, such as dynamic access, understandability, maintenance, and re-use.

To evaluate our approach and test our hypothesis, we apply the approach to different modularization use-cases in the context of ontology development projects. Aspects considered in these use cases will, at least, comprise project affiliation, temporal attribution, workflow affiliation, re-use, and module understandability. We then use quality metrics in order to assess the quality of the modularizations gained using our approach and compare it with existing approaches.

4 Approach

Aspect-oriented software programming (AOP) languages provide syntactic means for the separation of cross-cutting concerns in software code into dedicated modules called *aspects*. Cross-cutting concerns emerge from requirements on different levels, e. g., functional and non-functional requirements, such as authentication and logging. They concern the entire or a significant part of the system and are thus scattered across the system, diminishing code locality and hindering system evolution and reusability [20]. The novelty in AOP lies in the fact that not only the functionality is encapsulated in the aspect (then called *advice*) but also the related control flow, using *join points*. Join points define points in the target module's execution flow where the related advice should be executed.

Two particular principles in AOP are quantification and obliviousness.

4.1 Quantification

AOP allows for an intensional definition of join points by using quantified statements in the form

$$\forall m(p_1, \ldots, p_n) \in M : s(m(p_1, \ldots, p_n)) \rightarrow (m(p_1, \ldots, p_n) \rightarrow a(p_1, \ldots, p_n)),$$

where M is the set of all methods defined within the software product, s a predicate specifying the join point properties, $m(p_1, \ldots, p_n) \in M$ all methods adhering to the signature $m(p_1, \ldots, p_n)$, and $a(p_1, \ldots, p_n)$ the execution of the advice with all the parameters of each method, respectively [21]. The set of all join points defined by s is called a *pointcut*.

Applying this notion of join point selection to the problem of ontology modularization yields the expression

$$\forall P(C, D) \in \mathcal{O} : s(P(C, D)) \rightarrow (P(C, D) \rightarrow a(P(C, D))),$$

where $\mathcal{O}$ is an ontology, $P(C, D)$ a binary predicate and $a(P(C, D))$ a function that applies the aspect to $P(C, D)$. It is obvious that this expression is not allowed in DL, since (1) it quantifies over predicates, and (2) it contains a function. We solve (1) by using Common Logic[1], using sequence markers in order to stay within FOL. As for (2) we introduce the type *Aspect* which is supposed to work

[1] `http://standards.iso.org/ittf/PubliclyAvailableStandards/c039175_ISO_IEC_24707_2007%28E%29.zip`

similar to an annotation type in that it merely marks a statement from the original ontology for inclusion in the pointcut. The pointcut definition expressed in CL is thus as follows:

```
(forall (and (f Axiom)
        (if (S f) (and (Aspect a)
                (exists ((t Type)) (Attr a t)
                (Attr f a))))))
```

According to [22], ontology modules can be specified both intensionally and extensionally, while aspects in AOP systems can only be described intensionally. In this regard, we beg to differ, referring to the common practice in AOP languages to provide means for defining pointcuts manually, for example, by using tags. In a similar manner, the above definition of an ontology pointcut allows for the extensional definition of join points by manually adding Aspect annotations. This is of particular use for a priori modular ontology development if used, e. g., by an ontology editor with switchable contexts, each context representing an aspect. A user could then extensionally define an aspect-oriented ontology module by switching aspects.

4.2 Obliviousness and Harmlessness

The fact that control flow is handed over from the main module to the aspects, making it unaware of the (quantified or extensionally specified) assertions made about it by external aspects that might possibly be applied to it, is termed *obliviousness* [20]. The practical consequence of obliviousness is that the developer of a module is not required to have knowledge about or spend additional effort in anticipation of a possible application of an aspect to her module.

Danters et al. allude to the problem that obliviousness is only guaranteed on a syntactic level while external aspects have in fact the potential to alter the semantics of the target module, making them potentially dangerous [23]. They propose the adaption of the *harmlessness* property for aspect-oriented systems, defining a *harmless aspect* as an aspect which, when applied to a target module, does not alter the semantics of the target.

While it is obvious that the obliviousness property naturally holds for ontology modules, the harmlessness property does not. Consider a simple example where the addition of a restriction, expressed in a module defined by an aspect of customer-based extensions, alters the semantics of a concept and can even lead to an inconsistent ontology (see figure 2).

Nevertheless, it is agreed upon in the recent literature that it is desirable that an ontology module is uninvasive, i. e, its addition to an ontology has no side effects. Whether uninvasiveness of a module applied by the means of an aspect is a required feature or not depends on the specific use case.

[24] and [25] propose the notion of conservative extensions which guarantee that a module added to an ontology does not alter the meaning of the original ontology. In the same vein, we define a harmless aspect of an ontology as a

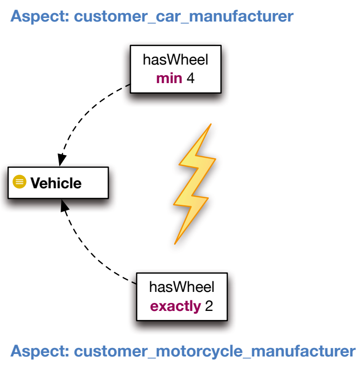


Fig. 2. Two harmful aspects, both changing the semantics of a concept, leading to an inconsistent ontology

module $\mathcal{M}$ that is the conservative extension of an ontology $\mathcal{O}$ with respect to the signature $\mathcal{S} = sig(\mathcal{O}) \bigcap sig(\mathcal{M})$ and the query language $\mathcal{QL} = \mathcal{M}$.

It has been noted that determining whether a module is a conservative extension of an ontology is a highly complex problem and even undecidable for expressive ontologies [24,25]. However, semantic locality is a sufficient condition for a conservative extension [12], and [26] suggests that the less complex syntactic locality constitutes a practically acceptable approximation.

4.3 Application of Aspects

Having established some important formal commonalities between aspects in software and ontology development, we propose a high-level architecture of an ontology repository which allows access to ontologies composed by different aspects which can be specified at access time. An application requesting an ontology can specify an ontology URI as well as multiple aspect names, also in the form of URIs. The repository stores the CL definitions of the aspects and, using this information, can compose the requested ontology with the appropriate activated aspects on the fly (see figure 3).

Furthermore, the repository should allow for storing ontologies along with extensional ontology specifications expressed in a logic language like, e. g., CL, as well as with intensional join point annotations.

In order to provide a common framework for the specification of different aspects, we propose an Aspect OWL ontology (see figure 4) which allows reasoning over aspects and, e. g., the application of multiple aspects subsumed under a class of aspects. The ontology has been developed according to the use cases we have collected by the time of this writing and intended to be extended as new use cases arise. The common base class *Aspect* subsumes all specialized aspects. It has two subclasses, *BuiltInAspect* which subsumes all possible built-in aspects of an ontology, such as ontology-language-specific ones. At the time of this writing, *BuiltInAspect* has exactly one subclass, *ReasoningComplexity*.

Individuals of this class represent different logics and their reasoning complexity, such as the different OWL 2 profiles. *ExternalAspect* subsumes all application-specific (functional or non-functional) aspects, such as provenance, trust, or customer specific features. The annotation property *hasAspect* may be used in order to attach an aspect instance to an ontological construct and thereby represents an extensional join point definition. The object property *associatedObject* relates the aspect instance to those ontological constructs which embody the concrete implementation of the aspect (advice).

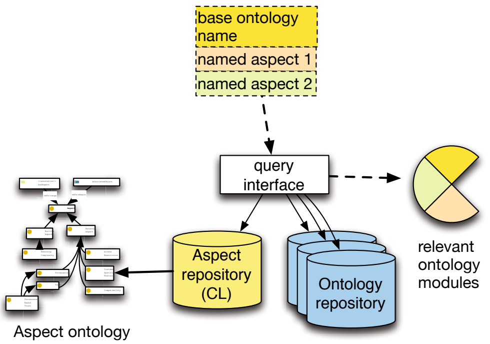


Fig. 3. An ontology repository with an aspect-aware query interface

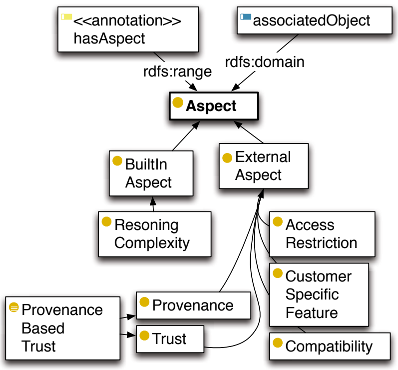


Fig. 4. The Aspect OWL ontology

As an example for the intensional definition of the built-in reasoning complexity aspect *tractability*, consider the following common logic fragment which defines a pointcut including all axioms that adhere to the OWL EL profile:

```
(forall (and (Relation r (r x y))
(if (or (r SubClassOf) (r EquivalentClasses) (r DisjointClasses) ...)
(and (TractabilityAspect a) (exists ((t Type)) (Attr a t) (Attr r a))))))
```

5 Research Methodology

Our PhD research will be carried out according to the Design Science Research methodological framework [27], in particular by following the six steps of the *Design Science Research Methodology (DSRM)* as proposed by [28]:

1. *Problem identification and motivation.* During this first step, we establish the relevance of our research by conducting a literature review and by collecting use cases for ontology modularization following cross-cutting requirements from business professionals engaged in development of ontology based solutions, supplementing the use cases we already have collected. The motivation for this research directly follows from the motivation for ontology modularization in general and a lack of a unified approach in this field, as described

in the literature. The diversity of the use cases will help us identify commonalities and differences between different modularization requirements. Based on these insights, we will be able to generalize the problem statement of cross-cutting concerns in ontologies and formulate research questions.

2. *Definition of objectives of a solution.* Based on the literature review and use cases from step 1, we will derive requirements for a possible solution. The literature review will recapitulate the state of the art in the fields of ontology and software development methodology, ontology modularization, as well as modular and, in particular, aspect-oriented software development. Based on the analysis of commonalities between software and ontology aspects described in section 4, we were able to achieve a first idea of the formal characteristics of the artifact to be produced in the context of this research as well as a positive prediction concerning the feasibility of the proposed solution. Further investigation is required in order to refine the requirements. In particular, since the research problem addressed in this PhD research is twofold (a priori and a posteriori modularization), the artifact will comprise formal, methodological and possibly algorithmic components.
3. *Design and development of a solution.* Based on the outcome of step 2, we will construct an artifact in the form of an aspect-oriented development method as a solution to the problem of modeling cross-cutting concerns in ontology development. The methodology will be based on a formal model of ontology aspects. Section 4 outlines a first proposal of such a formal model. Our method will allow ontology developers to formalize requirements from different stakeholders and map them to ontological constructs at development time. The formalism will allow ontology modules to retain meta-information about their nature, purpose, and original requirements they were derived from. Furthermore, a formalism will be constructed which allows the formal specification of an ontology module requirement, using a logic language (e. g. Common Logic) to enable a posteriori module extraction from monolithic ontologies.
4. *Demonstration.* During this step, we will provide a prototypical implementation as a proof that the requirements from step 2 are met and the problem can be solved by the artifact constructed in step 3. The prototype will be applied to different use cases in which the development of ontologies with containing cross-cutting concerns or the extraction of a module covering a specific concern from a monolithic ontology is required.
5. *Evaluation.* We plan to evaluate the artifact from step 4 in different ways. First, due to the formal nature of parts of the solution, an analytical (artificial) evaluation of the formalisms involved in the solution is necessary. This will, for example, comprise theoretical arguments and mathematical proofs in order to align the formalisms with requirements of ontology modules, such as semantic correctness and completeness. Second, the methodological aspects will need experimental evaluation. Therefore, we will conduct a controlled experiment involving the construction of modularized ontologies by experts, given a set of requirements expressed by different stakeholders. Third, in order to evaluate the quality of the ontologies and ontology modules produced

by our approach, we will again use analytical (natural) evaluation methods and apply established ontology quality metrics to them. Since the aim of this work is to improve ontology reusability, metrics adapted from the software engineering domain will be applied (cf., e. g. [29]). Furthermore, we will apply functional (black box) testing to applications in which the ontologies are used in order to evaluate their performance in real life scenarios.

6. *Communication.* Dissemination of progress and outcome of this work will be conducted by publishing in peer-reviewed workshops, conferences, and journals during the different stages.

6 Conclusion and Future Work

In this paper, we presented our ongoing PhD research. We showed that ontology modularization and aspect-oriented programming share interesting commonalities and hypothesized that the aspect-oriented paradigm can be applied to a priori modular ontology development as well as a posteriori module extraction. The next step will consist in providing a proof-of-concept system that dynamically interweaves aspects defined in the above manner.

The research question addressed by this research is how the application of the aspect-oriented paradigm affects the quality of ontology modularizations. Our hypothesis is that aspect-oriented ontology development yields useful ontology modules with respect to to cross-cutting modularization requirements, such as dynamic access, understandability, maintenance, and re-use. We expect that the intensional specification of ontology modules with pointcuts adds dynamicity and flexibility to modular development, making it easier to evolve modular ontologies in situations where evolution implies modularization requirement changes.

As an additional contribution of this research we expect a deeper understanding of the relation of software systems that inherently possess an execution semantics and knowledge representation models which naturally lack thereof. Furthermore, the current literature agrees upon the fact that is not yet well understood what defines a good modularization. We expect to gain a deeper understanding of ontology module quality and to contribute to finding a generalized definition.

Acknowledgements. This work has been partially supported by the "InnoProfile-Transfer Corporate Smart Content" project funded by the German Federal Ministry of Education and Research (BMBF) and the BMBF Innovation Initiative for the New German Länder - Entrepreneurial Regions.

References

1. del Vescovo, C., Parsia, B., Sattler, U., Schneider, T.: The modular structure of an ontology: an empirical study. In: Proceedings of the 2010 conference on Modular Ontologies: Proceedings of the Fourth International Workshop (WoMO 2010), pp. 11–24. IOS Press, Amsterdam (2010)

2. Grau, B.C., Parsia, B., Sirin, E., Kalyanpur, A.: Modularity and web ontologies. In: Doherty, P., Mylopoulos, J., Welty, C.A. (eds.) 10th International Conference on Principles of Knowledge Representation and Reasoning, pp. 198–209. AAAI Press (2006)
3. d'Aquin, M.: Modularizing Ontologies. In: Suárez-Figueroa, M.C., Gómez-Pérez, A., Motta, E., Gangemi, A. (eds.) Ontology Engineering in a Networked World, pp. 213–233. Springer, Heidelberg (2012)
4. Hoehndorf, R., Dumontier, M., Oellrich, A., Wimalaratne, S., Rebholz-Schuhmann, D., Schofield, P., Gkoutos, G.V.: A common layer of interoperability for biomedical ontologies based on OWL EL. Bioinformatics 27(7), 1001–1008 (2011)
5. Stvilia, B.: A model for ontology quality evaluation. First Monday 12(12) (2007)
6. Di Maio, P.: Toward "Just Enough" Ontology Engineering. Cutter Consoritium Business Intelligence Executive Report 9(3) (March 2009)
7. Group, I.A.W.: IEEE standard 1471-2000, Recommended Practice for Architectural Description of Software-Intensive Systems. IEEE (2000)
8. Grau, B.C., Parsia, B., Sirin, E., Kalyanpur, A.: Automatic Partitioning of OWL Ontologies Using ϵ-Connections (2005)
9. Cuenca Grau, B., Parsia, B., Sirin, E.: Combining OWL ontologies using ϵ-Connections. Web Semantics: Science, Services and Agents on the World Wide Web 4(1), 40–59 (2006)
10. Schlicht, A., Stuckenschmidt, H.: A Flexible Partitioning Tool for Large Ontologies. In: Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2008, vol. 01, pp. 482–488. IEEE Computer Society, Washington, DC (2008)
11. Coskun, G., Rothe, M., Teymourian, K., Paschke, A.: Applying community detection algorithms on ontologies for indentifying concept groups. In: Proceedings of the 5th International Workshop on Modular Ontologies, Ljubljana, Slovenia (September 2011)
12. Grau, B.C., Horrocks, I., Kazakov, Y., Sattler, U.: Extracting Modules from Ontologies: A Logic-Based Approach. In: [30], pp. 159–186, doi: 10.1007/978-3-642-01907-4
13. Konev, B., Lutz, C., Walther, D., Wolter, F.: Semantic Modularity and Module Extraction in Description Logics. In: Proceedings of the 2008 Conference on ECAI 2008: 18th European Conference on Artificial Intelligence, pp. 55–59. IOS Press, Amsterdam (2008)
14. Suntisrivaraporn, B.: Module Extraction and Incremental Classification: A Pragmatic Approach for $\mathcal{EL}^+$ Ontologies. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds.) ESWC 2008. LNCS, vol. 5021, pp. 230–244. Springer, Heidelberg (2008)
15. Kontchakov, R., Wolter, F., Zakharyaschev, M.: Logic-based ontology comparison and module extraction, with an application to DL-Lite. Artificial Intelligence 174(15), 1093–1141 (2010)
16. Grau, B.C., Honavar, V., Schlicht, A., Wolter, F. (eds.): Proceedings of the 2nd International Workshop on Modular Ontologies, WoMO 2007, Whistler, Canada, October 28, 2007. CEUR Workshop Proceedings, vol. 315. CEUR-WS.org (2008)
17. d'Aquin, M., Doran, P., Motta, E., Tamma, V.A.M.: Towards a parametric ontology modularization framework based on graph transformation. In: [16]
18. Thakker, D., Dimitrova, V., Lau, L., Denaux, R., Karanasios, S., Yang-Turner, F.: A priori ontology modularisation in ill-defined domains. In: Proceedings of the 7th International Conference on Semantic Systems. I-Semantics 2011, pp. 167–170. ACM, New York (2011)

19. Doran, P., Palmisano, I., Tamma, V.A.M.: Somet: Algorithm and tool for sparql based ontology module extraction. In: Sattler, U., Tamilin, A. (eds.) WoMO. CEUR Workshop Proceedings, vol. 348. CEUR-WS.org (2008)
20. Filman, R., Friedman, D.: Aspect-Oriented Programming Is Quantification and Obliviousness. In: Workshop on Advanced Separation of Concerns, OOPSLA (2000)
21. Steimann, F.: Domain Models Are Aspect Free. In: Briand, L.C., Williams, C. (eds.) MoDELS 2005. LNCS, vol. 3713, pp. 171–185. Springer, Heidelberg (2005)
22. Priss, U.: Facet-like Structures in Computer Science. Axiomathes 18(2), 243–255 (2008)
23. Dantas, D.S., Walker, D.: Harmless Advice. In: Conference Record of the 33rd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2006, pp. 383–396. ACM, New York (2006)
24. Grau, B.C., Horrocks, I., Kazakov, Y., Sattler, U.: Modular Reuse of Ontologies: Theory and Practice. Journal of Artificial Intelligence Research 31, 273–318, ACM ID: 1622664 (2008)
25. Konev, B., Lutz, C., Walther, D., Wolter, F.: Formal Properties of Modularisation. In: [30], pp. 25–66, doi: 10.1007/978-3-642-01907-4
26. Del Vescovo, C., Klinov, P., Parsia, B., Sattler, U., Schneider, T., Tsarkov, D.: Syntactic vs. Semantic Locality: How Good Is a Cheap Approximation? In: Schneider, T., Walther, D. (eds.) Workshop on Modular Ontologies (WoMO), pp. 40–50 (2012)
27. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28(1), 75–105 (2004)
28. Peffers, K., Tuunanen, T., Rothenberger, M., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24(3), 45–77 (2007)
29. d'Aquin, M., Schlicht, A., Stuckenschmidt, H., Sabou, M.: Criteria and Evaluation for Ontology Modularization Techniques. In: Stuckenschmidt, H., Parent, C., Spaccapietra, S. (eds.) Modular Ontologies. LNCS, vol. 5445, pp. 67–89. Springer, Heidelberg (2009)
30. Stuckenschmidt, H., Parent, C., Spaccapietra, S. (eds.): Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization. LNCS. Springer, Heidelberg (2009), doi:10.1007/978-3-642-01907-4

RDF Data Clustering

Silvia Giannini

SisInfLab & DEI, Politecnico di Bari, Via Orabona 4, 70125 Bari, Italy
s.giannini@deemail.poliba.it

Abstract. The Web is evolving from a Web of Documents to a Web of Data. Meanwhile, the development of Semantic Web applications opens the way for addressing complex information needs. In this scenario, clustering is identified as a crucial task for semantic mashups. After a thorough review of RDF clustering techniques, the paper addresses the open issues within the efficient exploitation of the knowledge contained in RDF data sources. Then, first promising attempts in exploring the applicability of community detection algorithms for RDF clustering are reported.

Keywords: RDF clustering, Linked Open Data, Semantic Web, data integration.

1 Introduction

The Web, seen as a global information space, is evolving from a Web of Documents to a *Web of Data*. This effort is actively pursued by the *Linking Open Data*[1] (LOD) project, that outlines best practices for publishing, sharing, and connecting structured data on the Web. These are summarized by four basic guidelines [1]: (i) use of global *Uniform Resource Identifiers* (URIs) for naming things (*e.g.,* documents, objects or even abstract concepts); (ii) use of a standardized access mechanism (HTTP URIs) to look up unique URIs and enable dereferencing of associated information and descriptions on the Web; (iii) adoption of the *Resource Description Framework*[2] (RDF) as the machine-readable and open standardized data format to represent semantics and informative content of resources; (iv) connection of data URIs among different data sources through the adoption of RDF links, in order to enrich the knowledge about a resource. RDF links represent the glue of the Web of Data, capturing a fair amount of knowledge and creating a complex network of interrelated resources [2]. An RDF link simply states that one piece of data has some kind of connection (*e.g.,* a relationship, an identity or a vocabulary link) to another piece of data. RDF links enable retrieval mechanisms, that is, the *navigation* among data sources exploited by Linked Data browsers, and *data crawling*, useful for searching the Linked Data space by using queries more expressive than the simple keyword

[1] `http://linkeddata.org`
[2] `http://www.w3.org/standards/techs/rdf#w3c_all`

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 220–231, 2013.

search [3]. The most recent statistic on the LOD Cloud (September 2011) reports that over 31 billions facts are currently stored as RDF triples in 295 data sources, with more than 500 millions links among them[3].

Against this critical mass of RDF data, mechanisms for efficiently storing, indexing and querying data have to be put in place [4,5]. Moreover, in order to build knowledge-intensive applications with LOD, machine learning techniques for mining the Web of Data need to be developed. Aimed at answering these compelling requirements, *RDF clustering* may be adopted for:

– the introduction of hierarchical path-indices in order to speed up the RDF query process [6,7];
– an efficient partitioning and storage of RDF data on different machines, reducing the exchange of data among nodes when processing query patterns spanning multiple partitions [5];
– the discovery of homogeneous groups of RDF resources, revealing discriminating features for each cluster of a data-set at hand, and ontology alignment [8,9,10].

A *cluster*, or *module*, can be defined as a set of homogeneous resources with large intra-cluster similarity and large inter-cluster dissimilarity. Therefore, the concept of *clustering* is related to the idea of grouping together resources *similar* to some extent [11].

The proposed research finally aims at the development of (i) a clustering technique suitable for RDF data, and (ii) a reasoning service over RDF data, able to identify groups of closely related resources in form of set of triples, and extract a summary RDF description of them, useful for further agent-oriented applications. This paper addresses the first part of research and investigates on possible inductive RDF clustering techniques. In particular, the brief state of art on RDF clustering methods exposed in Sec. 5 suggests focusing the research effort on a graph-mining technique that exploits the native graph model of RDF(S) data, and explores the features of the network that it originates. After introducing the RDF data-model (Sec. 2), the paper delineates the main concepts of graph-mining with community detection algorithms, and presents a preliminary attempt in exploring its application to RDF data clustering (Sec. 3). Promising results of the approach are illustrated in Section 4, where a community detection algorithm is applied on a synthetic RDF data-set. Conclusions are drawn in Section 6, with the definition of ongoing and future works.

2 RDF(S) Data Model

As an official standard of the World Wide Web Consortium (W3C), Resource Description Framework represents the universal graph-based data model to flexibly publish structured data on the Web of Linked Data, and, ultimately, into the Semantic Web [12,13,14]. It combines Web based protocols and formal semantics.

[3] `http://lod-cloud.net/state/`

In particular, *RDF-Schema* (RDFS) formalizes concepts to define a class-structure as part of a given RDF-graph, thus making the represented informative knowledge machine-understandable and processable. A single RDF(S) data source asserts facts about concrete or abstract entities of the real world. These are given as a set of (s, p, o)-triples, where (s, p, o) stands for $(subject, predicate, object)$. Triples are graphically represented by a directed edge-labeled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where s and o are nodes in $\mathcal{V}$, and $(s, o) \in \mathcal{E}$ is an edge, oriented from node s to node o and labeled with predicate URI p. Here, the formalization introduced by [4] is adopted, distinguishing between a *data-graph* and a *schema-graph* in the graph-based RDF(S) data model.

Definition 2.1 (RDF data-graph). *An RDF data-graph* $\mathcal{G}^d = (\mathcal{V}^d, \mathcal{E}^d, l^d)$ *is defined as a tuple where:*

- $\mathcal{V}^d = \mathcal{V}_E \cup \mathcal{V}_V$ *is a finite set of resource* ($\mathcal{V}_E$) *and literal* ($\mathcal{V}_V$) *vertices;*
- $l^d : \mathcal{V}^d \times \mathcal{V}^d \to \mathcal{L}^d$ *is an arc-labeling function returning, for a pair of nodes in* $\mathcal{V}^d$*, an element in* $\mathcal{L}^d = \mathcal{L}_R \cup \mathcal{L}_A \cup \{type\}$*, that is a finite set of relation* ($\mathcal{L}_R$) *and attribute* ($\mathcal{L}_A$) *labels;*
- $\mathcal{E}^d \subseteq \mathcal{V}^d \times \mathcal{V}^d$ *is a finite set of directed edges of type* $l(v_1, v_2)$*, where*
  (i) $l \in \mathcal{L}_R$, *if* $v_1 \in \mathcal{V}_E$, $v_2 \in \mathcal{V}_E$;
  (ii) $l \in \mathcal{L}_A$, *if* $v_1 \in \mathcal{V}_E$, $v_2 \in \mathcal{V}_V$;
  (iii) $l = \{type\}$, *if* $v_1 \in \mathcal{V}_E$, $v_2 \in \mathcal{V}_E$.

Please note that in a data-graph there are no distinctions between class and instance nodes in the set of resources modeled by $\mathcal{V}_E \subseteq \mathcal{V}^d$. On the other hand, the connection between data-graph and schema-graph is realized by labeled edges $type(v_1, v_2) \in \mathcal{E}^d$, where $v_2 \in \mathcal{V}_E$ coincides always with a vertex $v \in \mathcal{V}_C$ of the schema-graph $\mathcal{G}^s$ associated with $\mathcal{G}^d$, and defined below.

Definition 2.2 (RDF schema-graph). *An RDF schema-graph* $\mathcal{G}^s = (\mathcal{V}^s, \mathcal{E}^s, l^s)$ *is defined as a tuple where:*

- $\mathcal{V}^s = \mathcal{V}_C \cup \mathcal{V}_R \cup \mathcal{V}_A \cup \mathcal{V}_D$ *is a finite set of class* ($\mathcal{V}_C$)*, property* ($\mathcal{V}_R$ *and* $\mathcal{V}_A$) *and data-type* ($\mathcal{V}_D$) *vertices;*
- $l^s : \mathcal{V}^s \times \mathcal{V}^s \to \mathcal{L}^s$ *is an arc-labeling function returning, for a pair of nodes in* $\mathcal{V}^s$*, an element in* $\mathcal{L}^s = \{subClassOf, domain, range\}$*, that is a pre-defined set of labels;*
- $\mathcal{E}^s \subseteq \mathcal{V}^s \times \mathcal{V}^s$ *is a finite set of directed edges of type* $l(v_1, v_2)$*, where*
  (i) $l = \{subClassOf\}$, *if* $v_1 \in \mathcal{V}_C$, $v_2 \in \mathcal{V}_C$;
  (ii) $l = \{domain\}$, *if* $v_1 \in \mathcal{V}_R \cup \mathcal{V}_A$, $v_2 \in \mathcal{V}_C$;
  (iii) $l = \{range\}$, *if* $v_1 \in \mathcal{V}_R$, $v_2 \in \mathcal{V}_C$ *or* $v_1 \in \mathcal{V}_A$, $v_2 \in \mathcal{V}_D \cup \mathcal{V}_C$.

In the definition of the set $\mathcal{V}^s$, the property vertices are defined as the union of two subsets, $\mathcal{V}_R$ and $\mathcal{V}_A$, thus reflecting the distinction, at data-graph level, between relation labels and attribute labels.

In order to match the nature of the Web of Linked Data, it is assumed in the sequel that a schema-graph might be incomplete or might not exist for a given data-graph. However, the unified graph model adopted by RDF(S) makes it possible to apply the following clustering method to data sources providing also a schema definition.

3 Community-Detection Algorithms for RDF Clustering

The aim of community detection in graphs is to identify groups of vertices, which share common properties and/or play similar roles within the graph, and possibly their hierarchical organization, by only using the information encoded in the graph topology [11]. A widely accepted concept of *community* defines it as a subgraph of a network whose *nodes* are more *tightly connected* with each other than with nodes outside the subgraph [15]. Therefore, the similarity of nodes contained in a community is expressed in terms of cohesion degree of this subset of vertices. The idea of adopting community detection algorithms for RDF clustering descends from the evidence that, if two sets of entities are strongly related, they exhibit more connections than other sets of entities [16]. In detail, given an RDF data-graph $\mathcal{G}^d = (\mathcal{V}^d, \mathcal{E}^d, l^d)$, a classical community detection problem consists in finding a *partition* $\mathcal{C} = \{C_1, \ldots, C_n\}$ *of the set of nodes* optimizing a node similarity criterion, where it results $C_i \cap C_j = \emptyset$, $\forall i, j \in \{1, \ldots, n\}$, $i \neq j$. It means that traditional community detection algorithms work assuming absence of node-overlapping among clusters.

However, as in social networks an individual might simultaneously belong to several communities (family, work, sport, ...), an RDF resource might belong to different dimensions of analysis of the same informative domain described by the data-set. Therefore, the exploitation of community detection algorithms revealing overlapping communities (*i.e.,* structures in which a node might belong to more than one community) seems to be useful for RDF graph mining. Due to the possibility of each node to belong to different groups, highly overlapping communities can exhibit more external than internal connections. It follows that the distance metric adopted by traditional community detection algorithms is no more valid. In labeled graphs (like RDF graphs), each link models only one specific relation. Therefore, assuming that each link might belong to only one cluster, a traditional community detection algorithm can be applied this time to the set of edges, in order to discover groups of *tightly connected links* [17]. The problem of graph clustering with overlapping communities can be now formulated in terms of links: finding a *partition* $\mathcal{P} = \{P_1, \ldots, P_m\}$ *of the set of links* in $\mathcal{G}^d$ optimizing a link similarity criterion, such that $P_i \cap P_j = \emptyset$, $\forall i, j \in \{1, \ldots, m\}$. Starting from partition $\mathcal{P}$, a set $\mathcal{C}$ of m node clusters is obtained, putting in each cluster C_i, $i \in \{1, \ldots m\}$, nodes incident to each link contained in partition P_i. It means that $|\mathcal{V}^d| \leq \sum_{i=1}^{m} |C_i|$, *i.e.,* for some i and j in $\{1, \ldots, m\}$, it might result $C_i \cap C_j \neq \emptyset$.

Many networks also possess a hierarchical organization of modules. In the case of RDF graphs, it refers to the possibility of extracting concept descriptions of the information encoded in the RDF graph at different level of granularity, thus enabling approximate matching between a user request and the query results. A similarity measure established between links allows to build a dendrogram (*i.e.,* a tree diagram of the hierarchical organization of clusters) in which each leaf is a link from the original network, and branches represent link communities. In this dendrogram, links occupy unique positions whereas the hierarchy of nodes clusters, possibly overlapping, is naturally obtained with the same procedure described before for the conversion of a links partition $\mathcal{P}$ into a nodes partition $\mathcal{C}$.

In the following, they are illustrated the basic concepts of the graph clustering algorithm proposed by [18]. It has been selected among those available in literature for the capability of reconciling the principles of overlapping communities and hierarchy as aspects of the same phenomenon. A wide simulation campaign shows how the links clustering method developed by the authors overcomes every other tested method over several networks, varying in size and topology. It is chosen due to its simplicity and efficiency on large-scale networks, revealing the effective clusters structure described by external meta-data. Moreover, it is completely automatic and parameter free. The definition of similarity between links adopted in [18] considers only the local network topology as available information, *i.e.,* the neighborhood of a node is employed in the similarity evaluation. In particular, it is computed only the similarity between pairs of links that share a node k, since it is unlikely that disjoint links are more similar than adjacent links. The similarity $\mathcal{S}(e_{ik}, e_{jk})$ between links e_{ik} and e_{jk} sharing node k can be evaluated with the Jaccard index

$$\mathcal{S}(e_{ik}, e_{jk}) = \frac{|\mathcal{N}^{[i]} \cap \mathcal{N}^{[j]}|}{|\mathcal{N}^{[i]} \cup \mathcal{N}^{[j]}|}, \tag{1}$$

where $\mathcal{N}^{[i]}$ represents the set constituted by node i and its neighbors. A single-linkage agglomerative hierarchical clustering is then applied to derive the hierarchy of link communities. Each link is initially assigned to a different cluster; then, the pair of links with the largest similarity are merged, until all links are members of the same community (ties are agglomerated simultaneously).

Although meaningful structures exist at each level of the dendrogram, to extract the best community structure representing the data it is necessary to cut the dendrogram at a certain level. It is chosen according to the *partition density*

$$D_{\mathcal{P}} = \frac{2}{M} \sum_{c} m_c \frac{m_c - (n_c - 1)}{(n_c - 2)(n_c - 1)}, \tag{2}$$

where, given a link partition $\mathcal{P} = \{P_c\}_{c=1\ldots m}$, M is the total number of links in the network, m_c is the number of links of the c-th cluster of $\mathcal{P}$, and n_c is the number of nodes induced by links in P_c. The partition density is an objective function measuring the quality of a link partition in terms of links density inside communities, and it is evaluated at each level of the link dendrogram. It is then cut at the level identified by the maximum value of $D_{\mathcal{P}}$. Note that, the most time consuming phase is the computation of links similarity for each pair of link sharing a node. The worst case is represented by a complete directed graph, in which the number M of links is equal to $N(N-1)$, where N is the number of nodes of the network.

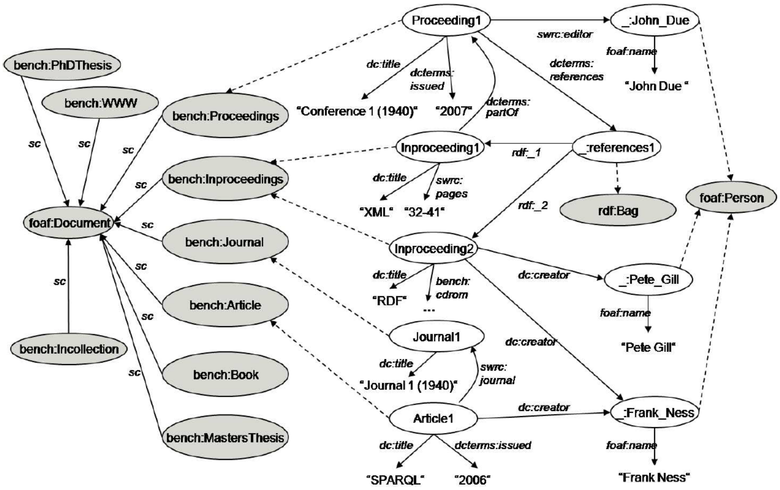


Fig. 1. A sample RDF graph generated by the benchmark [19]. On the logical level, the schema-graph (gray) and the data-graph (white) are distinguished. Dashed lines represent edges labeled with `rdf:type`; `sc` stands for `rdfs:subClassOf`. URIs and blank nodes are represented by ellipses, identifying blank nodes with the prefix _:; literals are represented by quoted strings.

4 Preliminary Results

In this section, the algorithm proposed by [18] is adopted to find a set $\mathcal{C} = \{C_1, \ldots, C_m\}$ of node-clusters, possible overlapping, over RDF(S) graphs. In order to evaluate the algorithm behavior over RDF(S) graphs, it has been primarily tested on a small synthetic data-set generated with the SP^2Bench benchmark [19]. The benchmark provides a data generator for the creation of arbitrarily large DBLP-like RDF documents, mirroring the key characteristics and social-world distribution of the real DBLP[4] data-set. It allows to obtain a controlled data-set, useful for a first manual inspection of resulting clusters and the definition of algorithm refinements for the adoption over real large-scale RDF graphs. Starting with a triple count limit fixed to 50, the generator ends up in a consistent state resulting in a total number of 266 triples, including data and schema information (see fig. 1). In detail, the description of 20 Articles and one Journal issue is created, together with schema definition. The execution of algorithm [18] on the data-set results in 53 clusters, one of which corresponds to a separated connected component of the RDF graph, originated by property vertices of $\mathcal{G}^s$ describing the semantics of the link labels $\mathcal{L}_R$ and $\mathcal{L}_A$ in $\mathcal{L}^d$. The remaining clusters, obtained cutting the links dendrogram at a level defined by the partition density optimization, can be classified in three categories:

---
[4] `http://dblp.uni-trier.de/`

1. Clusters of triples with the same subject s;
2. Clusters of triples with the same property-object pattern (p, o);
3. Clusters of triples with the same set $\{(p_1, o_1), \ldots, (p_f, o_f)\}$ of property-object patterns.

The first category of clusters aggregates nodes describing a resource s. Therefore, it equals the result of an instance extraction method, and it is identified through the signature s. Clusters of this type are represented by subgraphs of $\mathcal{G}$ containing the subject vertex, its outgoing links corresponding to properties listed in the cluster triples, and all the object vertices reached through them. Note that, due to the impossibility of links replication, not all subject outgoing links, and related object nodes, might be included in the cluster. If one or more outgoing links are included in another cluster, because of a higher similarity with links included in that cluster resulting from the links dendrogram cut based on partition density, they will not appear here (this issue can be resolved in a post-processing phase).

The second category of clusters aggregates different subject resources sharing the same property p with reference to an object resource o. These clusters are modeled by subgraphs of $\mathcal{G}$ containing the object vertex, its ingoing links labeled with the property p, and related subjects. Again, not all ingoing links of the object vertex o, labeled with p, might be included in the cluster, if they have been classified elsewhere.

Finally, the third category of clusters is the most interesting, since reveals composite aggregations of resources. Clusters of this type collect a set of t subject resources sharing the same set of f property-object patterns. Such kind of clusters are identified by subgraphs of the network composed by: the set of f object vertices characterizing the cluster; their f labeled ingoing links sharing the same subject node t, for each of the t subject vertices; the set of t incident subject vertices (here again, for each object vertex, only the ingoing property links listed in the cluster triples are considered). This category of RDF clusters can be interpreted as a collection of t clusters of the first type (there are t subsets of triples with the same subject), or as a collection of f clusters of the second type (they can be extracted f subsets of triples with the same property-object pattern). Clusters falling in the last two categories are identified through a signature composed by one (for the second category) or more (for the third) $(property, object)$ patterns.

The first category of clusters does not represent aggregation of homogeneous resources. However, it is still useful for RDF graphs summarization purposes, and clusters of this type could be employed by tasks or applications involving RDF instances extraction or resources description (*e.g.,* data browsers). The other two categories of clusters suggest the possibility of creating storage mechanisms based on path-indices. Moreover, they could be useful for link prediction or entity classification tasks.

Since non-trivial communities possess more than two nodes (or, equivalently, more than one link), a post-processing phase of the resulting links partition is needed. Clusters constituted by only one (s, p, o)-triple are merged with existing ones, analysing signatures. The triple is included in a cluster of the first or of the second type, if, respectively, s or (p, o) correspond to a *subject* or a (*property*, *object*) available signature (if links replication is allowed, both inserts are equally valid). If neither of the two previous cases occurs, the triple is merged with the cluster (or clusters) containing its subject node s as object. This choice has been suggested by the behavior shown by the algorithm against the presence of blank nodes, previously discarded for the definition of clusters categories. Moreover, clusters of the same category and with the same signature are eventually merged in the post-processing phase.

Applying clusters classification over the generated data-set, it results in: (i) 23 clusters of type 1; (ii) 3 clusters of type 2; (iii) 2 clusters of type 3 (remaining clusters are trivial). Clusters of type 1 collect the description of each of the 20 Articles and one Journal generated (two clusters are merged by the post-processing phase). In the second category of clusters, the set of all entities of type `Person`, together with two partitions of the schema-graph (the subclasses of the resource `Document` and resources of type `Property`) are listed. At the end, the two clusters of type 3 aggregate, respectively, all resources of type `Article` having `Paul Erdoes` as co-author and published in the same `Journal` issue, and all remaining resources of type `Article` of the same `Journal` issue (*i.e.* the articles that have not been written by Paul Erdoes).

For comparison purposes, the standard community detection algorithm reported in [15] has been tested on the same data-set. It is an heuristic method for the identification of nodes communities without overlapping, based on modularity optimization. The statistics shows that there are 21 communities, two of which belonging once again to the schema-graph $\mathcal{G}^s$. Each of the 19 remaining clusters can be classified as the description of a specific `Article` in the data-set. However, there are two exceptions: (i) the description of the `Journal` issue is located in the same cluster of an Article description; (ii) two Articles written by `Paul Erdoes` are clustered together. These are treated as misclassification cases, since in the data-set: (i) there are other resources of type `Article` published in the same `Journal` issue; and, (ii) there are other resources of type `Article` with `Paul Erdoes` as co-author. Moreover, due to the lack of overlapping among communities, there are missing information with reference to the results obtained by the first algorithm (*e.g.,* typing of resources).

Further tests performed with algorithm [18], over synthetic data-sets generated by *SP*^{2}*Bench*, reveal the same set of cluster categories. In particular, over 720 and 5362 triples, the algorithm returns, respectively, 277 and 3437 clusters. On the other hand, a degradation in clustering quality can be noticed for the algorithm without communities overlapping [15], when the size of the data-set increases. In particular, on the data-set of 720 triples, the statistic shows that there are three prominent clusters: 25.19% resources are of type `Person`; 24.81% resources are of type `Article`; and 15.23% resources has creator `Paul Erdoes`. The rest of nodes is partitioned in 14 smaller clusters. The identified communities are not so meaningful, revealing only a basic knowledge of DBLP domain entities. Results suffer the presence of nodes with high in- or out-degree (like classes `Person` and `Article`). In fact, a distinguishing feature of the DBLP data source, replicated by the benchmark also in small generated data-set, is to be a scale-free network, with few highly connected nodes which participate in a very large number of relations. These hubs prevent the emergence of submodules in the clustering process when overlapping is forbidden, imposing a restriction on the degree distribution (nodes must have approximately the same number of links), which contrasts with the network scale-free nature [20].

5 RDF Clustering: State of Art

The considerable amount of RDF data-sets being made available on the Web, and linked to each other, poses the question of how to efficiently exploit the knowledge therein. As depicted in the Introduction, clustering seems to be a good mining function in this perspective. First RDF clustering techniques presented in literature rely on traditional *data-clustering* methods [21]. However, the RDF data-model is not intuitively suited for standard machine learning algorithms. As a matter of fact, traditional data-clustering is based on instances, and not on large interconnected graphs. Therefore, a first step in the application of these methods over RDF graphs requires the extraction of instances, where the term instance refers to a subgraph relevant for a resource description. Unfortunately, *instance extraction* can be a time-consuming data-dependent phase (for a review of instance extraction methods and possible solutions to the problem of instance-based clustering of RDF data see [22]). Then, a classical hierarchical agglomerative or partitional algorithm can be applied to build a tree of clusters. In this step, an important role is assumed by the distance metric, that allows for evaluating the similarity between each pair of instances someway extracted. A classification of different *distance measures* for RDF instances clustering is adopted in [22].

First, the traditional *feature-vector based* clustering is revised. Paths of the RDF subgraphs representing instances are mapped to features, and the value of a feature is represented by the set of nodes reachable through that path. The distance between two instances is then computed with a vectors similarity notion inspired to the Dice coefficient.

Graph-based distance measures exploit the topology of the RDF graph in similarity evaluation. Conceptual similarity and relational similarity, respectively computed through the evaluation of nodes and edges overlap between the two subgraphs representing the pair of instances under evaluation, can be combined for the extraction of a distance value [23]. The work by [24] proposes instead kernel functions for RDF(S) data, *i.e.* specific for directed labeled graphs, based on the concepts of intersection graph and intersection tree between the instance-subgraph, or the instance-tree, of the two examined resources.

On the other hand, when it is available an ontological background, expressed with a semantic language, an *ontologically-based* distance measure can be adopted. Maedche and Zacharias developed and applied an ontological similarity metric for RDF data conforming to a well-defined ontology [25]. It is expressed as the weighted combination of three dimensions: the taxonomy, the relation and the attribute similarity. It does not require the phase of instance extraction, but it turned out to be more expensive than other similarity measures. A different approach for clustering RDF statements exploiting the ontological knowledge was proposed by Delteil et al. in [26]. This relies on the construction of a subsumption hierarchy, systematically generating the most specific generalization of all possible set of resources, using both the intension and extension of newly formed concepts. Unfortunately, both ontologically-based methods cited here are not applicable for real Semantic Web data, often noisy, incomplete and inconsistent (a list of assumptions that have to be verified by the RDF data-set at hand for these methods application is reported again in [22]).

The most noticeable limitation of traditional data-clustering techniques over RDF graphs is represented by the not fully exploitation of the native data model in the process - *i.e.,* the graph model - that makes it necessary the instance extraction phase. More recently, it has been proposed the investigation of other suitable methods for RDF clustering; in particular, the techniques that fall into the category of graph-mining. The goal of *graph clustering* is to partition graph vertices into several densely connected components, based on various criteria (vertex connectivity, neighborhood similarity, etc.). Many graph clustering methods are solely based on the network structure (see [27] for a comparison). A recent work by Alzogbi and Lausen [28] treats the problem of identifying similar structures inside an RDF graph, in order to reduce its size while keeping its structure invariate as much as possible. It combines bisimulation with an agglomerative data-clustering step. However, their work aims at finding set of nodes of an RDF graph which cannot be distinguished by looking at the sequence of predicates of their paths. It means that the method is only focused on paths, and not on property-object patterns. In many real applications, both the graph topological structure and the vertices', or relations', attribute similarity have to be taken into account. In fact, an ideal graph clustering method should generate clusters which have a cohesive intra-cluster structure with homogeneous vertices properties. Although they do not work directly on RDF graphs, the authors of [29] propose a graph augmented version, that makes it possible to take into account also the attribute similarity in the graph partitioning task, making the approach interesting for this state-of-art analysis. The augmented graph partitioning is then realized through a k-medoids clustering method, using a pairwise distance measure based on vertices neighborhood similarity. However, the algorithm relies on the a priori definition of the number of clusters to extract.

Community detection methods, while also concerned with dividing a graph into well-connected partitions, are more exploratory in spirit, and aim at extracting number and size of partitions from the data [11]. Therefore, they seem to represent a good candidate in refurbishing or improving RDF clustering methods. To the best of my knowledge there is no analysis of community detection algorithms behavior against RDF-graph for RDF clustering purposes.

6 Conclusions

In this paper, an analysis of existing RDF clustering methods for data-mining and management has been conducted. Then, a first attempt in automatically clustering RDF graphs with community detection algorithms has been presented. The results of first experiments are encouraging, and suggest the adoption of a technique allowing for communities overlapping. In particular, it has been put in evidence how the adopted method is able to perform, at the same time, instances extraction from RDF graphs, and resources clustering, without requiring any tuning. The hierarchical nature of the algorithm enables also the description of resulting communities at different levels of generalization.

A better definition of the post-processing phase, enabling also edges replication, is under study. In future works, the effectiveness and computational efficiency over real LOD data-set at Web-scale will be evaluated. Moreover, a systematic comparison with state of the art RDF clustering techniques will be conducted, for a deeper analysis of the proposed technique advantages.

References

1. Berners-Lee, T.: Linked data (2006), `http://www.w3.org/designissues/linkeddata.html`
2. Hausenblas, M., Halb, W., Raimond, Y., Heath, T.: What is the size of the semantic web? In: Proceedings of I-Semantics, pp. 9–16 (2008)
3. Bizer, C., Heath, T., Idehen, K., Berners-Lee, T.: Linked data on the web (ldow2008). In: Proceedings of the 17th International Conference on World Wide Web, pp. 1265–1266. ACM (2008)
4. Tran, T., Wang, H., Haase, P.: Hermes: Data web search on a pay-as-you-go integration infrastructure. Web Semantics: Science, Services and Agents on the World Wide Web 7(3), 189–203 (2009)
5. Zeng, K., Yang, J., Wang, H., Shao, B., Wang, Z.: A distributed graph engine for web scale rdf data. In: Proceedings of the 39th International Conference on Very Large Data Bases, pp. 265–276. VLDB Endowment (2013)
6. Kaushik, R., Shenoy, P., Bohannon, P., Gudes, E.: Exploiting local similarity for indexing paths in graph-structured data. In: Proceedings of the 18th International Conference on Data Engineering, pp. 129–140. IEEE (2002)
7. Wu, A.Y., Garland, M., Han, J.: Mining scale-free networks using geodesic clustering. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 719–724. ACM (2004)
8. Konrath, M., Gottron, T., Staab, S., Scherp, A.: Schemexefficient construction of a data catalogue by stream-based indexing of linked data. Web Semantics: Science, Services and Agents on the World Wide Web 16, 52–58 (2012)
9. Böhm, C., Lorey, J., Naumann, F.: Creating void descriptions for web-scale data. Web Semantics: Science, Services and Agents on the World Wide Web 9(3), 339–345 (2011)
10. Algergawy, A., Massmann, S., Rahm, E.: A clustering-based approach for large-scale ontology matching. In: Eder, J., Bielikova, M., Tjoa, A.M. (eds.) ADBIS 2011. LNCS, vol. 6909, pp. 415–428. Springer, Heidelberg (2011)

11. Fortunato, S.: Community detection in graphs. Physics Reports 486(3), 75–174 (2010)
12. W3C: `http://www.w3.org/tr/r2rml/` (September 27, 2012)
13. W3C: `http://www.w3.org/tr/rdfa-lite/` (June 07, 2012)
14. Augenstein, I., Padó, S., Rudolph, S.: LODifier: Generating linked data from unstructured text. In: Simperl, E., Cimiano, P., Polleres, A., Corcho, O., Presutti, V. (eds.) ESWC 2012. LNCS, vol. 7295, pp. 210–224. Springer, Heidelberg (2012)
15. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008(10), P10008 (2008)
16. Tran, T., Cimiano, P., Rudolph, S., Studer, R.: Ontology-based interpretation of keywords for semantic search. In: Aberer, K., et al. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 523–536. Springer, Heidelberg (2007)
17. Evans, T., Lambiotte, R.: Line graphs, link partitions, and overlapping communities. Physical Review E 80(1), 016105 (2009)
18. Ahn, Y.Y., Bagrow, J.P., Lehmann, S.: Link communities reveal multiscale complexity in networks. Nature 466(7307), 761–764 (2010)
19. Schmidt, M., Hornung, T., Lausen, G., Pinkel, C.: Sp^{2}bench: a sparql performance benchmark. In: IEEE 25th International Conference on Data Engineering, ICDE 2009, pp. 222–233. IEEE (2009)
20. Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabási, A.L.: Hierarchical organization of modularity in metabolic networks. Science 297(5586), 1551–1555 (2002)
21. Fanizzi, N., d'Amato, C.: A hierarchical clustering method for semantic knowledge bases. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part III. LNCS (LNAI), vol. 4694, pp. 653–660. Springer, Heidelberg (2007)
22. Grimnes, G.A., Edwards, P., Preece, A.D.: Instance based clustering of semantic web resources. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds.) ESWC 2008. LNCS, vol. 5021, pp. 303–317. Springer, Heidelberg (2008)
23. Grimnes, G.A., Edwards, P., Preece, A.D.: Learning meta-descriptions of the FOAF network. In: McIlraith, S.A., Plexousakis, D., van Harmelen, F. (eds.) ISWC 2004. LNCS, vol. 3298, pp. 152–165. Springer, Heidelberg (2004)
24. Lösch, U., Bloehdorn, S., Rettinger, A.: Graph kernels for RDF data. In: Simperl, E., Cimiano, P., Polleres, A., Corcho, O., Presutti, V. (eds.) ESWC 2012. LNCS, vol. 7295, pp. 134–148. Springer, Heidelberg (2012)
25. Maedche, A., Zacharias, V.: Clustering ontology-based metadata in the semantic web. In: Elomaa, T., Mannila, H., Toivonen, H. (eds.) PKDD 2002. LNCS (LNAI), vol. 2431, pp. 348–360. Springer, Heidelberg (2002)
26. Delteil, A., Faron-Zucker, C., Dieng, R.: Learning ontologies from rdf annotations. In: Workshop on Ontology Learning (2001)
27. Rattigan, M.J., Maier, M., Jensen, D.: Graph clustering with network structure indices. In: Proceedings of the 24th International Conference on Machine Learning, pp. 783–790. ACM (2007)
28. Alzogbi, A., Lausen, G.: Similar structures inside rdf-graphs (2013)
29. Zhou, Y., Cheng, H., Yu, J.X.: Graph clustering based on structural/attribute similarities. Proceedings of the VLDB Endowment 2(1), 718–729 (2009)

Semantically Enhanced Analysis of Enterprise Environment for the Needs of Business Networks Identification

Elżbieta Bukowska

Department of Information Systems,
Faculty of Informatics and Electronic Economy,
Poznan University of Economics al. Niepodleglosci 10, 61-875 Poznan, Poland
e.bukowska@kie.ue.poznan.pl

Abstract. This paper presents an outline of a method allowing to identify business networks via performing enhanced analysis and processing of data gathered from web sources, especially semantically enhanced databases. Thus, the proposed method, developed within an on-going PhD thesis, will operate on semantic profiles of market participants created especially for the needs of analysis. The clustering performed on profiles' elements will allow to identify similarities and dependencies between organizations and in consequence, identify existing business networks. The obtained knowledge will allow to address information needs of organizations that manifest themselves in the electronic economy setting.

1 Motivation and Research Questions

One of the main research areas in economics is the market, along with its subjects (e.g., organizations), objects that are being exchanged (goods and services) and relations between them. Recently a ,,static" model of business environment (i.e., a model dividing it into a proximal and distant environment) is being replaced by a more ,,dynamic" approach that focuses on relations between subjects [De Wit and Mayer, 2007] (Figure 1). Within this approach the market is represented as a business network [Jarillo, 1993, Campbell and Wilson, 1996, Hkansson and Snehota, 2006, Ratajczak-Mrozek, 2010, Ford et al., 2011].

A business network is defined as a set of relations connecting two or more subjects, characterized by a continuous interaction (a long-term interaction supported by informal contacts), interdependence (of resources, objects and actions), and a lack of clear boundaries and structures [Ratajczak-Mrozek, 2010]. What is important, all organizations on the market are a part of some business networks but some do not realize it.

There are two main approaches to define business network resulting in two types of business networks:

– *business network without a network leader* [Hkansson and Snehota, 2006, Ford et al., 2011] – a business network is created through interactions between organizations but none of those organizations can design network's strategy or define the direction of its development;

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 232–243, 2013.

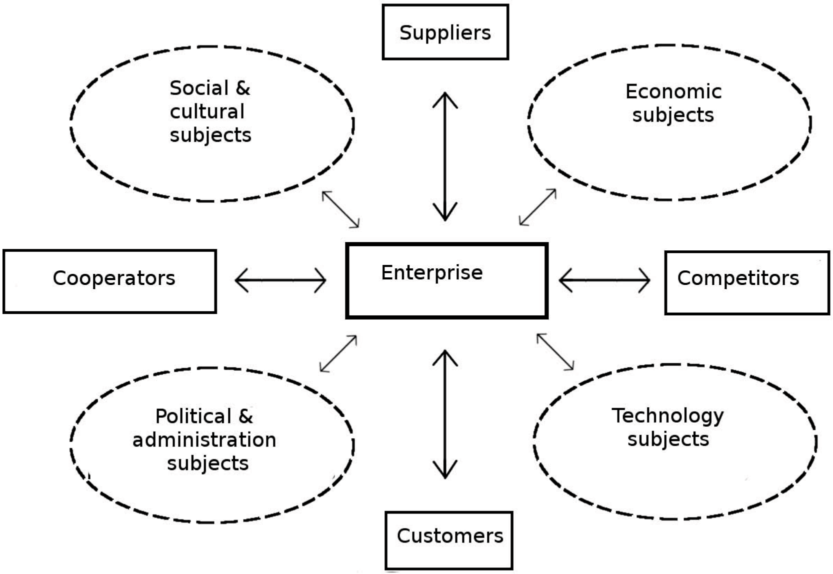


Fig. 1. Source: [De Wit and Mayer, 2007]

– *business network with a network leader* [Jarillo, 1993, Campbell and Wilson, 1996] – one organization acts as a network leader, i.e., not only it is aware of the network existence and structure, but also can influence the network, its strategy and development.

In order to be able to make valid decisions contributing to its success on the market, an organization needs information about its environment, i.e., its business network. However, there are several problems while acquiring information for the needs of a decision-making process. First, a decision maker usually requires information about a huge number of subjects, as well as objects related to them. Second, while formal relations are relatively easy to define, it is not clear how to find and name informal relations, which are sometimes even more important. Finally, the process of acquiring information about a business network from the Internet is usually complicated, especially if performed automatically. The last problem is caused by several aspects. The number of subjects that manifest themselves on the Web is enormous and getting bigger and bigger. There is too much data to process it manually (even with the search engines' support), finally there is a lack of tools to automatically monitor the Internet. Moreover, data about business networks changes frequently.

The motivation described above leads to the main research problem: *How to automatically discover relations between subjects that manifest themselves in the electronic economy setting, in order to identify business networks?* In order to solve this research problem, the following research questions need to be addressed:

– What kind of information sources are available in the Web? Are there any semantically enriched data sets? What is their structure?
– Can semantically enriched data and processing methods be used for identifying such relations?
– Which method of data analyses is the most suitable for this purpose?

The main goal of this paper is to provide an outline of the proposed method as well as share the initial findings from the carried out research. The paper is structured as follows. First the related work is dicussed. Section 3 describes an outline of the proposed method for business networks identification. Research methodology is presented in section 4. The last part describes obtained results and future work.

2 State of the Art

Investigating the above mentioned research question and designing the method to identify business networks requires application and consideration of concepts and artefacts developed within a few areas: data analysis and business intelligence, profiling, ontologies and semantic technologies. A short overview of the main aspects from this research fields follows.

Currently, the problem is not to obtain information, but to obtain the right information, at the right time and place. Thus, the need appears to deal with the information overload problem [Abramowicz, 2008]. As there is more and more information, we are not able to process it manually in a given time that we have. To facilitate this task Decision Support Systems (DSS) and Business Intelligence systems have been developed, but due to the lack of a coherent definition of company profile, DSSs are not able to provide full information on companies.

A profile might be defined as a representation in information system of a certain real or virtual object that is prepared in order to enable analysis and reasoning about a specific aspect of reality [Hildebrandt and Backhouse, 2005]. Profiles are widely used in search engines, information filtering, recommendation systems, Customer Relationship Management, in e-learning [Abramowicz et al., 2011] and moreover profiles might be used as expression of constant information needs [Abramowicz, 2008].

A company profile might be defined as a collection of information describing the whole enterprise, in order to provide basic facts about the company to the other entities [Hildebrandt and Backhouse, 2005, Abramowicz, 2008]. Many companies include their company profile at their web site, but there is no formal framework (or even best practices) how to structure it.

One of the few initiatives in the direction of creating specific-purpose company profiles is the research of Karol Wieloch aiming at building profiles based on information about company mentioned in text documents [Wieloch, 2011].

The contribution of the Semantic Web paradigm to the enterprises description is twofold. First, it provides ontologies that act like shared knowledge bases across the Web. Second, it provides a logic to infer how such terms may be combined to create more complex artefacts.

Tom Gruber defines an ontology as a formal, explicit specification of a shared conceptualization [Gruber, 1993]. Following [Fensel et al., 2001] the conceptualization refers to an abstract model of some phenomenon in the world that identifies the phenomenon's relevant concepts. Explicit means here that the type of concepts used and the constraints on their use are explicitly defined. Formal means that the ontology should be machine understandable. Finally, shared reflects the notion that an ontology captures consensual knowledge i.e., is not restricted to some individual but is accepted by a group.

An ontology consists of a set of concepts, axioms, and relationships that describe a domain of interest. Formally, following [Maedche and Staab, 2001], an ontology consists of five elements, namely: a set of concepts (C), hierarchy of concepts (Hc), a set of relations (R), a set of non-taxonomical relations (R→ CxC) and a set of axioms (Ao).

At the moment, semantically described data is just a small part of all data accessible in the Internet, but initiatives like Linked Open Data are gaining more and more attention [Bizer et al., 2009]. Linked Data is defined as ,,a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF."[1] Using Linked Data makes the Internet ceases to be a collection of loosely related HTML documents, but it is possible to enrich them with semantics.

The already performed analysis of the existing solutions allowed reaching a couple of conclusions – there is a vast quantity of information available, however, it is of dispersed nature and is stored in various sources. Lack of mechanisms allowing aggregating it and carrying out multi-perspective analysis taking into account various dimensions causes the need to invest lots of manual work in order to obtain required information. Once the data would be acquired, one of the approaches that may be followed to analyse the data is to use the clustering.

Clustering aims at dividing datasets into subsets (clusters), where objects in the same subset are similar to each other with respect to a given similarity measure, whereas objects in different clusters are dissimilar [Kriegel et al., 2009].

In general, clustering methods might be divided into two major subcategories: the centroid and the mediods algorithms. Centroid algorithms cluster dataset by distance between entities and before defined cluster- centers (called also gravity centres) [Dalal and Harale, 2011]. This method is probably not applicable due to the necessity to define cluster-centres before clustering. On the other hand, hierarchical clustering builds a cluster hierarchy (tree of clusters called dendrogram), might be applicable to the proposed research.

[1] `http://linkeddata.org/home`

There are many publications aiming at providing a comparison of different clustering algorithms, e.g., [Jain et al., 1999, Kriegel et al., 2009, Dalal and Harale, 2011]. Moreover [Carpineto et al., 2009] provides a survey of web clustering engines.

Despite huge efforts invested into developing clustering methods and their wide adoption in information systems, clustering of semantically described data is quite a new problem that differs significantly from the previous solutions. [Grimnes et al., 2008] pointed out two very important problems regarding semantic clustering: extracting instances from a large RDF graph and computing the distance between these instances.

One of the very few works about clustering of semantically described data is [Tagarelli and Greco, 2010]. Authors present the clustering framework SemX-Clust. It is not the only solution for clustering XML schemas, but only a few consider a semantic similarity as a distance measure, e.g., [De Meo et al., 2005, He et al., 2004, Lee et al., 2002].

An interesting solution was presented by [Hu et al., 2008]. Authors instead of using a simple bag of words for clustering, enhanced documents with semantics (synonyms, hypernyms, and associative relation based on Wikipedia links).

The clustering method operating on the semantically enhanced organization profiles proposed in this research will be built on top of the existing solutions: clustering methods and enterprise profiling. Due to the application of semantic technologies as well as incorporating existing semantic databases Web sources it constitutes, however, a novel and original solution.

3 An Outline of the Business Networks Identification Method

The proposed method will operate on dedicated profiles of various entities (organizations and objects). Automatic data processing and reasoning will be enabled by the use of semantics. At the moment there are several semantic data sources of good quality that might be used by the developed method (e.g., DBpedia[2], FreeBase[3]). Company profiles will be built based on structures of those semantic sources, enhanced with additional information that might be extracted from other databases and web pages. Profiles will be built using descriptions of organizations and products, existing on the Web (non-structural data and semantic databases). However, information extraction from Web sources is out of the scope of the carried out research – methods of information extraction from structured and unstructured sources are already developed, e.g., [Wieloch, 2011] and can be used to build organization profiles from defined sources. Profiles will be semantically enhanced: profile's elements will be linked to the developed domain ontology. Ontology will be built manually, possibly by integrating and expanding existing domain ontologies. Some relations between profile elements will be

[2] `http://dbpedia.org`

[3] `http://www.freebase.com/`

defined according to structures existing in semantic databases (e.g., from DBpedia: relations "owner of", "sponsor of", parentCompany of", etc.), other will be added manually to the developed ontology (e.g., "supplier of").

The developed method will employ the enhanced clustering algorithm to find business networks and clusters. The clustering method will operate on semantic profiles and will analyse not only entities similarities, but also relations between entities (organizations) profiles. An important problem that needs to be overcome is efficient processing of huge data sets. The method will be based on existing clustering methods for RDF, but extended in order to operate on non-hierarchical relations between data elements. Moreover, the clustering and reasoning method needs to consider the economic knowledge (i.e., mappings between relations in semantic data sources and real-world business relations have to be made). Reasoning on clustered profiles will enable identification of business networks and business clusters. Nonetheless, business network boundaries have to be defined.

Business network structure changes in time: some relations fade, other appear, new organizations join the network, etc. However, analysis of the business network dynamics is out of the scope of the current research and is planned as future work on the matter.

Method verification will be conduct using a set of test profiles, prepared according to the developed organization profile model. In the next step, the test profile set will be processed using the developed method in order to check the correctness and accuracy of the proposed solution. Solution will be evaluated in two aspects. First is method quality, i.e., whether obtained results meet identified information needs. Second aspect is experiment parameters: processing time, data volume, results correctness (precision, recall), etc.

In order to develop the envisioned solution, several challenges need to be addressed. Firstly, a method operates on entities' profiles thus, its effectiveness depends strongly on data availability and quality. As it was mentioned, methods of information extraction from Web sources are out of the scope of this research. Moreover, it is assumed that data sources are reliable and up to date. Increasing volume of semantic data sources indicates that soon enough they will be sufficient for the presented solution. Secondly, reasoning rules and methods must include knowledge from the field of economics. Mappings between relations from the existing semantic data sources and relations from real-world business networks must be defined. Analysis must consider hierarchical and non-hierarchical dependencies between data. The last big challenge is how to limit analysis of business network as one of the main characteristics of business networks is a lack of clear boundaries.

4 Research Methodology

The carried out research follows the pro-active research path. Following the methodology proposed by [Hevner et al., 2004] and [Österle et al., 2010], the below mentioned artefacts will constitute the results of the research process:

- concepts: terminology and definitions of the most important concepts in the domains of interest;
- models: a conceptual model of a company profile and sources of information on the companies in the Semantic Web settings;
- methods: the analysis method for semantically described data;
- instances: the implementation of the algorithm in order to verify the solution, based on instances of semantically described companies' profiles.

According to the followed research approach, the study plan is divided into three phases:

1. Problem diagnosis phase – identification of the state of knowledge in the area of existing models of company profiles, methods of semantic enrichment of profiles, and possible methods of analysis; this phase has already been partially conducted and research goals have been formulated;
2. Exploration phase – testing the possibilities of realization of project objectives, a review of available solutions and their possible adaptation to the objectives of the carried out research;
3. Implementation phase – implementation and verification of the proposed methods.

5 Results and Future Work

The outcome of the thesis will be a method that operates on semantically described profiles of subjects (organizations, companies) and objects (goods and services) and on relations between them, in order to identify business networks and business clusters.

According to the adopted methodology, the research started with problem diagnosis phase and state of the art was identified in order to formulate research goals. In the first part, company information needs were defined. [Citroen, 2011] based on interviews with several managers, designed a decision-making process model (Figure 2). The model is based on a rational approach thus, decision-makers try to gather as much relevant information as possible. Process phases are described with letters A-F. Data and information needed on each step are listed in squares. Arrows indicate the flow of interaction.

The decision-making process phases are as follows:

- A Preparation – problem definition, goals identification and analysis of available information (usually information at this stage is not enough to make a rational decision),
- B Analyses – retrieval and processing of additional information,
- C Specification of alternatives,
- D Limiting – analysis of results of implementing defined alternatives and choosing only those that might be successful,
- E Assessment,
- F Final – making the decision and its implementation.

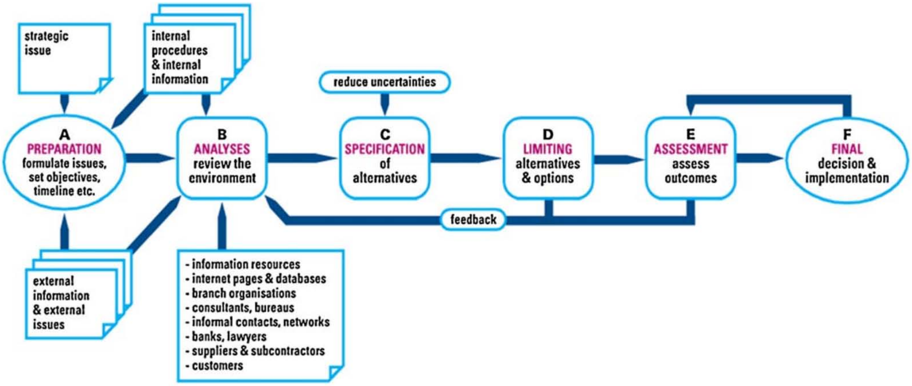


Fig. 2. Source: [Citroen, 2011]

Information needs appear mostly in phases A and B and address both internal information and information about business environment. Internal information is relatively easy to retrieve. Information about business environment should include:

- suppliers characteristics,
- customers characteristics,
- sector participants – organizations that operates in the same business sector (e.g., competitors),
- other organizations outside business sector,
- market objects (products, services) description.

Internet is one of the most common information source according to [Citroen, 2011]. With the spread of ICT and increasing volume of data indexed by search engines, it is believed that currently information retrieving is less problematic than information processing: analysis and interpretation.

Thus, the next research step was identification of data sources available on the Web, its structure, volume of data, semantic support, number of entities described. Several types of data sources were identified. The comparison is provided in table 1. It can be noticed that semantic data sources are characterized by lower volume of data, but contain more specific information, provided in a structured form. Semantic data sources will be used as a basis for the developed profile of organization. However, additional information from other sources (business directories, company home pages or news) might be added by using the existing information extraction methods (e.g., [Kaczmarek et al., 2010]) and enhancing it with semantic descriptions [Abramowicz et al., 2011].

Semantic data sources include some relations that appear in business networks. Examples of relations from DBpedia:

dbpedia:Gucci **is dbpedia-owl:owner of** dbpedia:Bottega_Veneta
dbpedia:Apple_Inc. **is dbpedia-owl:successor of** dbpedia:NeXT

dbpedia:Nestle **is dbpedia-owl:parentCompany of**
dbpedia:Gerber_Products_Company
dbpedia:Nestle **is dbpprop:sponsor of**
dbpedia:Circle_of_Life:_An_Environmental_Fable

Table 1. Web sources for organization information comparison

—**Source**		**Range**	**Author**	**Description**	**Type of data**
home pages		almost every organization has one	organization	unstructured	various
articles, news, press information		n/a	third parties	unstructured	various, distributed content
business directories	Panorama Firm	> 2,3 mln companies	company	half-structured	contact data, line of business, optional product portfolio
	YellowPages	n/a	company	half-structured	contact data, line of business, optional product portfolio
structured databases	Crunchbase	> 100 000 companies	users, company	semantic	detailed company description, line of business, contact data, owners, product portfolio, optional financial data
Linked Data	FreeBase	1125 ,,Company" instances	community	semantic	detailed company description, line of business, contact data, owners, product portfolio, optional financial data
	DBpedia	300 ,,Company" instances	community	semantic	detailed company description, line of business, contact data, owners, product portfolio, optional financial data

The identified data sources were confronted with information needs within the decision-making process. For each business environment's element main characteristics were defined as well as exemplary data sources from which the information might be retrieved. The results of this analysis are presented in table 2.

Table 2. Information needs and data sources

Business environment element	**Exemplary characteristics**	**Data sources**
customers	individuals: demographic statistics, geo-localisation, language; enterprises: contact data, scope of business	individuals: social portals, statistics from organization homepage; enterprises: Linked Data, structured databases, business directories;
suppliers	contact data, scope of business, organization structure	Linked Data, structured databases, business directories;
sector participants	contact data, scope of business, organization structure	Linked Data, structured databases, business directories;
organizations outside sector	contact data, scope of business, organization structure	Linked Data, structured databases, business directories;
market objects	functions, application, components, version history, popularity, buyers' rating	Linked Data, product recommendation portals, applications markets;

Current work is focused on the development of a semantically described company profile and selection of a representation technique. Future work will include analysis of existing data analysis methods, including clustering methods, in order to assess their suitability for the semantically described data, as well as a development of data analysis method for semantically described company profiles. Simultaneously evaluation and validation of the developed artefacts will be performed.

6 Conclusions

The main results of the research will be the upper level ontology constituting a company's profile and linked to other sources available in the Semantic Web, together with the data analysis method addressing information needs of companies operating on the electronic markets. The most important effect of the research is therefore to enable the use in the economy the full potential of the artefacts already existing on the market, i.e., models (by providing semantically described company profiles), products (by providing sets of semantically described data) and methods (by providing data analysis algorithms). Thus, the results will also contribute to the realization of the Semantic Web vision.

Acknowledgments. The work published in this article was supported by the project titled: ,,Clustering Method to Support Strategic Decision-Making Process in the Semantic Web" financed by the National Science Centre (NCN) (decision no. DEC-2012/05/N/HS4/01875).

References

[Abramowicz, 2008] Abramowicz, W.: Filtrowanie Informacji. Wydawnictwo Akademii Ekonomicznej w Poznaniu (2008)

[Abramowicz et al., 2011] Abramowicz, W., Bukowska, E., Kaczmarek, M., Starzecka, M.: Semantic-enabled efficient and scalable retrieval of experts. In: Proceedings of International Conference on Information, Process, and Knowledge Management (eKNOW 2011), Gosier, Guadeloupe, pp. 30–35. IARIA (2011)

[Bizer et al., 2009] Bizer, C., Heath, T., Berners-Lee, T.: Linked data – the story so far. International Journal on Semantic Web and Information Systems, Special Issue on Linked Data 5(3), 1–22 (2009)

[Campbell and Wilson, 1996] Campbell, A., Wilson, D.: Managed Networks. In: Networks in Marketing. Sage Publications, USA (1996)

[Carpineto et al., 2009] Carpineto, C., Osiński, S., Romano, G., Weiss, D.: A survey of web clustering engines. ACM Comput. Surv. 41(3), 17:1–17:38 (2009)

[Citroen, 2011] Citroen, C.L.: The role of information in strategic decision-making. International Journal of Information Management 31, 493–501 (2011)

[Dalal and Harale, 2011] Dalal, M.A., Harale, N.D.: A survey on clustering in data mining. In: Proceedings of the International Conference & Workshop on Emerging Trends in Technology, ICWET 2011, pp. 559–562. ACM, New York (2011)

[De Meo et al., 2005] De Meo, P., Quattrone, G., Terracina, G., Ursino, D.: An approach for clustering semantically heterogeneous XML schemas. In: Meersman, R., Tari, Z. (eds.) OTM 2005. LNCS, vol. 3760, pp. 329–346. Springer, Heidelberg (2005)

[De Wit and Mayer, 2007] De Wit, B., Mayer, R.: Synteza strategii. Polskie Wydawnictwo Ekonomiczne, Warszawa (2007)

[Fensel et al., 2001] Fensel, D., Horrocks, I., Harmelen, F.V., McGuinness, D.L., Patel-Schneider, P.F.: Oil: An ontology infrastructure for the semantic web. IEEE Intelligent systems. Special Issue on the Semantic Web 16(2) (2001)

[Ford et al., 2011] Ford, D., Gadde, L., Hakansson, H., Snehota, I.: Managing Business Relationships. Wiley (2011)

[Grimnes et al., 2008] Grimnes, G.A., Edwards, P., Preece, A.: Instance based clustering of semantic web resources. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds.) ESWC 2008. LNCS, vol. 5021, pp. 303–317. Springer, Heidelberg (2008)

[Gruber, 1993] Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. In: Guarino, N. (ed.) Proceedings of Padua Workshop on Formal Ontologies (1993)

[He et al., 2004] He, B., Tao, T., Chang, K.C.-C.: Organizing structured web sources by query schemas: a clustering approach. In: Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management, CIKM 2004, pp. 22–31. ACM, New York (2004)

[Hevner et al., 2004] Hevner, A., March, S., Park, J., Ram, S.: Design science in information systems research. MIS Quarterly 28(1), 75–105 (2004)

[Hildebrandt and Backhouse, 2005] Hildebrandt, M., Backhouse, J.: D7.2: Descriptirve analysis and inventory of profiling practices. Technical report, FIDIS (Future of Identity in Information Society) Deliverable (2005)

[Hu et al., 2008] Hu, J., Fang, L., Cao, Y., Zeng, H.-J., Li, H., Yang, Q., Chen, Z.: Enhancing text clustering by leveraging wikipedia semantics. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2008, pp. 179–186. ACM, New York (2008)

[Hkansson and Snehota, 2006] Håkansson, H., Snehota, I.: No business is an island: The network concept of business strategy. Scandinavian Journal of Management 22(3), 256–270 (2006)

[Jain et al., 1999] Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Comput. Surv. 31(3), 264–323 (1999)

[Jarillo, 1993] Jarillo, J.: Strategic Networks. Taylor & Francis (1993)

[Kaczmarek et al., 2010] Kaczmarek, T., Zyskowski, D., Walczak, A., Abramowicz, W.: Information extraction from web pages for the needs of expert finding. Studies in Logic, Grammar and Rethoric, Logic Philosophy and Computer Science 22(35), 141–157 (2010)

[Kriegel et al., 2009] Kriegel, H.-P., Kröger, P., Zimek, A.: Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. ACM Trans. Knowl. Discov. Data 3(1), 1:1–1:58 (2009)

[Lee et al., 2002] Lee, M.L., Yang, L.H., Hsu, W., Yang, X.: Xclust: Clustering xml schemas for effective integration. In: Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM), pp. 292–299 (2002)

[Maedche and Staab, 2001] Maedche, A., Staab, S.: Ontology learning for the semantic web. IEEE Intelligent Systems (2), 72–79 (2001)

[Österle et al., 2010] Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., Sinz, E.J.: Memorandum on design-oriented information systems research. European Journal of Information Systems 20(1), 7–10 (2010)

[Ratajczak-Mrozek, 2010] Ratajczak-Mrozek, M.: Sieci biznesowe a przewaga konkurencyjna przesibiorstw zaawansowanych technologii na rynkach zagranicznych. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Pozna (2010)

[Tagarelli and Greco, 2010] Tagarelli, A., Greco, S.: Semantic clustering of xml documents. ACM Trans. Inf. Syst. 28(1), 3:1–3:56 (2010)

[Wieloch, 2011] Wieloch, K.: Budowanie profili przedsibiorstw z wykorzystaniem utosamiania odwoa semantycznych do przedsibiorstw w polskich tekstach ekonomicznych. PhD thesis, Poznan University of Economics (2011)

Conceptual Semantic Representation of 3D Content*

Jakub Flotyński and Krzysztof Walczak

Poznań University of Economics,
Niepodległości 10, 61-875 Poznań, Poland
{flotynski,walczak}@kti.ue.poznan.pl
http://www.kti.ue.poznan.pl

Abstract. The complexity of 3D content makes its creation accessible to professional IT developers equipped with specific software tools and hardware devices, but it is generally inaccessible to non-expert users. The Semantic Web approach enables description of web resources with commonly used concepts. However, the use of semantic concepts may also facilitate creation of interactive 3D content. In this paper, a new approach to conceptual semantic representation of 3D content is proposed. The presented solution permits modelling of 3D content at an arbitrarily high level of semantic abstraction with the use of domain-specific ontologies. Thanks to the compliance with well-established solutions for semantics and 3D content representation, the proposed approach can facilitate widespread creation, dissemination and reuse of 3D content by non IT-professionals in a variety of application domains.

Keywords: 3D web, Semantic Web, 3D content, semantic modelling, ontology.

1 Introduction

Widespread use of interactive 3D technologies and multimedia systems, including virtual reality (VR) and augmented reality (AR), has been recently enabled by significant progress in hardware performance, rapid growth in the available network bandwidth, as well as the availability of versatile input-output devices. The primary element of VR/AR applications is three-dimensional content. The creation of interactive 3D content is typically more complex than the creation of typical web resources, as it may concern a variety of aspects—geometry, structure, space, appearance, logic and behaviour. Hence, this process is currently accessible mainly to professional developers equipped with specific 3D modelling tools and hardware devices (e.g., 3D scanners), and it is generally inaccessible to non IT-professionals. The increasing popularity of 3D/VR/AR systems in various application domains requires development of efficient modelling methods and tools that are easy-to-use by domain experts who are not required to have technical skills in 3D modelling and programming.

The Semantic Web standards enable description of diverse types of web resources, including 3D content, with commonly used concepts. However, the use of semantic concepts may also facilitate the modelling of interactive 3D content.

* This research work has been supported by the Polish National Science Centre Grant No. DEC-2012/07/B/ST6/01523.

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 244–257, 2013.

The main contribution of this paper is a semantic mapping model for conceptual semantic representation of interactive 3D content. The proposed solution enables modelling of 3D content by non-IT professionals at an arbitrarily chosen (arbitrarily high) level of semantic abstraction. The model leverages domain-specific ontologies, liberating domain experts from technical concerns that are typical for 3D content. The model is platform- and standard-independent, but it conceptually complies with well-established 3D content and semantic representation standards, facilitating widespread creation, dissemination, reuse and reasoning on 3D content in a variety of application domains.

The remainder of this paper is structured as follows. Section 2 provides an overview of the current state of the art in the domain of semantic description and semantic creation of 3D content. Section 3 contains a brief introduction to a method of semantic modelling of 3D content. Section 4 focuses on the new mapping model, which is a part of the modelling method. Section 5 presents an implementation of the proposed model. Section 6 discusses an example of the conceptual semantic design of a 3D room with a domain-specific ontology. Finally, Section 7 concludes the paper and indicates the possible directions of future research.

2 Related Works

In this section, Semantic Web technologies and methods of semantic creation of 3D scenes are considered.

2.1 Semantic Web

The primary technique for describing the semantics of web content is the Resource Description Framework (RDF) [1]—a standard devised by the W3C. RDF introduces basic means for making statements about resources. The RDF Schema (RDFS) [2] and the Web Ontology Language (OWL) [3] are W3C standards based on RDF, providing higher expressiveness for semantic descriptions of web resources. OWL defines a set of profiles, which differ in complexity and decidability [4]. The Semantic Web Rule Language (SWRL) [5] is an extension to OWL that permits semantic Horn-like rules. The Reaction Rule Markup Language (RuleML) [6] enables declarative description of reaction rules, in particular in the event-condition-action paradigm.

2.2 Semantic Creation of 3D Scenes

Several works have been devoted to semantic creation of 3D content. In [7], an approach to designing interoperable RDF-based Semantic Virtual Environments, with system-independent and machine-readable abstract descriptions has been presented. In [8], [9], [10], a rule-based framework using MPEG-7 has been proposed for the adaptation of 3D content, e.g., geometry and texture degradation, and filtering of objects. The content can be described with different encoding formats (in particular X3D), and it is annotated with an indexing model. In [11], the integration of X3D and OWL using scene-independent ontologies and the concept of semantic zones have been proposed to enable querying 3D scenes at different levels of semantic detail.

In [12], a method of structured design of VR content has been proposed. In [13], [14], [15], an approach to generating virtual worlds upon mappings of domain ontologies to particular 3D content representation languages (e.g., X3D) has been considered. The following three content generation stages are distinguished: specification of a domain ontology, mapping of the domain ontology to a 3D content representation language, and generation of the final presentation. The solution stresses spatial relations (position and orientation) between objects in the scene.

Several works have been conducted on the modelling of behaviour of VR objects. In [16], the Beh-VR approach and the VR-BML language have been proposed for the creation of 3D content with behaviour (interactions and animations). The proposed solution aims at simplification of behaviour programming for non-IT professionals. Another method facilitating the modelling of content behaviour [17], [18], [19] provides a means of expressing primitive and complex behaviours as well as temporal operators. A tool-supported design approach to defining the behaviour of objects in X3D scenes has been presented in [20]. Finally, a rule-based ontology framework for feature modelling and consistency checking has been proposed in [21].

3 Overwiev of the Method of Semantic Modelling of 3D Content

Although several approaches have been proposed for conceptual creation of 3D scenes, they lack general solutions for semantic modelling of interactive 3D content, its components, properties and relations, at an arbitrarily chosen level of abstraction, by using various domain-specific ontologies.

The paper presents an outline of a method of semantic modelling of interactive 3D content with the focus on a mapping model of semantic domain-specific concepts to concrete semantic 3D content components and properties. In the presented method, modelling of 3D content may be performed at three distinct stages, which are partly dependent—design of concrete semantic 3D content components, mapping domain-specific concepts to the concrete semantic 3D content representation, and conceptual design of 3D content based on domain-specific concepts (Fig. 1). Every stage uses an appropriate semantic model.

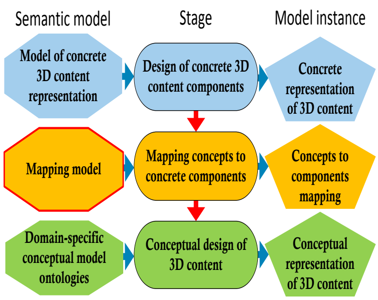


Fig. 1. Semantic modelling of 3D content

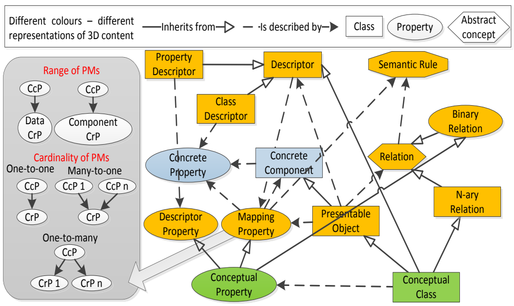


Fig. 2. Semantic mapping model for conceptual 3D content representation

At the first stage of modelling, a concrete representation of 3D content at a low level of abstraction and a high level of detail is created. The representation is created according to the multi-layered concrete model of interactive 3D content, which has been proposed in [22]. The model enables separation of concerns between six layers corresponding to distinct aspects of 3D content and different stages of 3D content design—*Geometry*, *Structure*, *Appearance*, *Scene*, *Logic* and *Behaviour* layers. The layers define concrete components (CrCs) and concrete properties (CrPs) that allow for describing different aspects of 3D content and which are common for well-established 3D content representation languages (e.g., VRML [23] or X3D [24]). The primary *Geometry Layer* introduces basic uniform individual geometrical CrCs and CrPs, e.g., lines, planes and meshes. The second *Structure Layer* introduces complex structural CrCs that assemble geometrical CrCs, enabling definition of spatial CrPs for them, e.g., position, orientation and size. The *Appearance Layer* adds appearance to geometrical and structural CrCs, e.g., colour, transparency and texture. The *Scene Layer* extends structural CrCs to navigable scenes with viewpoints. The *Logic* and *Behaviour* layers enrich CrCs that have been defined in the previous layers, with logic and behaviour, in particular animations and interactions. The layers are partly dependent—every layer uses only its concepts and concepts specified in its lower layers. Modelling of 3D content with the semantic model can be performed at any layer. For instance, design of a complex 3D scene with behaviour involves CrCs from all of the layers, while reusable 3D objects that are to be injected into different complex 3D scenes may be, e.g., only structural CrCs without appearance, thus designed at the *Structure Layer*.

The resulting concrete representation is a knowledge base compliant with the concrete model, which includes CrCs and CrPs corresponding to different layers of the content. Since the CrCs and CrPs cover aspects typical for 3D content, this stage of modelling is performed by a developer with technical skills in 3D modelling.

At the second stage of modelling, semantic domain-specific concepts that are defined in the selected domain-specific ontology, are mapped to the concrete 3D content representation created at the previous stage. The mapping is an ontology compliant with the mapping model, which will be discussed further. The mapping extends the domain-specific ontology with mapping concepts. Like the concrete representation, the mapping is created by a developer using a modelling tool – once for a particular domain-specific ontology, and it may be reused for representing various conceptual objects (CpOs) and conceptual properties (CpPs) that conform to this ontology.

At the last stage of modelling, conceptual semantic 3D content representation at an arbitrarily chosen (arbitrarily high) level of abstraction is created. The conceptual representation is a knowledge base compliant with the selected domain-specific ontology. The conceptual representation consists of conceptual objects (CpOs)—instances of domain-specific classes, which are described by conceptual (domain-specific) properties (CpPs). Complex CpOs, which include sub-CpOs, specify viewpoints and navigation modes, are referred to as scenes. The conceptual representation consists of CpOs and CpPs that are abstract in the sense of their final presentation as, in general, they could be presented in different manners (e.g., a 2D graphics, 3D model, voice, etc.). The representation may be created by a non IT-professional who has a knowledge of the domain-specific ontology selected for the modelling process. All conceptual representations, which are built upon common domain-specific ontologies, may reuse the same concrete representations and the same mappings, thus becoming presentable in three-dimensions.

The proposed approach has several important advantages in comparison to the available solutions for modelling of 3D content. First, the presented approach facilitates creation and analysis of 3D content at different levels of semantic abstraction. Second, the conformance to well-established Semantic Web standards and tools simplifies creation, dissemination and reuse of conceptual objects and scenes. It fosters the reuse of existing domain ontologies and knowledge bases. In addition, referring to the domain-specific meaning of particular 3D objects may improve creating, searching, exploring and reasoning on 3D content by domain experts who no longer need to go into details specific to diverse aspects of 3D content. Like in high-level (e.g., object-oriented) programming, the presented approach permits the creation of reusable CrCs and CpOs, which may be combined in different scenes. Next, the semantics of CrCs and CpOs can specify applicability and compatibility between different objects which are used in the designed content. Furthermore, 3D content described by the concrete semantic representation is platform and standard-independent, and it may be transformed to final 3D representations encoded in different languages, depending on particular requirements, e.g., the context of interaction, user location, preferences as well as hardware and software used.

4 Mapping Domain-specific Concepts to a Concrete Representation

Although high-level conceptual semantic description of virtual environments is sufficient for typical semantic processing and analysis, it must be extended to enable semantic representation of interactive 3D content. Such representation usually requires additional levels of detail that allow to accurately reflect the semantic concepts used by different components and properties of 3D content. For instance, stating that one car outruns another car is sufficient for a typical semantic analysis, but, to be presented, it requires an animation that changes the positions of both objects, which represent the cars.

In this section, a new semantic mapping model for conceptual representation of interactive 3D content is proposed. The model is an element of the semantic modelling method, which has been outlined in the previous section. The model enables mapping of domain-specific concepts to a concrete 3D content representation, thus permitting 3D content creation at an arbitrarily high level of abstraction.

The proposed mapping model (Fig. 2) links domain-specific concepts with CrCs and CrPs, which are a low-level detailed representation of these domain-specific concepts. To clearly separate the stages of mapping and conceptual design and to minimally affect the domain-specific concepts used, mapping concepts are created as super-concepts of domain-specific concepts and they indicate appropriate representational CrCs and CrPs. Domain-specific concepts hereby become mapping concepts, and individuals of domain-specific classes (CpOs) are individuals of mapping classes. Furthermore, statements made on a mapping concept are relevant to its domain-specific sub-concepts without introducing direct modifications of these domain-specific concepts. Such extension of the domain-specific ontologies does not affect the way in which conceptual design is performed by a domain expert, who does not need to take into account the low-level concrete representation of the domain-specific concepts used.

In the proposed approach, conceptual design of 3D content is based on CpOs, which are instances of domain-specific classes. CpOs may be described and linked by conceptual properties (CpPs). Mapping domain-specific concepts to CrCs and CrPs enables representation of CpOs and CpPs by combinations of these CrCs and CrPs. In this way, CpOs (instances of domain-specific classes) and their CpPs become presentable. The proposed model does not preclude representations of CpOs as combinations of other CpOs. Such representations can be defined during the conceptual design by typical means provided by the Semantic Web standards (such as combination by inheritance or restriction), thus they are not discussed in this paper. Mapping concepts that define complex dependencies are described with auxiliary *semantic rules* (SRs). Specifying representations for particular domain-specific concepts improves efficient modelling and reusability of these concepts in contrast to defining individual representations for particular CpOs. A CpO in the modelled content has all of the presentational effects of its classes and super-classes.

The modelled 3D content is semantically complete in the sense that no semantic concepts or individuals are added or removed during its presentation – neither in the conceptual nor in the concrete representation, only the values of individual properties may change. For instance, the motion of an object may be stopped by turning off its animation, but not by removing the animation property.

The introduced mapping concepts may form hierarchical structures with multiple inheritance. The 3D representation of a descendant concept extends (and not restricts) the representations (sets of presentational effects) of its parent concepts with additional presentational effects. Presentational effects implied by descendant concepts take precedence over the effects of their parent concepts. For example, objects made of silver may have the same reflectiveness but different colours than other metal objects.

While the semantic representations of static 3D content (without logic and behaviour) are in general decidable, introducing complex descriptions of logic and behaviour can make them undecidable, which affects possible reasoning on the semantic representations but still enables proper presentation of the content.

The following sections explain the details of the proposed mapping model with the taxonomy and semantics of its particular entities. Since the introduced mapping concepts are ascendants of domain-specific concepts, the presented discussion is also relevant to these domain-specific concepts.

4.1 Presentable Objects

A *presentable object* (PO) is the primary entity of a conceptual semantic representation of 3D content. Every class from a domain-specific ontology whose instances need to have individual representations in the created 3D content, must be specified as a sub-class of a PO class. Each PO class is a sub-class of the *geometrical component* or the *structural component* class of the multi-layered 3D content model. During the design of CrCs, a developer may arbitrarily extend PO classes by determining various CrPs to represent the geometry, structure, space, appearance, logic or behaviour of POs, thus making their child domain-specific classes also sub-classes of the *geometrical*, *structural*, *spatial*, *appearance*, *logical* or *behavioural component* classes of the concrete representation model. As the CrPs concern specific aspects related to 3D content, they are not intended to be manipulated by a domain expert (a non-IT professional) during the conceptual design. Therefore, only CrPs that are common for different instances of POs should be specified as such inherent properties of their PO class, e.g., accurate reusable models of plants and chemical compounds may be reused in different contexts always with the same values of properties. Properties of the concrete model that often need to be differentiated for different POs are specified using other concepts of the proposed mapping model.

During the conceptual design, POs in the content may be described with the other mapping concepts of the proposed model—*mapping properties*, *descriptors* and *relations*. In contrast to POs, the remaining concepts reflect aspects of 3D content that have no individual representations in the final 3D scene.

4.2 Properties of Presentable Objects

In the proposed semantic model, POs may be described by *mapping properties* (MPs). Each MP is a super-property of some CpPs and it links them to representational CrPs. The values of representational CrPs are functions of the values of their primary MP, e.g., the RGB triple '1 0 0' corresponds to the 'red' colour. Depending on the complexity of the concrete representation of an MPs, complex MPs may require auxiliary SRs, which determine the values of the representational CrPs upon the values of these MPs. The value of each MP is a literal of a simple data type (e.g., integer, string or date).

Different categories of *property mappings* (PMs) are distinguished with regards to the range of representational CrPs as well as the cardinality of PMs. These two groups are mutually independent, as each PM belongs to both of them.

Range of Property Mappings. Two types of PMs are distinguished in the proposed model in terms of the range of the representational CrPs that are used in the mapping—*data* PM and *component* PM. A *data* PM links CpPs to CrPs whose ranges are literals of a simple data type (e.g., integer, string, date, etc.). A *component* PM links CpPs to CrPs whose ranges are CrCs. *Component* PMs enable describing features of POs that cannot be sufficiently described by *data* PM. For instance, a 3D object moving on a surface is connected to a CrC that reflects the animation and specifies a set of necessary values, such as time, speed, recurrence, etc. A *component* PM requires an auxiliary SR determining appropriate CrPs of the associated CrC with regard to the primary CpP. Each *combined* PM is a combination of *data* and *component* PMs.

Cardinality of Property Mappings. The cardinality of links in a particular PM determines the numbers of CpPs and CrPs that are used in the PM. Three types of PMs are distinguished. A *one-to-one* PM links a single CpP to a single CrP. This category of PMs is convenient for CpPs whose concrete representations depend only on the values of these CpPs, and not on any other CpPs of the described PO, and vice versa—these CpPs do not influence any other CrPs. *One-to-one* PMs are sufficient for determining independent features of POs, e.g., colour, pattern, size, etc. A *one-to-one equivalent* PM is the simplest PM, which copies the value of the primary CpP to the CrP (e.g., a direct RGB colour description).

A *many-to-one* PM links a number of CpPs to a single CrPs. This category of PMs is convenient for CrPs whose values are determined by a number of CpPs. *Many-to-one* PMs are used for describing advanced and relative features, e.g., the colour of an object may be determined by both its material and temperature. *Many-to-one* PMs have more complex descriptions than *one-to-one* PMs and they require additional auxiliary SRs that determine the exact aggregated value of the CrP by linking this value to the different associated CpPs.

A *one-to-many* PM links a single CpP to a number of CrPs. The CpP aggregates the semantics of multiple associated CrPs. *One-to-many* PMs are convenient for conceptual semantic description on a high level of abstraction that hides the details of the low-level concrete representation and aggregates multiple low-level CrPs. For instance, a PO made of gold is yellow, not transparent and it reflects light. *One-to-many* PMs require SRs for determining the values of the CrPs upon the primary CpP value. A *many-to-many* PM is a combination of the previous categories of PMs.

4.3 Descriptors of Presentable Objects

In some cases, MPs are not sufficient for semantic modelling of 3D content when POs need to be described by domain-specific concepts that gather multiple domain-specific CpPs and that need to be defined and be legible in detail to a modelling user. Such a requirement is not satisfied by MPs that use *one-to-many* PMs, which aggregate multiple CrPs of the concrete model by SRs. SRs are specified in the concrete representation by a developer, and thus they are hidden to a user modelling on the conceptual level.

Semantic *descriptors* are a functional extension of *one-to-many* PMs, which are used for describing POs. A *descriptor* determines an arbitrary set of CrPs or CpPs. Although these properties are associated with the *descriptor*, they describe the PO that the *descriptor* is applied to, i.e. the *descriptor* only carriers the properties. Two types of *descriptors* are distinguished.

Class Descriptors. A *class descriptor* (CD) is a class, for which a set of CrPs is specified. A CD may be used as a super-class for arbitrarily selected PO classes, or as a class for arbitrarily selected POs. Unlike PO classes, CDs are not used to create independent 3D objects existing within a scene. Instead, they are used to enrich POs with additional CrPs that cannot be defined inherently for a particular PO class, because they need to be assigned selectively only to some instances of PO classes. For example, a class of interactive rotating POs includes only selected POs that rotate after being touched. The CrPs of a CD are not modified during the conceptual design.

Property Descriptors. A *property descriptor* (PD) is a semantic individual that determines a set of MPs. A PD is connected to the described PO by a *descriptor property*. Although a PD may be described by numerous MPs, these MPs are related to the described PO. For example, furniture can be made of different types of wood, each of which is described by a few properties such as colour, shininess, texture, etc.

These MPs need to be specified by a domain expert individually for different pieces of furniture. The MPs of a PD may be modified during the conceptual design.

4.4 Relations of Presentable Objects

The proposed model provides *relations* (RLs) to permit semantic modelling of dependencies between POs. Every RL has at least two parts (participants), which are connected to one another by mutual dependencies related to some aspects of 3D content—geometry, structure, space, appearance, logic or behaviour—reflected by CrPs. In each RL, at least one part affects other parts and at least one part is affected by other parts, and each part is either affecting or affected. In the concrete representation of an RL, CrPs of at least a single PO depend on CrPs of at least one another PO.

Cardinality of Relations. In terms of cardinality, *binary* RLs and *n-ary* RLs are distinguished. A *binary* RL links two POs and it is represented with an auxiliary SR that subordinates selected CrPs of the affected part with selected CrPs of the affecting part. In *uni-directional* RLs one part is affecting and the other one is affected. For instance, an object (affected part) that is placed on a table (affecting part) takes its height as one of the position coordinates without changing any properties of the table. In *bi-directional* RLs both parts are affecting and affected. For example, a bullet hitting a wall breaks up and partly damages the wall.

In contrast to *binary* RLs, *n-ary* RLs allow for modelling complex dependencies between a number of POs, e.g., an RL that defines the relative position between three objects does not only indicate these POs, but also specifies their relative orientations and distances between them.

Presentational Variants of Relations. The use of auxiliary SRs in RLs does not only allow for describing a mutual influence of parts of RLs, but also provides the possibility to distinguish different *presentational variants* of the modelled RLs (presentational polymorphism of semantic RLs). While the semantics of an RL remains the same (what is important when searching, exploring and reasoning on the scene in the typical semantic manner) and it is not desirable to distinguish different semantic sub-RLs, its 3D representations may be different depending on, e.g., types of the described parts or their properties. For instance, sitting on a sofa differs from sitting on a floor. Moreover, different subjects sit in different manners, e.g., a dog or a human. While the 'sitting' RL remains the same for different semantic analyses, it requires different final 3D representations, depending on the parts of the RL.

5 Implementation of the Proposed Mapping Model

The proposed mapping model for conceptual representation of 3D content can be implemented as an ontology with the well-established Semantic Web technologies (RDF [1], RDFS [2], OWL [3] and Reaction RuleML [6]). The implementation of the particular concepts of the model is explained below.

Each PO class is an `owl:Class`. It may become a super-class of any class of a domain-specific ontology whose individuals need to have independent representations in the final 3D scene. Properties which are inextricably connected to all POs of a particular PO class, are specified using the `owl:hasValue` and the `owl:someValuesFrom` restrictions on the level of the definition of this PO class.

Each MP is an `owl:DatatypeProperty` supported by an auxiliary RuleML semantic `Rule`. The `rdfs:domain` of an MP indicates the appropriate domain classes, its `rdfs:range`—the acceptable data types. *One-to-one equivalent* PMs are implemented by the `owl:equivalentProperty`.

Each CD is an `owl:Class` with `owl:hasValue` restrictions indicating the desirable values of particular CrPs, which are gathered by the CD. Each PD class is an `owl:Class` with `owl:someValuesFrom` restrictions. Each *descriptor property* is an `owl:ObjectProperty` with an `rdfs:domain` and an `rdfs:range` set to the described PO classes and the describing PD class respectively.

Each *binary* RL is an `owl:ObjectProperty` supported by RuleML `Rules`. The `rdfs:domain` and the `rdfs:range` of a *binary* RL indicate the parts of the RL. Each *n-ary* RL is an `owl:Class` whose individuals determine the parts of the *n-ary* RL and its desirable CrPs.

Inheritance of the mapping concepts is implemented with `rdfs:subClassOf` property (for PO classes, descriptors and n-ary RLs) and `rdfs:subPropertyOf` (for MPs, *descriptor properties* and *binary* RLs).

Due to the restrictions on class members that are used for POs and CDs, at least the expressiveness of the OWL EL or OWL RL profiles is required for the proposed model. Hence, in general, the semantic representations of static 3D objects and static 3D scenes (without logic and behaviour) are decidable with the polynomial combined complexity for the majority of reasoning problems [4]. The representations of dynamic 3D content (with logic and behaviour) may be undecidable because of the use of SRs [5]. Although, in general, the undecidability makes it impossible to find solutions for some queries against the semantic representations of 3D content, the gain achieved by logic and behaviour description is more important for efficient modelling of complex 3D content.

6 Example Conceptual Design of an Interactive 3D Scene

In this section, an example of conceptual semantic design is presented to demonstrate the main benefit of the semantic representation model proposed—an easy-to-use method of modelling 3D content by non IT-professionals. The example scene presents a room, e.g., for e-commerce flat presentation or design systems. The two other stages of the semantic modelling of 3D content—the design of concrete 3D content components and mapping concepts to the concrete representation—are, in general, rather complex processes, which are typically performed once by a developer for a particular set of domain-specific ontologies. These processes have been outlined in Section 3, and they are not considered in this example.

A conceptual scene description encoded by a non-IT professional is presented in Listing 1.1 (the RDF-Turtle format). Its final 3D representation is presented in Fig. 3[1]. Since the scene is a knowledge base, which may be based on a domain-specific ontology, it can be built with any semantic modelling tool (e.g., Protégé). Since the scene is abstract in the sense of its final presentation, it does not cover any aspects that are directly related to 3D content. Such aspects are addresses by a developer during the design of concrete 3D content components.

```
1:  room rdf:type Room.
2:  floor rdf:type Floor.
3:  table rdf:type Table.
4:  chair rdf:type Chair.
5:  flowerpotChair rdf:type Chair.
6:  tableFlowerpot rdf:type Flowerpot.
7:  chairFlowerpot rdf:type Flowerpot.
8:  decoration rdf:type Decoration.
9:  table isMadeOf "wood".
10: decoration isMadeOf "metal".
11: tableFlowerpot rdf:type
      RotatingObject.
12: chairAppearance rdf:type
      AppearanceDescriptor.
13: chair appearance chairAppearance.
14: chairAppearance shininess "0.5␣0.4
      ␣0.3".
15: chairAppearance material "leather"
      .
16: room hasFloor floor.
17: table standsInTheMiddleOf room.
18: chairFlowerpot standsOn
      flowerpotChair.
19: tableFlowerpot standsOn table.
20: decoration standsOn table.
21: chairRelativePosition hasPart
      chair, flowerpotChair ; distance
      "10".
```

Listing 1.1. A conceptual semantic representation of a 3D scene

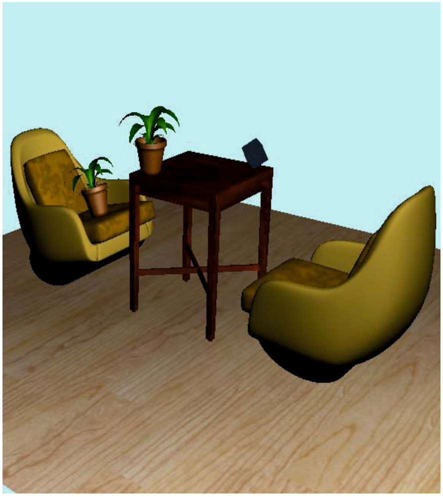


Fig. 3. A final 3D representation of the scene

Lines 1-8 form the scene (a room) and POs, which are incorporated in it. It is assumed that the mapping has already been performed for the scene and it introduces mapping concepts, which are super-concepts of the domain-specific concepts, so the discussion of this example concerns the mapping concepts instead of their domain-specific descendants. The MPs indicate materials that the `table` and the `decoration` are made of (9,10). The `isMadeOf` MP uses a *one-to-many* PM, as it determines the texture or the colour of the PO and its reflectiveness. The `RotatingObject` CD assigned to the `tableFlowerpot` (11) enables a single rotation of the flowerpot after touching it. The `chairAppearance` PD (12-15) specifies the `shininess` and the `material` of the `chair`, which may be managed by a domain expert.

[1] In the final 3D presentation of the scene, the 3D models of a table (`http://resources.blogscopia.com/2011/05/19/bar-table-and-chairs/`), a chair (`http://resources.blogscopia.com/2010/05/11/armchair-2-download/`) and a flowerpot (`http://resources.blogscopia.com/2010/04/22/flowerpot/`) have been used.

To enable placing furniture in the room, a reference surface must be given (a floor—16). The `standsInTheMiddleOf` *binary* RL places the table in the middle of the floor respecting its size (17). The `tableFlowerpot` and the `decoration` may be placed in an arbitrary point on the `table`, as they are described by the generalized `standsOn` *binary* RL (19,20), which is an ascendant RL of the `standsInTheMiddleOf`. Both flowerpots in the scene stand on some furniture (they are in the same type of RL with different POs—18,19), but the visual presentations are different in these two cases. The `chairFlowerpot` tilts the soft surface of the `flowerpotChair`, while the `tableFlowerpot` does not affect the hard surface of the `table`. The relative position of the chairs is described by the `distance`, which is set to 10 units in the n-ary `chairRelativePosition` RL (21).

In the presented example, the transformation of the conceptual scene representation to its final 3D representation encoded in VRML has been performed manually. However, the development of a tool for automatic semantic transformations is planned. The conceptual representation presented in Listing 1.1 includes 15 times less characters than its final equivalent. The presented solution enables creating representations of 3D content that are much more concise than representations encoded in typical 3D content representation languages.

7 Conclusions and Future Works

In this paper, a new mapping model for conceptual semantic representation of 3D content has been proposed. The proposed approach leverages the existing Semantic Web standards to enable modelling of 3D content by non-IT professionals at an arbitrarily chosen level of abstraction, using domain-specific ontologies and knowledge bases. The proposed approach may simplify 3D content creation, dissemination and reuse in multiple application domains, such as building commercial and museum 3D exhibitions, games and advanced business presentations.

The proposed mapping model has some limitations. First, it is inconvenient for conceptual representations of simple 3D objects and scenes that are neither reused nor shared, and that are accessed by authors (who know their semantics). In such cases, referring to the high-level semantics of the modelled content is not necessary. Second, designing 3D content with the presented approach requires explicit specification of all of the objects and their properties that need to be presented in the resulting scene. Methods of dynamic semantic composition of 3D content can be proposed to permit implicit conditional query-based assembly of 3D scenes from different reusable components.

A modelling tool supporting the proposed model is planned to be developed. Then performance metrics will be calculated and the model will be validated and evaluated in terms of the efficiency of 3D content creation and encoding. Modelling 3D content by a group of users can allow to verify the approach and to compare the times required for 3D content creation to the times obtained with widely used modelling tools, such as 3ds Max or Blender. The sizes of documents describing conceptual 3D content representations can be compared to equivalent representations encoded in well-established 3D content representation and programming languages, such as X3D or ActionScript.

Other possible directions of future research incorporate several facets. First, automatic transformation of conceptual representations to final 3D content representations encoded in different languages should be implemented. Second, the presented example of semantic modelling assumes a uni-directional transformation of a conceptual scene to a final 3D scene. To permit semantic management and exploration of 3D content in real-time, a persistent link between the primary conceptual objects and the components of the final generated 3D content should be maintained. Third, an additional means can be proposed to represent descriptors that determine the order of the semantic properties included. For instance, the 3D representation of a rotating object that moves may differ from the 3D representation of a moving object that rotates. Finally, the context of user–system interaction (e.g., user location, preferences, client device, etc.) can be considered to enable multi-platform presentations of semantic 3D content on a multitude of available mobile and desktop devices.

References

1. Resource Description Framework (RDF), `http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/` (retrieved August 13, 2013)
2. Resource Description Framework Schema, `http://www.w3.org/TR/2000/CR-rdf-schema-20000327/` (retrieved August 13, 2013)
3. Web Ontology Language Reference `http://www.w3.org/TR/owl-ref/` (retrieved August 13, 2013)
4. OWL 2 Web Ontology Language Profiles, 2nd edn., `http://www.w3.org/TR/owl2-profiles/` (retrieved August 13, 2013)
5. SWRL: A Semantic Web Rule Language Combining OWL and RuleML, `http://www.w3.org/Submission/SWRL/` (retrieved August 13, 2013)
6. Reaction RuleML, `http://ruleml.org/reaction/` (retrieved August 13, 2013)
7. Otto, K.A.: Semantic Virtual Environments. In: Special Interest Tracks and Posters of the 14th Int. Conference on World Wide Web, Chiba, Japan, May 10-14, pp. 1036–1037 (2005)
8. Bilasco, I.M., Villanova-Oliver, M., Gensel, J., Martin, H.: Semantic-based Rules for 3D Scene Adaptation. In: Proceedings of the Twelfth International Conference on 3D Web Technology, Umbria, Italy, April 15-18, pp. 97–100 (2007)
9. Bilasco, I.M., Gensel, J., Villanova-Oliver, M., Martin, H.: On Indexing of 3D Scenes Using MPEG-7. In: Proceedings of the 13th Annual ACM International Conference on Multimedia, Singapore, November 06-12, pp. 471–474 (2005)
10. Bilasco, I.M., Gensel, J., Villanova-Oliver, M., Martin, H.: 3DSEAM: a model for annotating 3D scenes using MPEG-7. In: Proceedings of the Eleventh International Conference on 3D Web Technology, Columbia, MD, USA, April 18-21, pp. 65–74 (2006)
11. Pittarello, F., Faveri, A.: Semantic description of 3D environments: a proposal based on web standards. In: Proceedings of the 11th International Conference on 3D Web Technology, Columbia, USA, April 18-21, pp. 85–95 (2006)
12. Walczak, K.: Flex-VR: Configurable 3D Web Applications. In: Proceedings of the International Conference on Human System Interaction HSI, Kraków, May 25-27, pp. 135–140 (2008) ISBN: 1-4244-1543-8
13. De Troyer, O., Bille, W., Romero, R., Stuer, P.: On Generating Virtual Worlds from Domain Ontologies. In: Proceedings of the 9th International Conference on Multi-Media Modeling, Taipei, Taiwan, pp. 279–294 (2003) ISBN 957-9078-57-2

14. Bille, W., Pellens, B., Kleinermann, F., De Troyer, O.: Intelligent Modelling of Virtual Worlds Using Domain Ontologies. In: Proceedings of the Workshop of Intelligent Computing (WIC), held in conjunction with the MICAI 2004 Conf., Mexico City, pp. 272–279 (2004)
15. De Troyer, O., Kleinermann, F., Pellens, B., Bille, W.: Conceptual modeling for virtual reality. In: Tutorials, Posters, Panels and Industrial Contributions at the 26th International Conference on Conceptual Modeling, vol. 83, pp. 3–18. Australian Computer Society, Inc., Darlinghurst (2007) ISBN: 978-1-920682-64-4
16. Walczak, K.: Beh-VR: Modeling Behavior of Dynamic Virtual Reality Contents. In: Zha, H., Pan, Z., Thwaites, H., Addison, A.C., Forte, M. (eds.) VSMM 2006. LNCS, vol. 4270, pp. 40–51. Springer, Heidelberg (2006)
17. Pellens, B., De Troyer, O., Bille, W., Kleinermann, F., Romero, R.: An Ontology-Driven Approach for Modeling Behavior in Virtual Environments. In: Meersman, R., Tari, Z. (eds.) OTM-WS 2005. LNCS, vol. 3762, pp. 1215–1224. Springer, Heidelberg (2005)
18. Pellens, B., De Troyer, O., Bille, W., Kleinermann, F.: Conceptual Modeling of Object Behavior in a Virtual Environment. In: Proceedings of Virtual Concept 2005, pp. 93–94. Springer, Biarritz (2005) ISBN 2-287-28772-8
19. Pellens, B., Kleinermann, F., De Troyer, O.: A Development Environment using Behavior Patterns to Facilitate Building 3D/VR Applications. In: Proceedings of the 6th Australasian Conference on Interactive Entertainment. ACM, Sydney (2009)
20. Pellens, B., De Troyer, O., Kleinermann, F.: CoDePA: a conceptual design pattern approach to model behavior for X3D worlds. In: Proceedings of the 13th International Symposium on 3D Web Technology, Los Angeles, CA, USA, August 09-10, pp. 91–99 (2008)
21. Zaid, L.A., Kleinermann, F., De Troyer, O.: Applying semantic web technology to feature modeling. In: Proceedings of the 2009 ACM Symposium on Applied Computing, Honolulu, Hawaii, USA, pp. 1252–1256 (2009) ISBN: 978-1-60558-166-8
22. Flotyński, J., Walczak, K.: Semantic Multi-layered Design of Interactive 3D Presentations. In: Proceedings of the Federated Conference on Computer Science and Information Systems, Kraków, Poland, September 8-11 (accepted for publication, 2013)
23. ISO/IEC 14772-1:1997. The Virtual Reality Modeling Language, `http://www.web3d.org/x3d/specifications/` (retrieved August 13, 2013)
24. Extensible 3D (X3D) (2008), `http://web3d.org/x3d/specifications/` (retrieved August 13, 2013)

Creation of Interactive AR Content on Mobile Devices

Dariusz Rumiński and Krzysztof Walczak

Poznań University of Economics,
Niepodległości 10, 61-875 Poznań, Poland
{ruminski,walczak}@kti.ue.poznan.pl
http://www.kti.ue.poznan.pl

Abstract. This paper presents a method of context-based mobile on-site creation of complex interactive augmented reality (AR) content. The method significantly simplifies the content creation process by providing a WYSIWYG authoring mode for AR scenes. In the paper, an AR authoring tool, called MARAT—Mobile Augmented Reality Authoring Tool, is presented. The tool allows users without programming skills to create interactive AR presentations on mobile devices directly on-site, where the AR presentations will be accessed by end-users. The use of the MARAT application for building AR-based virtual museum exhibitions is presented.

Keywords: augmented reality, mixed reality, mobile authoring, virtual exhibitions, MARAT.

1 Introduction

Remarkable progress observed in the last years in the performance of consumer-level hardware has enabled widespread use of interactive 3D multimedia applications. The progress is particularly evident in the domain of mobile devices, such as smartphones and tablets. These devices are nowadays equipped with multi-core processors, large amounts of memory, high-quality displays and multi-modal interaction devices, such as accelerometers, cameras, microphones and GPS sensors. Mobile devices became general-purpose computing platforms well suited for deployment of various kinds of multimedia applications. Moreover, rapid growth in the available bandwidth of wireless networks, which is now sufficient to deliver large amounts of data required by interactive 3D multimedia applications, makes the use of mobile devices for this kind of applications even more appealing.

Multimedia applications that became particularly popular on mobile devices are augmented reality (AR) applications. Augmented reality is a term describing technology that enables superimposing computer-generated content (e.g., interactive 3D virtual objects) in real-time on a view of a real environment. Using a mobile device equipped with a camera, for instance, it is possible to superimpose images of additional objects or additional information on a view of the real world.

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 258–269, 2013.

To enable widespread adoption of AR technologies in various application domains, not only end-users must have simple ways of accessing the AR content on-site, but also the presentation designers must have intuitive, easy to use methods of creating AR presentations. However, most research works in the field of AR have focused on developing AR software without particular emphasis on the content authoring phase, especially in the mobile domain. In most cases, the augmented reality applications are based on content, which is either hard-coded in a concrete application or enable AR authoring only by the use of some editing applications running in the PC environment, which lacks spatial context information. Such methods of creating AR applications are not efficient and not convenient for use in most application domains.

In this paper, we describe a mobile application, called MARAT – Mobile Augmented Reality Authoring Tool, which addresses this issue. On the one hand, the application supports content designers in creating and managing AR presentations in an intuitive way, on mobile devices, directly on-site. On the other hand, the application can be used by end-users to access the AR presentations – to display AR content and to interact with it. We also discuss the use of MARAT for building virtual museum exhibitions.

The reminder of this paper is organized as follows. In Section 2, we start by briefly introducing related works in the domain of authoring AR environments. Later, in Section 3 we describe the ARCO virtual museum system, which is used in connection with MARAT for designing virtual museum exhibitions. In Section 4, we present the goals and requirements of the MARAT application, its implementation, and usage. In this section, we also list events and actions that can be used in the process of creating interactive AR content. Finally, we conclude the paper and indicate some directions for future works.

2 Authoring AR Environments

Authoring AR environments, in addition to the typical difficulties associated with the creation of interactive 3D content, requires additional effort needed to associate the synthetic 3D content with real objects and places. A number of research works have been devoted to the problem of authoring AR environments. For instance, DART – Designer's Augmented Reality Toolkit – is one of well-known authoring toolkits for AR [1]. DART is implemented as an extension to the Adobe Director and is mostly used by professional content producers. DART supports both visual programming and a scripting interface.

Numerous augmented reality applications are based on the ARToolkit library, which requires advanced programming skills and technical knowledge to create applications [2]. GUI-based authoring applications based on ARToolkit that provide user-friendly functionality for placing virtual objects in mixed reality scenes without coding have been presented in [3,4,5,6,7,8]. These solutions, however, target only desktop computer environments.

Another example is the Layar Creator web application, which helps to author multimedia content displayed on marker images [9]. A user needs only to upload to a server an image or a PDF file that will serve as a marker for accessing additional content. The content is accessible in the form of buttons representing special functions, such as accessing a web page or displaying an image or a movie. A user can drag and drop buttons on the marker image and adjust the position and orientation of the displayed AR content. After that, the AR content can be accessed by aiming a mobile device on the marker image.

Only few studies have been performed in the domain of AR content authoring for mobile devices. ComposAR-Mobile is a cross-platform content authoring tool for mobile phones and PCs [10]. It allows users with programming skills to create simple mobile AR applications on a PC. The system consists of two modules – a desktop PC authoring tool based on the ComposAR tool [11] and an AR viewer for both PCs and mobile phones. When a user finishes the authoring process, the first module generates an XML file, which describes the AR scene content. Then, the AR viewer application reads the XML file and renders the AR scene.

Langlotz et al. presented an on-site authoring system for mobile augmented reality [12]. The system targets inexperienced users by permitting creation of content directly in an AR environment. For example, a user can sketch 2D drawings and simple 3D objects using a touch screen. In addition, there is a possibility to create geo-referenced models of real objects, such as buildings. The main drawback of this authoring approach is that the created models are static – a user does not have any possibility to scale, rotate or animate the models.

Stiktu is another mobile AR authoring application [13]. It enables attaching simple content, such as text messages and images, to particular views of real places. In other words, the content, which has been attached to the views "sticks to them". This tool is appropriate for users without programming skills. Everybody, who has a compatible mobile device can create content in this way. The main disadvantage of this application is the lack of a possibility of creating complex objects, such as interactive 3D models.

The Aurasma application enables augmenting arbitrary real-world views using a mobile device's camera [14]. The tool is relatively easy to use and users do not need to have programming skills. First, a user needs to select a digital object from a library, e.g., an image, a video or a 3D model. Next, the user needs to capture the image that will serve as a marker for accessing the digital content. The last step is to position the content by adjusting and rotating the digital object until it gets the most appropriate fit. After saving the setup, a so-called 'Aura' is created. The user can make his/her 'Aura' public, so it becomes publicly available. Aurasma implements also social tools that provide functions that help to share 'Auras' with friends.

In addition to the tools presented above, there are numerous AR SDKs, which can be used to build AR applications on mobile devices. One of the well known platforms is Vuforia [15]. The Vuforia platform, a product of Qualcomm Technologies, uses computer vision based image recognition methods and offers a wide set of features and capabilities to create AR applications. It allows to write native applications for iOS, Android and Unity platforms. Another well-known platform is Wikitude, which consists of a mobile browser application and an SDK [16]. The Wikitude World Browser enables displaying information about points of interest (POIs) in the vicinity of the current user location. POIs are grouped into 'Worlds' which can be assigned to different categories. The browser also serves as a navigation tool to the POIs, which are overlaid on the real-world view. Additionally, Wikitude uses the Vuforia's platform, which gives developers tools for image recognition.

3 Augmented Reality Museum Exhibitions

In the cultural heritage domain, augmented reality technology provides museums with an appealing way of presenting their collections. Interactive 3D virtual and augmented reality exhibitions offer museums and other cultural heritage institutions a convenient alternative method of presenting their existing or reconstructed artefacts. AR enables presentation of objects that otherwise could not be presented, because they are too fragile or too expensive to acquire by the museum, because they were destroyed or lost, or because they are in a distant location. Moreover, augmented reality enables users to interact with the virtual objects, which is usually not possible with real museum objects. The forms of interaction can range from simple object manipulation to complex scenarios engaging users in some educational game plot. Interactive and engaging augmented reality exhibitions may be used to present digitized artefacts, reconstructed interiors, architectural objects and even the whole cities. Museums start to realize the opportunities and many of them invest in 3D digitization of their collections.

ARCO – Augmented Representation of Cultural Objects – is a virtual museum system that consist of a set of tools designed to help museums create, manage and present 3D virtual and augmented reality exhibitions of museum artefacts accessible on-line for use both inside and outside museums [17,18]. The overall architecture of the ARCO system is presented in Figure 1.

The ARCO system consists of three main layers:

- content production,
- content management, and
- content presentation.

The content production layer encompasses all tools and processes required for the creation of digital representations of museum artefacts. Acquisition of 3D shapes can be performed by modelling museum artefacts with a modelling package, such as 3ds max, or by scanning them with the use of a 3D scanning system. The 3D models obtained from the object modelling tools may require further refinement (e.g., reconstruction of missing parts or fixing of polygon meshes), which are carried out using a modelling package.

Digital representations of objects are then stored in the ARCO database and managed using the ARCO Content Management Application (ACMA). Each digitized cultural object is represented as a set of media objects and associated metadata.

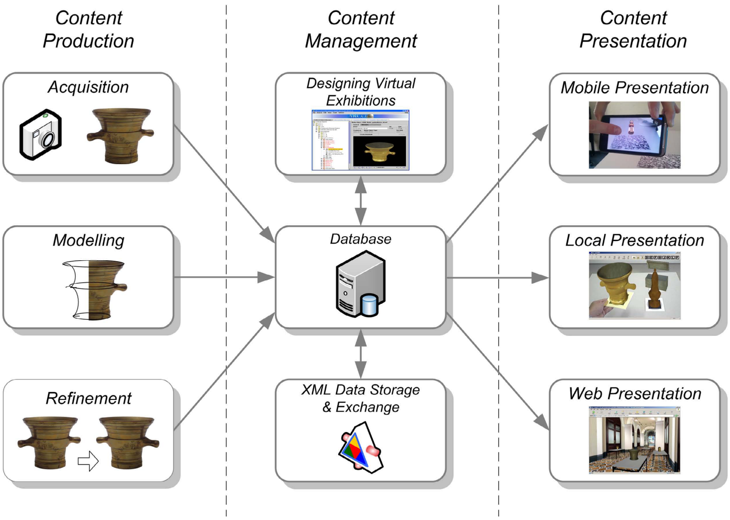


Fig. 1. The overall architecture of the ARCO system

Content presentation in ARCO can be performed in the form of virtual and augmented reality exhibitions accessible both locally—through stationary and mobile devices—and remotely on the web. For content presentation, ARCO uses the concepts of presentation templates, presentation spaces, and presentation domains [18]. The use of presentation templates – content templates and behaviour templates – enables separation of the process of designing and programming complex virtual scenes from the process of setting up actual virtual exhibitions. All the visualization and interaction rules necessary to build virtual exhibitions and most of the graphical properties of the exhibitions are encoded in the presentation templates. A content designer can then create a virtual exhibition simply by building a hierarchy of presentation spaces, assigning cultural objects and presentation templates to the spaces, and setting presentation parameters of the objects and the templates.

Presentation spaces correspond to logical, temporal or spatial parts of a presentation. For example, a presentation space may represent a single 3D exhibition room or a set of objects presented on a single AR marker image.

Presentation domains correspond to different presentation contexts for the virtual exhibitions (e.g., different groups of users or different hardware/software environments). The same presentation space can be presented differently in different presentation domains by the use of different templates. For this purpose, in each presentation space, there may be a different set of presentation templates for each presentation domain.

4 The MARAT Application

4.1 Overview of MARAT

MARAT – Mobile Augmented Reality Authoring Tool – is an easy-to-use mobile authoring application for augmented reality presentations. The application can be used as an extension to the ARCO system described in the previous section.

One of the key requirements for the application was implementation of an authoring method that is easy to use and does not require programming skills, and therefore can be used by people without technical background, such as museum curators or designers responsible for creating and managing virtual exhibitions. The application is intended also for use by end-users, e.g., exhibition visitors, to access the AR content. Typically, users do not have the rights to create or modify AR presentations. These users run MARAT in the read-only mode without access to the functionality that enables modification of the AR content.

A content designer, who is responsible for creating and managing virtual exhibitions, uses the mobile augmented reality application to assign virtual objects to real locations (image markers) and to set their presentation and interaction properties. MARAT enables setting object presentation and interaction properties in a user-friendly interactive WYSIWYG mode. Both the assignments and the properties of objects are stored on the ARCO server and therefore are visible for every user of the MARAT application.

Visitors, equipped with mobile devices with the MARAT application installed, can browse a virtual exhibition. They can view the virtual exhibition through mobile devices equipped with cameras and can interact with the objects. The results of interaction, however, are not stored in the ARCO data repository, and therefore are not persistent.

In the MARAT data model, a presentation space corresponds to a single image marker. The presentation space may contain presentation templates – content templates and behaviour templates – responsible for generating particular presentations of cultural objects, e.g., their appearance and behaviour in the augmented reality scenes. Different presentation domains – which may use different presentation templates for the same set of objects in a presentation space – may be used to differentiate the method of presentation for different users (e.g., exhibition designers vs. visitors) or for different target hardware/software environments (e.g., tablets vs. smartphones).

4.2 Implementation

The MARAT application is based on the Vuforia (Qualcomm) computer vision library [15] to recognize and track planar images and 3D objects in real time. The MARAT application is written in Java and C++ and runs on the Android platform [19]. The architecture of MARAT is presented in Figure 2. The application consists of a number Java modules that are responsible for the application's logic, communication with ARCO servers, and the rendering process. The Java modules are integrated with C++ modules using the JNI – Java Native Interface. The C++ modules implement functions related to image analysis and position tracking.

The MARAT application communicates with the ARCO Application Server using web services, implemented in SOAP, to set and retrieve information about assignment of virtual objects to image markers and presentation properties of objects, and with the ARCO Exhibition Server, using standard HTTP requests, to retrieve the virtual exhibition content.

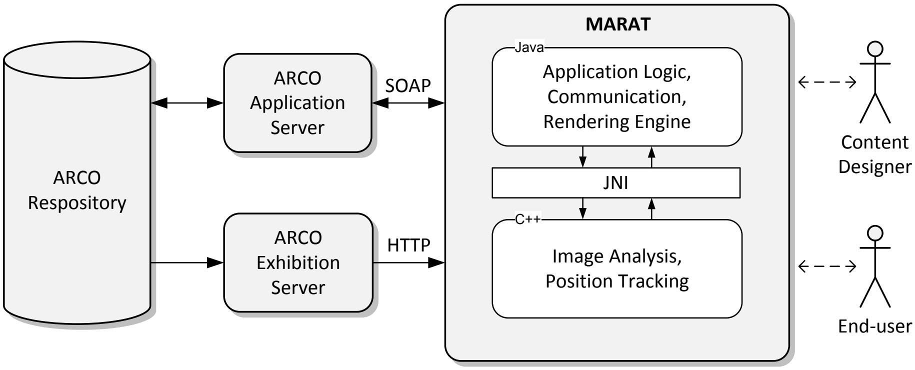


Fig. 2. The architecture of the MARAT application

4.3 Using MARAT to Prepare and Access Virtual Exhibitions

As a first step of creating an augmented reality exhibition using MARAT, the exhibition designer needs to prepare a set of markers. A marker is a planar image, which best should have a complex and sharp pattern. The images may be printed and placed in the physical space, where the exhibition will be accessed. In some cases, pictures of real surfaces in the exhibition space may be also appropriate as markers. The marker image is the basic element, which allows to project a virtual object in a correct perspective to the user's view.

After preparation of the set of markers, the exhibition designer needs to decide on the assignment of cultural objects available in a particular ARCO folder to concrete markers in the exhibition space. In order to do that, the content designer aims a mobile device with MARAT at the chosen marker. When the application recognizes the image, the first unassigned virtual object appears on the screen. The content designer can change it to a different virtual object (Fig. 3). To do that, the exhibition designer can choose to display a list of virtual objects from the menu and can choose another object from the list. Then, the orientation and the scale of the object should be set. Finally, the exhibition designer can save these settings on the remote server by using ARCO web services.

Fig. 3. Assignment of objects to an image marker

The MARAT application allows exhibition designers as well as visitors to interact with virtual objects in an intuitive manner. When a user points the camera of a mobile device towards a registered marker image, the assigned 3D virtual object appears on the screen. The application tracks position and orientation of the marker image in real-time, so that the user's perspective on the virtual object corresponds with his/her perspective on the marker image.

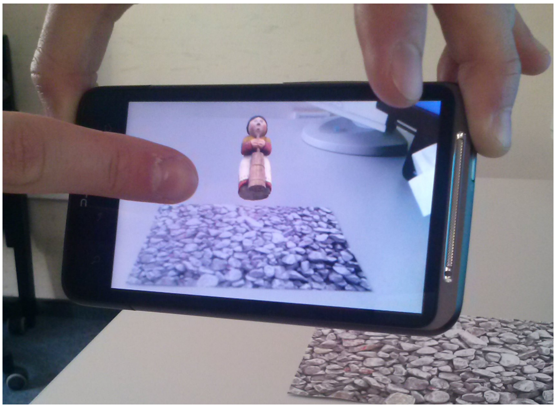


Fig. 4. A pan gesture used to rotate a virtual object

The application recognizes multi-touch gestures to enable users to efficiently interact with 3D models. When a 3D model appears on the display, a user can manipulate the object in a simple way. The user can change object orientation by panning his/her finger on the touch screen. Figure 4 demonstrates how a user can rotate a 3D object in two axes – horizontally and vertically.

When a user moves his/her finger on the touch screen from left to right or opposite, the application recognizes this move and converts the covered distance into an angle. The 3D virtual object is rotated by the calculated angle in the X axis. The application behaves in an analogous way when a user moves his/her finger up or down on the Y axis.

The second gesture that is implemented in the application is the pinch gesture. This type of multi-touch gesture is used to change the scale of a 3D virtual object. To scale an object, the user has to put two fingers onto the touch screen. The change in distance between the fingers is a factor that is used to control the 3D virtual object scale (Fig. 5).

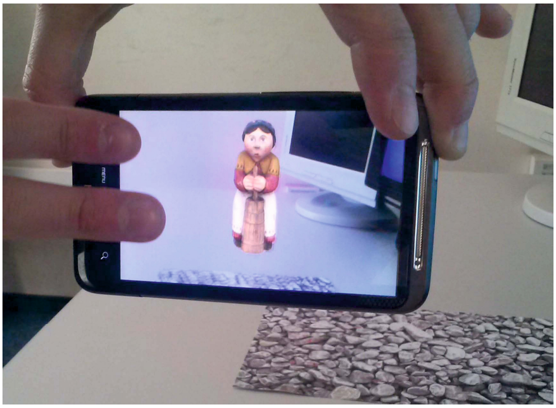


Fig. 5. A pinch gesture used to scale a 3D virtual object

4.4 Designing Interactive AR Content

One of the challenges, when designing complex interactive content through WYSIWYG interfaces without coding, is setting object interaction properties. To solve this problem, in MARAT, a list of predefined events and actions has been programmed. Events indicate specific conditions, such as the beginning or end of the object presentation, a user changing the viewpoint, moving or touching the virtual object, etc. With each event, an action can be associated. Possible actions include displaying a message, playback of some audiovisual content, opening a URL, and switching the presented object to another one. When designing an AR presentation, a designer can indicate active events and set their parameters (e.g., the position of a user which triggers some interaction) and can select actions that are to be performed when the event is registered (e.g., displaying some additional information about the object). Both the event and the action need to be parametrized. For this purpose, MARAT offers easy to use forms specific for events and actions. This method of programming interactions has some limitations, but has the advantage that interactions can be designed without any coding by simply manipulating the mobile device and registering events and actions through a GUI.

The following list presents the names and meaning of events available in the current version of the MARAT application. Events are associated with specific virtual objects.

– **onPresentationBegin** – generated when the application begins to display the virtual object on a marker image. This event allows listeners to respond before the content is shown, e.g., execute the *showQuiz* action when the object is presented.
– **onPresentationEnd** – generated when the application ends to display the virtual object on a marker image.
– **onTouch** – generated when a user clicks on the presented virtual object.
– **onScaleChange** – generated when a user changes the scale of the presented virtual object.
– **onPositionChange** – generated when a user changes position of the presented virtual object.
– **onRotationChange** – generated when a user rotates the presented virtual object.
– **onUserPositionChange** – generated when the position of the mobile device changes.

Below, the list of actions available to an AR content designer is presented. Each action can be associated with a specific event and a specific range of parameter values.

– **replaceObject** – replaces the currently shown object with another one.
– **showMsg** – displays a text message on the screen.
– **showQuiz** – displays a quiz defined in the ARCO database.
– **gotoURL** – opens a web browser and loads the given URL.
– **playVideo** – plays a video sequence.
– **playSound** – plays an audio sequence.
– **playAnimation** – initiates animation within the virtual object.
– **translate** – changes the position of the displayed virtual object.
– **rotate** – changes the orientation of the displayed virtual object.
– **scale** – changes the scale of the displayed virtual object.

Figure 6 shows a user interface that allows a designer to set interaction parameters of an object. For example, the designer can select the *onPresentationBegin* event. Then he or she can choose the *showQuiz* action and can select from a menu an *id* of the quiz. The selected quiz will be shown when the virtual object appears on the screen (Fig. 6, right).

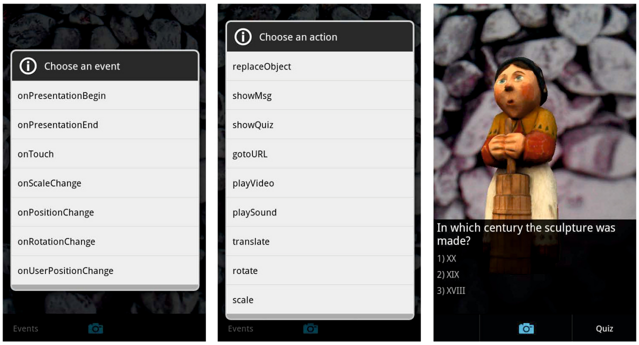


Fig. 6. Setting object interaction parameters

5 Conclusions

In this paper, the Mobile Augmented Reality Authoring Tool has been presented, which allows designers to create interactive AR presentations in a WYSIWYG mode. The initial tests performed on the application prototype indicate that the proposed method of creating AR presentations is promising as it is easy to use and enables creation of content directly on-site, where the AR content will be accessed by users.

Tests carried out on several mobile devices and various 3D objects clearly demonstrate that the currently available processing power of mobile devices is sufficient for displaying complex 3D interactive content. Operations such as switching 3D objects assigned to a marker, changing their orientation and scale are performed very fast – without visible delay. This means that, in a museum context, users will be able to use their mobile devices without requiring the museum to provide specific equipment. The application can run on any number of mobile devices at the same time without the visitors disturbing each other, when they view the virtual exhibition.

The next step of development will be adaptation of MARAT to the iOS system so that it can be accessible to a wider audience. We also plan to focus on development of more advanced method of augmentation, which will omit the process of creation of markers.

Future research works will focus on easy-to-use methods of parameterisation of both the events and the actions to avoid manual entering of parameter values. Methods of combining the described events and actions into more complex interactive scenarios directly within the AR interface will be also considered. A method of presentation of the scenario structure and logic will have to be elaborated to enable meaningful manipulation of the designed scenarios.

Acknowledgements. This research work has been supported by the Polish National Science Centre Grant No. DEC-2012/07/B/ST6/01523.

References

1. MacIntyre, B., Gandy, M., Dow, S.: DART: A Toolkit for Rapid Design Exploration of Augmented Reality Experiences. In: Proc. of ACM Symp. on User Interface Software and Technology (UIST), pp. 197–206 (2004)
2. ARToolKit, `http://www.hitl.washington.edu/artoolkit`
3. Anagnostou, K., Vlamos, P.: Square AR: Using Augmented Reality for urban planning. In: Proc. of the 3rd International Conference on Games and Virtual Worlds for Serious Applications, pp. 128–131 (2011)
4. Lee, G.A., Kim, G.J., Billinghurst, M.: Immersive authoring: What You eXperience Is What You Get (WYXIWYG). Commun. ACM 48(7), 76–81 (2005)
5. Wang, M.-J., Tseng, C.-H., Shen, C.-Y.: An Easy to Use Augmented Reality Authoring Tool for Use in Examination Purpose. In: Forbrig, P., Paternó, F., Mark Pejtersen, A. (eds.) HCIS 2010. IFIP AICT, vol. 332, pp. 285–288. Springer, Heidelberg (2010)
6. Seichter, H., Looser, J., Billinghurst, M.: ComposAR: An Intuitive Tool for Authoring AR Applications. In: Proc. of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR), pp. 147–148 (2008)
7. Grimm, P., Haller, M., Paelke, V., Reinhold, S., Reimann, C., Zauner, R.: AMIRE - authoring mixed reality. In: Proc. of the First IEEE International Augmented Reality Toolkit Workshop, p. 2 (2002)
8. Zauner, J., Haller, M.: Authoring of Mixed Reality Applications Including Multi-Marker Calibration for Mobile Devices. In: Proc. of the 10th Eurographics Symp. Virtual Environments (EGVE), pp. 87–90 (2004)
9. Layar Creator, `http://www.layar.com/blog/2012/06/05/introducing-layar-creator`
10. Wang, Y., Langlotz, T., Billinghurst, M., Bell, T.: An Authoring Tool for Mobile Phone AR Environments. In: Proc. of the New Zealand Computer Science Research Student Conference, pp. 1–4 (2009)
11. Dongpyo, H., Looser, J., Seichter, H., Billinghurst, M., Woontack, W.: A Sensor-Based Interaction for Ubiquitous Virtual Reality Systems. In: Proc. of the International Symposium on Ubiquitous Virtual Reality (ISUVR), pp. 75–78 (2008)
12. Langlotz, T., Mooslechner, S., Zollman, S., Degendorfer, C., Reitmayr, G., Schmalstieg, D.: Sketching up the world: in situ authoring for mobile Augmented Reality. Personal and Ubiquitous Computing 16(6), 623–630 (2012)
13. Stiktu, `http://stiktu.com`
14. Aurasma, `http://www.aurasma.com`
15. Vuforia, `http://www.qualcomm.com/solutions/augmented-reality`
16. Wikitude Developer – Devzone, `http://developer.wikitude.com`
17. Walczak, K., Cellary, W., White, M.: Virtual Museum Exhibitions. IEEE Computer 39(3), 93–95 (2006)
18. Walczak, K.: ARCO: Building Virtual Museum Exhibitions with Flex-VR. In: Styliaras, G., Koukopoulos, D., Lazarinis, F. (eds.) Handbook of Research on Technologies and Cultural Heritage: Applications and Environments, pp. 427–445. IGI Global (2011)
19. Android Platform, `http://www.android.com`

Feature Extraction Based on Semantic Sentiment Analysis

Mohammed Almashraee

Freie University Berlin, Department of Computer Science and Mathematics,
Institute of Computer Science, Koenigin-Luise-Strae 24/26, D-14195 Berlin, Germany
almashraee@inf.fu-berlin.de

Abstract. The abundance of information available on the Web raises the need for techniques that are able to analyze and make a better use of such huge information. Extracting features from unstructured text and assign for each feature its associated sentiment in a clear and efficient way is the goal of this paper. Aspect or feature sentiment analysis is the suitable level of sentiment classification especially for dealing with the domain of products and their related features. We introduce two important and text-related techniques for this purpose namely sentiment classification and semantic Web technology. Sentiment analysis extracts people's opinions in an automatic manner, while semantic Web technology provides a useful improvement by enriching data with additional meaning (semantically annotated) and makes product features better understood by human and automatically processed by machines. For evaluation, we apply our sentiment analysis method twice. Once on a plain corpus, and once again on the same corpus after applying semantic annotation to its data. We also compute the precision and recall principles to evaluate our information extraction method over the corpus in its both cases.

Keywords: opinion mining and sentiment analysis, semantic Web, machine learning, feature extraction, semantic annotation.

1 Introduction

The Web is no longer as it was in the past. Recently, it supports direct interaction between users and its various resources and services such as social media Networks (Twitter, Facebook, LinkedIn, etc.), reviews (movie reviews, product reviews, etc.), blogs, online commercial shops (Amazon, eBay, Google Shopping, etc.) and more. The number of reviews especially for popular products grows rapidly. Peopleś opinions have become an important issue for making decisions, not only for individuals but also for government and commercial sectors. Having such sheer volume of data from different information sources, make it difficult to take a useful and satisfactory decision due to important factors: 1. People cannot read a huge amount of available data, 2. Data on the Web is unstructured, semi-structured and heterogeneous in nature. 3. Information about the same product is often spread over a large number of Web sites. 4. Product features are presented in different formats. 5. Within a single document or a review or within

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 270–277, 2013.

the whole corpus, some products are referred to them using several different names. These factors affect the performance of the resulting output of opinion mining and sentiment analysis concerning the domain of the online products. There are mainly several levels for classifying sentiments. Document level (could be a review, an article, a tweet, etc.), sentence/phrase level, and aspect/feature level. These levels are used according to the domain of interest. In our paper we are using the feature level, since we are collecting opinions about several aspects for the same product and within the same document (definitely from different documents and sources as well). In this paper and among the mentioned problems above, we are proposing a solution to the product feature extraction. Our contributions in this paper is "to annotate our created corpus semantically using existing knowledge bases. This step is useful to define and specify the relations between the product features and to provide a good solution to the problem of feature extraction for sentiment detection."

This paper is organized as follows. In Section 2, we discuss the state of the art and depict the different mechanisms used to extract product features for sentiment analysis. In Section 3, we introduce some basic concepts. In Section 4, we describe the problem statement. The project architecture and the proposed solution are illustrated in Section 5. Evaluation is in Section 6 and finally, the conclusion and future work are presented in Section 7.

2 State of the Art

The extensive volume of electronic information available online draws the attention of researchers to think about useful methods to treat these data in a way that make machines understand and automatically perform useful tasks upon them. This on-line information varies in nature depending on their sources. Essentially they are of three kinds: structured text, semi-structured text, and completely unstructured text. In order to be able to process such text information (mainly semi-structured and unstructured text) we need to apply annotation to our collected data. Opinion mining or sentiment analysis introduced first in [1] is mainly researched in two different approaches. Supervised approach [2–5] and unsupervised approach [6–8]. In both approaches opinions are obtained according to different manipulation levels. Document level [9, 10], sentence level [11–13], and feature level [7, 8, 14]. Pang discusses in [2] the major issues related to social media analytics and the use of different methods concerning machine learning (ML) techniques. He studies the effectiveness of some algorithms from ML techniques in sentiment classifications such as Bayesian models [3] especially the Naive Bayes (NB) method [4], the maximum entropy classification [15], and the support vector machines (SVM) which is first proposed by Vapnik [5]. Liu proposes in [16] a framework to analyze and to compare consumer opinions of competing products; the authors also implement a system prototype called Opinion Observer, which uses supervised rule-discovery techniques to extract product features and their corresponding pros and cons. Another survey is conducted [19] in which the authors discuss and compare the effectiveness of several ML techniques for opinion mining. Their investigation and results indicate that the most commonly used ML mechanisms are SVM and NB and that they outperform several other techniques. Unsupervised approach or so called the semantic approach provides an accurate solution to the sentiment classification task. It digs deeply towards the semantic of the given text. It is also useful in case that there is no much data available for analysis. Turney in [6] introduces an unsupervised algorithm to classify reviews into recommended or not recommended. His algorithm is based mainly on Pointwise Mutual Information and Information retrieval PMI-IR to compute the semantic orientation of the given review. An unsupervised information extraction system called OPINE is introduced by Popescu [17], it uses the relaxation-labeling technique to determine the semantic orientation of opinion words according to the extracted features and to the specific review sentences. The above technology provided a lot of improvement to the problem of collecting peopleś opinions from different information sources. However, all the previously stated work use syntactic process to extract features mainly the linguistic features, e.g., n-grams, parsing to generate part of speech (POS) tags, WordNet, etc. These methods are not good enough at extracting good features, especially when we are dealing with unstructured or even semi-structured data. There is still a need to move the results from the syntactical point of view to the semantic perspective. Ramakrishnan in [18] reported that: features have been defined either manually or by using some special purpose programs so far, and they do not scale well to handle both new domain specific information, and the heterogeneity of the data. Our idea is based on a previous paper presented in [8], in which on-line product features are extracted syntactically to obtain the POS of each word within the sentence and apply for each feature its associated polarity. However, in this paper we apply a more effective approach by combining sentiment analysis with semantic technologies algorithms on product feature extraction and polarity detection.

3 Basic Concepts

The aim of this section is to introduce important concepts that are mentioned in our paper.

3.1 Data Structures and Representations

Data representation is a process of providing a good environment for data to be accessed and manipulated fast. There are different structures:

- Database scheme structure; which called a structured data representation. This kind of structure provides a smooth access and manipulation to the data stored in its scheme, e.g., Oracle, MySQL, etc.
- Semi-structured data; in this type of data representation, we can have direct storage manipulation to data, but limited querying ability, e.g., HTML and XML.

– Semantic representation; this is the newly available and structured data representation. It is usually used by Semantic Web technology applications to interpret their related information and store them in a triple (Subject, Predicate, and Object) format.
– Feature vector representation; this is the most common used representation in which, some extracted features presented as a vector. This representation allow different methods (such as, support vector machines, Nave Bayes, association rule mining, decision trees, hidden Markov models, maximum entropy models, etc.) to build useful models to solve related problems.

3.2 Opinion Mining and Sentiment Analysis

It is a sub field of text mining, which analyzes peopleś opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. Opinion mining drawn the attention of researchers from different areas as illustrated in Figure 1 such as Machine Learning (ML), Information Retrieval (IR), Text Mining (TM), and Natural Language Processing (NLP).

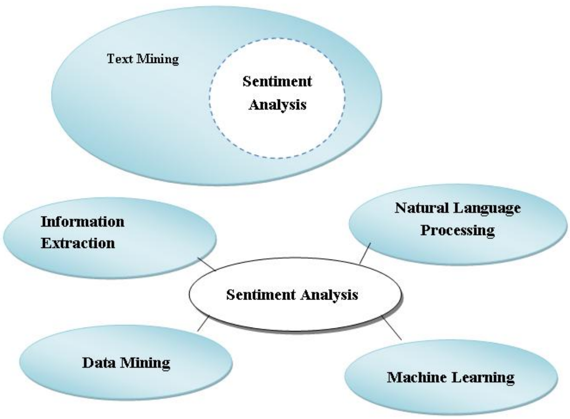


Fig. 1. Opinion Mining and its research areas

3.3 Annotation

Annotation is a process of adding information to already existing data. Annotation can be used, e.g., when there exist a big volume of semi or unstructured data. The task is to convert the unstructured data into a well structured data. There are two types of data annotations. Syntactic annotation, in which information is added according to some lexical tools such as, POS tagging, n-grams, among others. On the other hand, semantic annotation is about providing semantic definitions (add meaning) to the annotated object. Annotations could be done manually which is a time consuming task. Instead, making machines able to automatically annotate data is the best option.

4 Problem Statement

The main research question for this paper is: How to improve the outcome of the opinion mining or sentiment analysis using Semantic Web technology, i.e., how to make data semantically interpretable and hence understandable for machines in an automated way, and how to move the results from the syntactical point of view to the semantic perspective. There are many challenges in opinion mining and sentiment analysis. The challenges for which we are proposing new solutions are as follows:

- Information about the same product is often spread over a large number of Web sites.
- Features are presented in different formats.

In this paper we will consider the above challenges and try to select the required features carefully to convey the required results.

5 Proposed Structure

The proposed structure is divided into two main parts as shown in Figure 2. First part is the corpus annotation, in which we enrich our available domain specific knowledge base (corpus) by adding knowledge that is semantically annotated. The second part is the sentiment classification task, in which product features are extracted and their associated sentiments are obtained.

5.1 Corpus Annotation

This step is used to add meaning to several terms in the corpus using the available knowledge bases (ontology) related to e-commerce. This process will provide a good guide for the sentiment analysis process to get clear explanation about the text in the corpus. One important task before annotation is to apply text preprocessing task to normalize our data in the plain corpus. The annotator then annotate the extracted features that are generated from the dependency parsing algorithm.

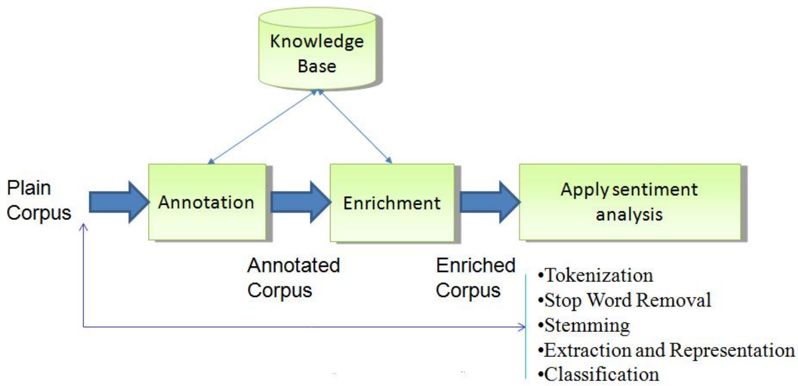


Fig. 2. Semantic Sentiment Analysis Structure

5.2 Sentiment Analysis

A directed graph dependency parsing algorithm is applied for each sentence in our annotated corpus, in which words of the sentence are represented as nodes and the grammatical relations and dependencies between these words are represented as edges. This will give a clear description about the available links between the product features each with its related opinion word and other related terms.

Data Representation. In this stage, we build our corpus by aggregating the most relevant documents to the query. This is done using term frequency - document inverse frequency (tf-idf) weighting scheme as in Equation(1)

$$W_{t,d} = (1 + \log tf_{t,d}) \times \log_{10} \frac{N}{df_t} \tag{1}$$

where W(t,d) represents the weight of in common feature between the document and the query

All document vectors are further normalized since not all the collected documents have an equal length Equation (2)

$$\mid x \mid = \sqrt{\sum_i x_i^2} \tag{2}$$

The most relevant documents are selected according to the similarity measure between the collected documents and the given query. Cosine measure is used as document scoring measure see Equation (3)

$$\cos(q, d) = q.d = \sum_{i=1}^{|v|} q_i d_i \tag{3}$$

Classification. The Naive Bayes (NB) is the most commonly used classifiers for classifying text. NB is based on the Bayes theory in which we have a set of documents $D = d_1, d_2, ..., d_n$ with n the number of documents, and a set of predefined classes $C = c_1, c_2, .., c_m$, with m the number of the categories or the classes, such that

$$P(c_i/d_j) = \frac{P(d_j/c_i) \times P(c_i)}{P(d_j)} \tag{4}$$

where

$$P(c_i/d_j)$$

is the probability that a document exist given a category.

6 Evaluation Methodology

The results will be evaluated by computing the precision and recall. This evaluation is to be performed in two levels. In the first level: the evaluation is applied to the corpus in its two states (plain and annotated) to show the process of semantic annotation. In the second level: we will compare the evaluation results of our method against existing gold-standard e-commerce on-line products.

7 Conclusions and Future Work

Feature extraction is the essential step in sentiment analysis task. In this paper we are applying a novel combination of semantic technologies with sentiment analysis algorithms on product feature extraction and polarity detection. Justifying the research contribution claims of our paper by a proof-of-concept implementation and thorough evaluation is our future work. In addition to that, different algorithms beside classifications are worth for investigation such as association rule mining, clustering, Hidden Markov models, etc. Their results concerning feature generation for sentiment analysis need to be compared. The more efficient method is then adapted to our system.

Acknowledgement. Supervised by Prof. Dr. Adrian Paschke.

References

1. Dave, K., Lawrence, S., Pennock, D.M.: Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: Proceedings of the 12th International Conference on World Wide Web, WWW 2003, pp. 519–528. ACM, New York (2003)
2. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL 2002 Conference on Empirical Methods in Natural Language Processing, EMNLP 2002, vol. 10, pp. 79–86. Association for Computational Linguistics, Stroudsburg (2002)

3. Samsudin, N., Puteh, M., Hamdan, A.: Bess or xbest: Mining the malaysian online reviews. In: 2011 3rd Conference on Data Mining and Optimization (DMO), pp. 38–43 (June 2011)
4. Abney, S.: Part-of-speech tagging and partial parsing. In: Corpus-Based Methods in Language and Speech, pp. 118–136. Kluwer Academic Publishers (1996)
5. Abney, S.: Statistical methods and linguistics. In: The Balancing Act: Combining Symbolic and Statistical Approaches to Language, pp. 1–26. MIT Press (1996)
6. Turney, P.: Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews, pp. 417–424 (2002)
7. Jia, W.-J., Zhang, S., Xia, Y.-J., Zhang, J., Yu, H.: A novel product features categorize method based on twice-clustering. In: 2010 International Conference on Web Information Systems and Mining (WISM), vol. 1, pp. 281–284 (2010)
8. Almashraee, M., Diaz, D.M., Unland, R.: Sentiment classification of on-line products based on machine learning techniques and multi-agent systems technologies. In: Industrial Conference on Data Mining - Workshops (2012)
9. Melville, P., Gryc, W., Lawrence, R.D.: Sentiment analysis of blogs by combining lexical knowledge with text classification. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2009, pp. 1275–1284. ACM, New York (2009)
10. Zhang, Z., Ye, Q., Zhang, Z., Li, Y.: Sentiment classification of internet restaurant reviews written in cantonese. Expert Syst. Appl. 38, 7674–7682 (2011)
11. Hatzivassiloglou, V., Wiebe, J.M.: Effects of adjective orientation and gradability on sentence subjectivity. In: Proceedings of the 18th Conference on Computational Linguistics, COLING 2000, vol. 1, pp. 299–305. Association for Computational Linguistics, Stroudsburg (2000)
12. Riloff, E., Wiebe, J.: Learning extraction patterns for subjective expressions. In: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, EMNLP 2003, pp. 105–112. Association for Computational Linguistics, Stroudsburg (2003)
13. Riloff, E., Patwardhan, S., Wiebe, J.: Feature subsumption for opinion analysis. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, EMNLP 2006, pp. 440–448. Association for Computational Linguistics, Stroudsburg (2006)
14. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2004, pp. 168–177. ACM, New York (2004)
15. Berger, A.L., Della Pietra, V.J., Della Pietra, S.A.: A maximum entropy approach to natural language processing. Comput. Linguist. 22, 39–71 (1996)
16. Liu, B., Hu, M., Cheng, J.: Opinion observer: analyzing and comparing opinions on the web. In: Proceedings of the 14th International Conference on World Wide Web, WWW 2005, pp. 342–351. ACM, New York (2005)
17. Popescu, A.-M., Etzioni, O.: Extracting product features and opinions from reviews. In: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT 2005, pp. 339–346. Association for Computational Linguistics, Stroudsburg (2005)
18. Ramakrishnan, G., Joshi, S., Balakrishnan, S., Srinivasan, A.: Using ilp to construct features for information extraction from semi-structured text. In: Blockeel, H., Ramon, J., Shavlik, J., Tadepalli, P. (eds.) ILP 2007. LNCS (LNAI), vol. 4894, pp. 211–224. Springer, Heidelberg (2008)
19. Khan, K., Baharudin, B.B., Khan, A., e-Malik, F.: Mining opinion from text documents: A survey. In: 3rd IEEE International Conference on Digital Ecosystems and Technologies, DEST 2009, pp. 217–222 (June 2009)

Aspect-Aware Identification of Opinion Phrases Polarity Based on Summaries of Consumer Opinions about Products and Services

Jacek Małyszko

Poznan University of Economics
jacek.malyszko@kie.ue.poznan.pl

Abstract. In this paper, we propose a novel approach to identification of opinion words polarity (so-called polarity induction) for opinion words lexicon extraction. The method employs summaries of many reviews of a single product (or service) as a prediction of polarity of opinion words describing different aspects of the said product. In the article a preliminary experimentation is presented on the basis of which we can expect that the proposed approach can be used in polarity identification.

Keywords: opinion mining, sentiment analysis, polarity induction, natural language processing.

1 Introduction

Consumer reviews are opinions on goods and services, expressed by regular consumers and usually published on the web. Such opinions may be expressed in various ways, for example in a form of the so-called star rating, list of pros and cons or as a free text written in natural language. The goal of opinion mining (also known as sentiment analysis) is to automatically process texts to determine what opinions are expressed in these texts, and thus to overcome the increasing problem of information overload resulting from the fact that the amount of information exceeds processing capacity of decision makers. In this situation, decision makers are persons making decisions concerning purchase of goods or services.

Basic resources used in opinion mining process include opinion word lexicons, that is lists of words and phrases that are used by reviewers in order to express their opinion concerning different products and their aspects, where by aspect we understand components and attributes of these products [1]. Examples of opinion words are *beautiful*, *poor* and *unpredictable*. Opinion words are characterized by so-called semantic orientation, also called polarity, which means that they have either positive or negative meaning in a specific context to which they pertain (optionally with different intensity levels) [1].

Lexicons of opinion words can either be prepared manually or automatically, for example based on corpus analysis and information extraction techniques. Still, in the long term and for extensive applications, manual preparation of

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 278–289, 2013.

such lexicons is unfeasible. The reason for this is the fact that, although there are some words that bear the same polarity regardless the item they describe, such a polarity usually depends on the context in which the given opinion word appears [2]. As an example of such a situation in the paper [1] it has been pointed out that the adjective *unpredictable* is negatively oriented in an automotive review, e.g. in phrase like *unpredictable steering*, but it can have a positive orientation in a movie review, e.g *unpredictable plot*. It is possible not only in texts describing products from different categories, but also in reviews of the same product when a given word is used to describe different aspects of this product (for example the word *low* in such phrases as *low price* and *low screen resolution* in a mobile phone review).

The goal of this paper is to propose a new method for identification of polarity of opinion words in specific contexts which constitutes one of basic tasks of automatic opinion words lexicon extraction. We understand such context as an aspect of product, that the given opinion word describes, as well as category, to which the reviewed product belongs. The method uses information from summaries of easily-processable opinions about products expressed in the form of lists of pros and cons. In this paper we explain in detail the main idea behind the proposed method and present results of preliminary experiments which aimed at statistical verification of whether the proposed idea can be used to solve the specified problem.

The paper is structured as follows: first, a literature analysis is presented to give the reader a broader understanding of opinion words lexicon extraction task; next, in section 3 we describe our approach to opinion words polarity induction. In the Experiments section preliminary statistical verification of the proposed approach is presented. The article ends with a short discussion of directions for future research and a summary.

2 Related Work

As stated in the Introduction, for each category of products and for each aspect of products in a given category opinion words and phrases can have different polarities. Moreover, as products constantly evolve and new features are being introduced, manual maintenance of such lexicons would be a daunting task. Therefore there are attempts to introduce automatic generation of opinion word lexicons. Some extensive state of the art analysis of this area can be found in papers [2] and [3]. In this section we will focus on presenting some of previous work that in our opinion is important to better understand the specified problem.

There are two main types of approaches to opinion words lexicon extraction task [1,4]:

- dictionary-based, where lexicon is generated based on analysis of existing dictionaries and similar lexical resources, such as WordNet (for example papers [5,6]),
- corpus-based, aimed at extracting lexicon from a large corpus of texts with subjective statements.

The accuracy of dictionary-based approach relies heavily on the used dictionary, which currently cannot capture relations between opinion words and different categories and aspects of products. Thus, the only way to learn context-aware opinion words lexicon is to use corpus-based approach [1]. In this paper, we focus on the corpus-based methods.

Generally, context-aware corpus based lexicon generation process can be divided into three separate steps [2]:

1. extraction of phrases used by reviewers to assess different products and their aspects,
2. identification of products and aspects assessed in these phrases,
3. polarity induction, that is assigning polarity to extracted phrases.

Extraction of opinion phrases usually requires some pre-processing, such as tokenization and part of speech tagging as described in [7]. Based on that, certain types of phrases can be extracted from the text, for example based on a simple pattern matching. The idea is based on an observation that often opinion words occur in specific sequences together with other words[8]. For example, in [7] the patterns were based on part of speech tags and any sequence of words with pre-defined tags was extracted (for example an adverb and an adjective).

Identification of what product or service and which aspect of it is described in the given text extract constitutes an intermediate step in aspect-aware lexicon generation process. This can be done based on a lexicon of aspect names where each aspect of the product or service can be referred to using different names (the so called lexicalization of aspects [9]). Such a lexicon also needs to be prepared somehow: either manually (for example [10]) or automatically, for example based on analysis of reviews expressed as list of pros and cons [1]. Yet another task is to recognize to which aspect a given opinion word refers as more than one aspect can be mentioned in a given sentence [11]. Broader analysis of this issues lies beyond the scope of this paper.

Having the phrases extracted and aspects identified, the last step is to determine the semantic orientation, or polarity, of such phrases. Based on that, we can assign different words or phrases to lexicons of corresponding aspects with appropriate polarity, so that we know, if the word is used to express favourable or unfavourable opinions about the aspect. Important pre-processing step is here to handle negations, opinion shifters (for example in phrases *It works* versus *It hardly works*), so-called *but* clauses etc. [1]

Many techniques have been proposed for identification of polarity of words and phrases. Some classical approaches employ statistical associations between words with unknown polarity and words with known positive or negative meaning such as *good* and *bad*. Such statistical association can be understood for example in terms of pointwise mutual information [12]. In another approach, presented in [13], polarity is identified on the basis of analysis of conjunctions as they are assumed to transfer polarity between conjoined adjectives. Here, a seed list of words with known polarity may be used [2]. Yet another approach proposed in [14,15] is based on the observation that in single textual reviews reviewers sometimes use more that one word to assess some aspect and if we know polarity of one of this words, we can transfer it to other words used to assess this aspect.

From the point of view of this paper, the most important type of approaches to polarity detection is to utilize some additional information about the analyzed textual reviews, such as ratings or lists of pros and cons. If the product or service is, apart from textual review, assessed with some additional, easily-processable method (for example star ratings of overall product, as in [3,7]), then such additional information can be used to predict the semantic orientation of words used in the text. In [9] a corpus was used, where star ratings were assigned to different aspects of reviewed entities (in this case restaurants). If it was found that some aspect is assessed in both textual part and star rating in a single review, this rating could be used to determine polarity of the word or phrase used to express opinion on the aspect in the text. Another similar approach is to process short text snippets with known polarity, for example pros and cons given by the reviewers, as in paper [4]. Next, Chi-Square test was used to verify if the probability that the given opinion word appears in positive phrases and negative phrases is equal. If these probabilities proved unequal, positive polarity is assigned if the word appears more often in pros; otherwise negative polarity is assigned.

3 Proposed Approach

The goal of this paper is to propose a new method for polarity induction, which would use additional methods of opinion expression, namely lists pros and cons. In such lists different aspects of products and services may be assessed. Moreover, because such pros and cons usually take the form of very short texts, it is relatively easy to process them [1]. In the proposed work, we focus on lists of pros and cons with predefined options, where for each category of products (such as mobile phones, washing machines etc.) there exists a predefined set of possible pros and cons (a similar method of opinion expression is available in a Polish consumer reviews portal `cokupic.pl`). Such lists are even easier to process, but they still enable users to assess a wide range of product aspects.

The proposed approach is somewhat similar to the method presented in [9]. As it has been stated, in [9] star ratings of predefined aspects were used to assign polarity to opinion words that were used in textual part of the review to assess the same aspects. In case of lists of pros and cons we have only two possibilities: a given aspect is considered to be an advantage or disadvantage of the product or service (optionally, given aspect is not assessed by the given reviewer). With reviews consisting of both a list of pros and cons and of a textual part the main idea is to analyze correlations between easily processable lists of pros and cons and phrases in textual parts of reviews in order to identify polarity of words and phrases used by reviewers to express their opinions about various products and their aspects.

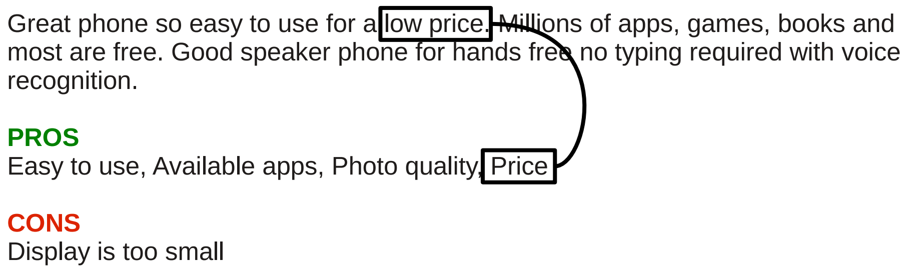


Fig. 1. An example of correlation of information between textual review and list of pros and cons

The simplest way to employ such reviews is to analyze correlations between opinions expressed in free text and in a list of pros and cons at the level of a single review, as presented in picture 1. For example, if in a single review the given aspect in the list of pros and cons was assessed as an advantage of the product, we can expect, that words used in the textual part to assess the same aspect have positive orientation. Such an approach would be analogous to one proposed in [9]. The problem with this approach is that very often the aspects evaluated in these two parts of a review (in textual review and in lists of pros and cons) are different. In [16] we have shown that in a corpus of reviews retrieved from the `cokupic.pl` portal, on average only about 44% of aspects assessed in the textual part of a review is assessed in the list of pros and cons as well, and only 27% of aspects reviewed in lists of pros and cons is described in the textual part at the same time. In other words, information in the part of a review written in free text is often different from the information in list of pros and cons. Regardless of what is the reason for such a difference, it makes the described analysis vulnerable to loss of information which is not included in both parts of a given review simultaneously. Moreover, information from one review cannot be simply used to process other reviews, as discussed in [14].

To overcome this problem, we propose to summarize opinions of many different users about the same product or service to obtain information about an overall view of reviewers on various aspects of the product or service and use such

Table 1. Summary of users opinion on different aspects of Alcatel-ot311 mobile phone based lists of pros and cons retrieved from http://cokupic.pl/produkt/Alcatel-alcatel-ot311

Aspect	**Pros**	**Cons**
ease of use	1	0
functionality	0	16
battery life	3	11
sound quality	1	6
price	12	0

summaries as a prediction of words polarity. Summarizing many reviews of a single product has been proposed for example in [17] and example of such a summary is presented in table 1. We understand such summary in the following way. For each aspect assessed in lists of pros and cons, we count how many users considered this aspect as an advantage and how many as disadvantage of the product. Thus for each aspect we have two numbers, as presented in table 1. To our knowledge, such summaries of reviewers opinion on a product so far have not been used in polarity identification task.

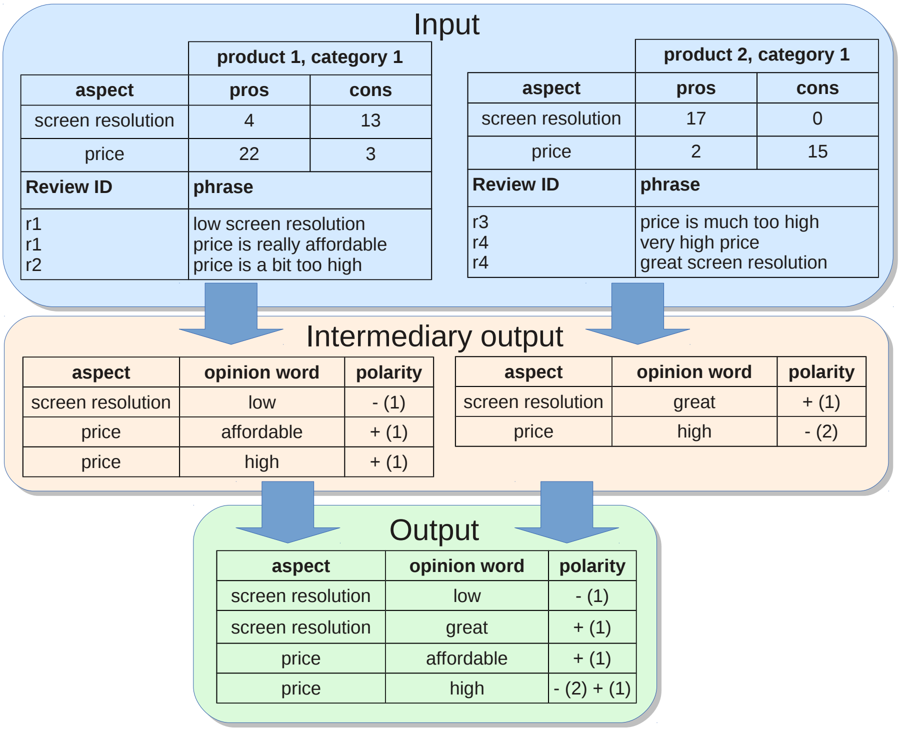


Fig. 2. An overview of the main idea behind proposed approach

The main rationale behind our approach is that if in lists of pros and cons of a certain product or service most reviewers marked a given aspect as an advantage, we can expect that most words describing this aspect in textual reviews also have a positive orientation in the context of this aspect. Picture 2 presents an overview of the proposed approach. Let us assume that we have summaries of opinions about two aspects of two different products as input to our method (but these products are from the same category). The aspects under review are *screen resolution* and *price*. The summaries were generated based on opinions expressed as lists of pros and cons. Apart from that let us assume that using some state-of-the-art method we already have extracted phrases in which these two aspects were assessed from textual reviews of these products (these phrases are presented in the same tables as summaries, in the lower parts of the tables).

In Intermediary Output we have assigned polarity to each phrase based on the values in the summaries of opinions about aspects that are assessed in these phrases. If a given aspect received more pros, we assign + (1) to a given word; if it received more cons, - (1) is assigned. If a given word was extracted more than once for a given product and a given aspect, assigned polarities are added. Please note, that at this stage it happened that we assigned a wrong polarity to one of the opinion words. In one of textual reviews, the reviewer wrote that *price is a bit too high* which clearly is a disadvantage, but because our sole source of information is the summary of opinions about a given product, we assigned + (1) to this opinion word (for this product the *price* aspect was usually assessed as an advantage).

The final step of our approach (the Output field in the picture) is to merge lexicons and polarities of all products from a given category. For each aspect we now have a list of opinion words that were used in all textual reviews of products from this category. Again, polarities are summed up which can help to eliminate errors from the previous step. As we can see, word *high* in the context of *price* has a more negative polarity now, and we can thus treat this word as having negative semantic orientation. We believe that in case of a large corpus such inference would be even more distinct. It is important to note, that categories of products must be fine-grained, because in case of broad categories (e.g. *consumer electronics*) presented approach can give poor results. Still, in case of narrow categories (for example *mobile phones* and *television Sets*) we argue, that the presented approach is justifiable.

4 Experiments

In the previous section, we proposed to use summaries of opinions about products or services generated from multiple reviews expressed in a form of lists of pros and cons in order to predict the polarity of words used to assess these aspects in textual reviews of these entities. The goal of this section is to perform a preliminary analysis of available data in order to determine whether the proposed approach is promising.

4.1 Goal of the Experiment

The goal of the experiment is to verify whether information derived from multiple textual reviews of a given product is similar to the information in summaries of opinions about the same product published in a form of pros and cons. If this is the case we can expect that summaries of lists of pros and cons can be used to predict the polarity of words used to express opinions about aspects in textual reviews of this product.

4.2 Available Data

The experiment was conducted on a corpus of mobile phones reviews retrieved from the `cokupic.pl` website on the 1st of February 2013. In each review in the corpus there were two parts – a textual review and a list of pros and cons. There were 23 predefined pros and 23 predefined cons that users might use to assess the product in a lists.

Table 2. Comparison of summaries of reviewers' opinion in textual form and lists of pros and cons for selected aspects of Alcatel-ot311, based on http://cokupic.pl/produkt/Alcatel-alcatel-ot311

Aspect	**Text**		**Pros/Cons**	
	Pros	**Cons**	**Pros**	**Cons**
ease of use	5	4	1	0
functionality	1	12	0	16
battery	5	6	3	11
sound quality	0	1	1	6
price	3	0	12	0

A sample of 151 reviews of 12 products has been drawn from 77936 reviews of 1858 products (mobile phones) in the corpus. The size of sample was chosen arbitrarily due to the fact that in the experiment analysis of documents from the sample required a significant amount of manual work and 151 was selected as a trade-off between large sample size and feasibility of the task.

Textual parts of these reviews were manually annotated in the following manner: a human annotator was reading textual parts of reviews (lists of pros and cons have been hidden) and was assigning new pros and cons from the predefined list to a review. In other words, the annotator was building a new list of pros and cons based solely on the information which reviewers included in the textual part without knowing what information was included by the reviewers in a list of pros and cons. Thus, for each review we obtained two lists of pros and cons: one made by the reviewer and the other produced by the annotator based on information from the textual review, as presented on table 2.

4.3 Course of the Experiment

Having two summaries for each product, where one summary was based on annotations of textual reviews and the other on lists of pros and cons provided by the actual reviewers, we proceeded to compute the distance between such summaries using a defined distance function $d(s_{i,t}, s_{j,l})$, where $s_{i,t}$ is a summary of opinions about product i generated from textual reviews, and $s_{j,l}$ is a summary of opinions about product j based on lists of pros and cons (where i and j can refer to the same product). After computing the distance between summaries using this function, we obtained some numerical values representing the distance between these two summaries.

To determine the interpretation of such values and check whether summaries of reviews of the same products are indeed similar to each other, we computed distances between summaries of the same products and between summaries of different products. As a result, we obtained a distance matrix for all summaries of textual reviews and all summaries of lists of pros and cons, as presented in table 3. From statistical point of view there were two populations:

1. population of distances between summaries of textual reviews and summaries of lists of pros and cons for the same products (these are diagonal values in table 3);
2. population of distances between summaries for different products (non-diagonal values).

Table 3. Table of distances between summaries from collections T and L

	s_{1L}	s_{2L}	...	s_{nL}
s_{1T}	$d(s_{1T}, s_{1L})$	$d(s_{1T}, s_{2L})$	...	$d(s_{1T}, s_{nL})$
s_{2T}	$d(s_{2T}, s_{1L})$	$d(s_{2T}, s_{2L})$	...	$d(s_{2T}, s_{nL})$
...	...	...	...	...
s_{nT}	$d(s_{nT}, s_{1L})$	$d(s_{nT}, s_{2L})$	...	$d(s_{nT}, s_{nL})$

Next, we analyzed the populations to see if the distances between summaries of the same products tend to be statistically smaller than distances between summaries of different products. We performed such an analysis with two statistical tests; one parametric and one non-parametric:

– test for differences between means of two populations,
– Mann-Whitney U test.

If the tests revealed that distance between summaries of the same products are statistically smaller that in the other population, it would mean that textual reviews and lists of pros and cons of the same products have similar information. Thus, lists of pros and cons can be used to predict polarity of opinion words used to assess aspects in textual reviews.

4.4 Results

First, we conducted test for differences between means of two populations. Let us denote mean of the first population (mean distance between summaries of the same products) as μ_1 and mean of the other one as μ_2. The following hypothesis were tested:

– $H_0 : \mu_1 = \mu_2$ (means for both populations are equal)
– $H_1 : \mu_1 < \mu_2$ (mean for the first population is lower)

Due to small number of elements in the first population, we had to test some additional hypothesis to verify whether the assumptions for the test (concerning the normal distribution in populations and equality of variations) hold. Results of both tests (Kolmogorov-Smifnov and F test for equality of variations) proved that these requirements are met. Next, test for differences between means was performed with Student's T statistics at 0.01 significance level which gave us -2.6 as critical value. The test statistics was -4.68. Thus, the hypothesis about equality of means of both populations must be rejected what means that the mean value of distances between summaries of the same products is statistically lower tl

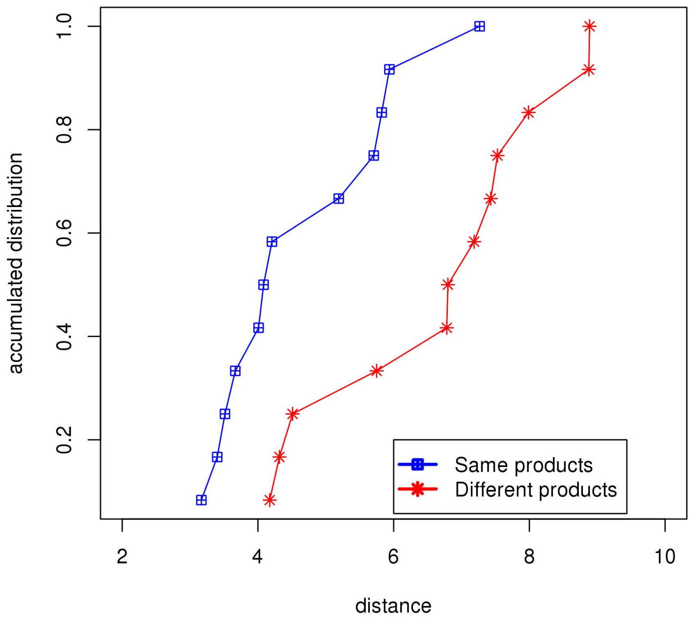


Fig. 3. Comparison of distributions of distances between summaries of the same and different products, where one summary was based on textual reviews and the other on lists of pros and cons

Next we conducted Mann-Whitney U test. The distributions of populations analyzed in the test have been presented in picture 3. Sample size in both populations was 12 (a sample of this size was picked randomly from the population of distances between different products). In the picture it is clearly visible, that the distances between summaries of the same products tend to be smaller than distances between summaries of different products. The test was used to formally confirm this observation. The following hypotheses were tested:

- H_0: in the analyzed populations distributions of distances between summaries of the same products and between summaries of different products do not differ significantly,
- H_1: distances between summaries of the same products are statistically smaller than between summaries of different products.

For sample size of 12 in both populations and for significance level $\alpha = 0.025$ the critical value was 37. The test statistics value was 21. Thus, again the H_0 was rejected, which means that distances between summaries of the same products are smaller than distances between summaries of different products.

4.5 Interpretation of Results

Both test revealed that summaries of textual reviews and summaries of lists of pros and cons for the same products are similar to each other. Based on that we can expect, that easy to generate summaries of lists of pros and cons contain information that can potentially be used to determine polarity of words in textual reviews of these products.

5 Summary and Future Research

In the last section of this paper, we present a summary of the presented idea as well as a plan of the future work.

The goal of our work is to prepare an artifact in the form of a method for solving the problem of context-aware opinion word polarity induction. This method may be used in opinion word lexicon extraction, as part of sentiment analysis task, allowing to solve the problem of information overload faced by consumers trying to make purchase-related decisions based on analysis of consumer reviews published on the Web.

The developed method will be evaluated twofold. First, we will focus on comparing the polarities assigned to opinion words lexicon with the use of the developed method with human annotations to calculate the accuracy of the proposed approach and compare it with selected baseline accuracy. Also, the accuracy of sentiment analysis performed using the generated lexicon will be compared with results of opinion mining conducted according to other selected approach.

If successful, the proposed approach will serve as a basis for the PhD dissertation of the author.

References

1. Liu, B.: Opinion mining and sentiment analysis. In: Web Data Mining. Data-Centric Systems and Applications, pp. 459–526. Springer, Heidelberg (2011)
2. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2), 1–135 (2008)

3. Lu, Y., Castellanos, M., Dayal, U., Zhai, C.: Automatic construction of a context-aware sentiment lexicon: an optimization approach. In: Proceedings of the 20th International Conference on World Wide Web, WWW 2011, pp. 347–356. ACM, New York (2011)
4. Kaji, N., Kitsuregawa, M.: Building lexicon for sentiment analysis from massive collection of html documents. In: EMNLP-CoNLL, pp. 1075–1083 (2007)
5. Godbole, N., Srinivasaiah, M., Skiena, S.: Large-scale sentiment analysis for news and blogs. In: Proceedings of the International Conference on Weblogs and Social Media (ICWSM), pp. 219–222 (2007)
6. Hassan, A., Radev, D.: Identifying text polarity using random walks. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL 2010, pp. 395–403. Association for Computational Linguistics, Stroudsburg (2010)
7. Rill, S., Scheidt, J., Drescher, J., Schütz, O., Reinel, D., Wogenstein, F.: A generic approach to generate opinion lists of phrases for opinion mining applications. In: Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, WISDOM 2012, pp. 7:1–7:8. ACM, New York (2012)
8. Grefenstette, G., Qu, Y., Evans, D., Shanahan, J.: Validating the coverage of lexical resources for affect analysis and automatically classifying new words along semantic axes. In: Shanahan, J., Qu, Y., Wiebe, J. (eds.) Computing Attitude and Affect in Text: Theory and Applications. The Information Retrieval Series, vol. 20, pp. 93–107. Springer, Netherlands (2006)
9. Higashinaka, R., Prasad, R., Walker, M.A.: Learning to generate naturalistic utterances using reviews in spoken dialogue systems. In: Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics. ACL-44, pp. 265–272. Association for Computational Linguistics, Stroudsburg (2006)
10. Lula, P., Wójcik, K.: Sentiment analysis of consumer opinions written in polish. Economics and Mangement 16, 1286–1291 (2011)
11. Jijkoun, V., de Rijke, M., Weerkamp, W.: Generating focused topic-specific sentiment lexicons. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 585–594. Association for Computational Linguistics (2010)
12. Turney, P.D., Littman, M.L.: Measuring praise and criticism: Inference of semantic orientation from association. ACM Trans. Inf. Syst. 21(4), 315–346 (2003)
13. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics, EACL 1997, pp. 174–181. Association for Computational Linguistics, Stroudsburg (1997)
14. Qiu, G., Liu, B., Bu, J., Chen, C.: Expanding domain sentiment lexicon through double propagation. IJCAI 9, 1199–1204 (2009)
15. Qiu, G., Liu, B., Bu, J., Chen, C.: Opinion word expansion and target extraction through double propagation. Computational Linguistics 37(1), 9–27 (2011)
16. Bernatowicz, A., Małyszko, J.: Recenzje konsumenckie w Internecie. Społeczny kontekst publikowania opinii i analiza spójności różnych sposobów ich wyrażania (2013); to be published on Studia Oeconomica Posianiensia Journal
17. Liu, B., Hu, M., Cheng, J.: Opinion observer: analyzing and comparing opinions on the web. In: Proceedings of the 14th International Conference on World Wide Web, WWW 2005, pp. 342–351. ACM, New York (2005)

A Rule-Based Agent Framework for Weakly-Structured Scientific Workflows

Zhili Zhao and Adrian Paschke*

Freie Universität Berlin, Institut for Computer Science,
AG Corporate Semantic Web, Königin-Luise-Str. 24/26, 14195 Berlin, Germany
{zhili,paschke}@inf.fu-berlin.de

Abstract. Instead of focusing on the orchestration of low-level computational intensive tasks in efficiency critical structured scientific processes, as most of available scientific workflow systems do, this work explicitly addresses the coordination and decision-centric higher levels of weakly-structured scientific workflows supporting the choreography of human agents and IT services/agents in their scientific problem solving tasks. From a technical perspective, this work proposes a rule-based agent framework, which exploits the benefits of both the declarative knowledge representation with rules and ontologies and the abstraction with multi-agent technology to support the dynamic and adaptive execution of weakly-structured scientific workflows. Based on the evaluation in terms of the workflow patterns, the rule-based workflow specification in this work demonstrates a high expressiveness level required to represent different types of tasks. The concrete use cases from different scientific domains implemented by the framework also demonstrate that the work fulfills the requirements of weakly-structured scientific workflows.

Keywords: Weakly-structured Scientific Workflows, Declarative Rules, Multi-agent Systems, User Interaction.

1 Problem Statement and Research Question

Scientific workflows have become an important information technology supporting computer-intensive scientific research activities. They assist scientists to perform data management, analysis and simulation of *in silico experiments*[1]. Compared to business workflows, which are already supported by many competing specifications and BPM standards, scientific workflows haven't been widely adopted and supported yet. Among the reasons are the extra requirements of scientific workflows, such as explicit data/information flow, exact reproducibility, agility to quickly adapt to changed knowledge and human/MACHINE decisions, team coordination for distributed problem solving [15]. To address these requirements, the existing workflow technologies in the business domain cannot be easily employed and need to be significantly adapted and extended. Furthermore, the current focus of the existing solutions for business processes as

* Supervisor: Adrian Paschke

[1] http://www.taverna.org.uk/

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 290–301, 2013.

well as scientific workflows is on the orchestrated and pre-structured execution of computational intensive and data-oriented tasks, instead of the goal-oriented and decision-centric tasks that need the coordination of scientists or computer agents (so called expert system agents) as a team supported by semi-automated weakly-structured workflows.

Such weakly-structured workflows are processes, in which there are goal-oriented tasks that require agile runtime decisions during their execution. They may involve the choreography-style coordination of multiple participants to achieve particular goals or have complicated logic to express different policies and cater to dynamic execution environments. They could be modeled at a high abstraction level by graphical representation tools (e.g. BPMN), but the inherent complex and flexible behavior of underlying task executions cannot be easily implemented by the standard workflow execution systems. In current state-of-the-art, partial solutions for some of the aforementioned issues have been proposed, such as: increasing the flexibility of service composition [1,5], incorporating knowledge tasks and objects into workflow models [11]. Nevertheless, many of the core issues in the implementation and execution of such weakly-structured scientific workflows are still largely unsolved. Compared to the more focused computational scientific workflows, the weakly-structured scientific workflows, with their focus on knowledge-intensive agile tasks, additionally require: (1) *Rich Process Representation*: weakly-structured scientific workflows need fleixble task dependencies that support the adaptation in new and exceptional situations. Besides simple process control-flow descriptions (e.g. a task is enabled after the completion of a preceding task), it also needs to represent advanced behavioral logic, which needs dynamic recognition of operational as well as (agent) knowledge-based states to achieve intelligent routings at runtime. (2) *Expressing Domain Knowledge-intensive Decision*: the process execution of weakly-structured scientific workflows often contain the domain decisions, which are often made in terms of knowledge-intensive conditions and may involve multiple sub-decisions. In order to automate the workflows, it is necessary to express these decisions and enable machines to deal with them automatically. (3) *Flexibility*: the workflows are dynamic and need to react to changes according to current circumstances. In other words, a workflow should be allowed to modify according to dynamic situations. (4) *Human Interaction*: many scientific workflow systems are designed to automate scientific processes and improve efficiency. However, humans are still required to perform messy tasks and steer workflow execution to deal with unforeseen problems at runtime. (5) *Exact Reproducibility*: the provenance plays an important role in verification, explanation, reproduction, and informed reuse of data used and produced by scientific workflows, especially by the weakly-structured scientific workflows, which have non-deterministic knowledge-intensive decision logic.

The contribution of this work is in the execution phase of the scientific workflow life cycle [8] with a purpose of explicitly supporting

weakly-structured scientific workflow execution. Several research questions are answered in this work:

1. **How to devise an expressive logical formalism to specify process representations**: The most common rule-based formalisms used to specify workflows employ Event-Condition-Action (ECA) rules to express control flow dependencies. ECA rules are usually defined with a global scope in the knowledge base of a reactive system (e.g. in active databases). However, scientific workflows are usually executed in a certain coordination context, rather than in global event occurrences which trigger global reactions as in ECA rules. Moreover, these rule-based workflow formalisms focus on exploiting the benefits of reactive logic, but do not provide a high expressiveness required for describing the decision logic of weakly-structured scientific workflows.
2. **How to express and enrich complicated domain knowledge-intensive decisions with semantics**: The dynamic decision making in weakly-structured scientific workflows needs higher levels of semantic understanding of decision goals and situation awareness. To do so, declarative representation of the domain decision logic and further semantic enrichment need to be addressed.
3. **How to support adaptive workflow execution**: The adaptability denotes to which extent workflow processes are allowed to be automatically or manually modified according to changed situations at runtime [6]. To achieve this several questions need to be addressed: (1) how to dynamically discover appropriate resources used to perform a task according to current circumstances at runtime; (2) how to implement flexible mechanisms to handle exceptions at runtime, such as: dynamic replacement of an exceptional service; (3) how to perform the goal-oriented tasks, which are done with uncertainties.
4. **How to incorporate user interactions in workflow execution**: Although some scientific workflow systems (e.g. Taverna [10]) can invoke Web services and hence it would be possible to wrap human behavior by Web services, none of existing scientific workflow systems provide features to specify human tasks in workflows [15]. To support the user interaction, firstly it is necessary to provide a well-defined human task definition between a workflow engine and human task owners, who can complete human tasks. Moreover, what users need are not only the support of human task execution, but also the asynchronous interactions with the workflow system, e.g. for long running activities such as discussions or exhaustive knowledge searches.

2 Related Work

Many workflow languages are procedural and describe an executable process involving control flow and data flow among participant services. For example, WS-BPEL allows Web Services to be composed into more sophisticated services using a workflow process. AO4BPEL [3] extends WS-BPEL by increasing the modularity of a system and allowing the separation of crosscutting concerns and changes. YAWL (Yet Another Workflow Language) is sometimes seen as an alternative to BPEL [16]. It supports dynamic adaptation of workflow models through the notion of worklets, i.e. each task is associated with a set of self-contained sub-processes, by which a dynamic runtime selection can be made depending on the context of the particular work instance. However, there are many weakly-structured applications in the real world, especially in the scientific domain, and these block-structured process-oriented languages are not very flexible enough to describe these agile or decision-centric workflows.

For the purpose of supporting a flexible workflow execution, there are research efforts attempting to exploit the benefits of agent systems. Adam Barker et al. [1] capture the service composition with the MultiAgent Protocol (MAP), which allows the typical features of scientific workflow requirements to be understood in terms of pure coordination and to be executed in an agent-based, decentralized, peer-to-peer architecture. Each agent taking part in the interaction adopts a role, by which the agent references a reasoning Web Service that implements all the decision procedures required for that role type. [2] presents an approach to specify a workflow as a multi-agent system, which can intelligently adapt to changing environmental condition. However, the specifications employed in these efforts to specify agent behavior still rely on the procedural languages, e.g. BPEL in [2].

The rule-based approach is another way to support flexible service composition and model the process logic with declarative rule languages. Rule-based languages, especially ECA rules, which provide active real-time or just-in-time reactions to events and are just right for the detection and reaction to events in dynamic computing environments. In [5], Marc Frincu et al. look at scientific workflows from a distributed system perspective and argue that rule-based workflow composition has advantages to handle issues with respect to scalability, failure tolerance, data integrity and scheduling. They give an overview of typical workflow issues and solutions and present a simple ECA rule-based workflow formalism for self-adaptation and auto-generation. [4] also describes an ECA-based workflow management system for the service composition. An automatic event composition algorithm is developed to automate the event processing and validate the manual activities composition at design time. [17] proposes a declarative and rule-driven framework "FARAO" for a dynamic service composition. The "heart-and-soul" of FARAO constitutes business rules that prescribe the way in which services can actually be aggregated into processes. In contrast to these efforts, this work not only provides a flexible workflow description, but also supports the goal-oriented tasks that need to be executed on the basis of dynamic decisions which the agents can take in order to adapt to unforeseen uncertainties at runtime and to dynamically changed (scientific) knowledge about scientific problems.

3 Proposed Solution

In computer science, an agent is a software entity, which is capable of autonomous actions on behalf of a user or other programs to achieve design objects. Besides the autonomy, the agent also can perceive their environments and automatically respond to changes (*reactivity*), start new actions on their own to pursue their own or given objects (*proactivity*), and engage in conversations with other agents in cooperative ways (*social ability*) [19]. On the other hand, the rule-based technology provides a declarative approach to programming and it has the advantages like supporting intuitive understanding by exploiting a limited set of primitives, but with a precise underlying formal semantics, direct support for business and scientific policies, flexibility by alternative execution paths in face of exceptions, adaptability of by easy insertion and retraction of rules and reusability by their property of being isolated from the process context [17]. Moreover, one especially interesting application of the rule-based technology is that it can be directly used as the basis for the agent behavior control. This work proposes a rule-based agent framework, which exploits the benefits of both worlds to support the weakly-structured scientific workflow execution. Figure 1 shows the overall conceptual framework, which is characterized as providing support in two levels for the workflow description:

1. the workflow *definition*: a workflow is described via the composition of a group of abstract tasks, which only describe certain scientific goals and are independent of any specific implementation.
2. the workflow *execution*: during the workflow execution, each abstract task is allocated to an agent, which decides on its own to accomplish the task and then notifies other agents to perform subsequent tasks via sending event messages. Based on the interactions between the agents, they negotiate and collaborate to achieve complex goals.

More precisely, the reactive logic is specified by *reaction rules* and conversation-based *messaging reaction rules*. The *reaction rules* concern the actions in response to events and actionable situations. They specify the conditions under which actions must be taken and describe the effects of action executions [13]. The *messaging reaction rules* concern the message-driven conversations between agents, i.e. sending and receiving event messages associated with the conversation identifiers. Moreover, the *reaction rules* in general are in coupled with *derivation rules*, which specify decision-centric steps represented complex scientific policies. With the combination of these declarative rules and agents, scientific processes are implemented in terms of message-driven conversations and represent their associated interactions via constructs for asynchronously sending and receiving event messages.

For the purpose of enriching the workflow description with semantics, an upper-level ontological workflow metadata model is given to define general concepts (e.g. agent, task, data, role, service) and their relationships, such as: each task is performed by one or more agents, data are the input and output of a

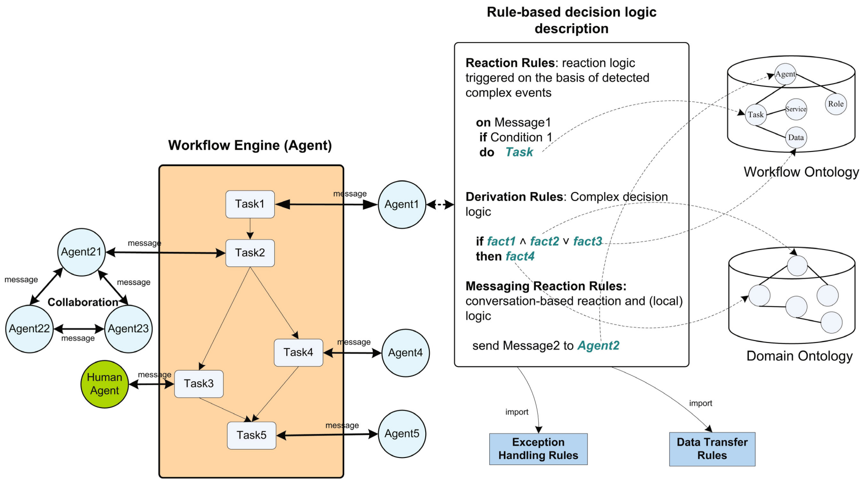


Fig. 1. Rule-based Agent Scientific Workflow Framework

task, every agent plays one or more roles, a role is responsible for one or more tasks, each task uses a service as its realization. These general concepts defined in the upper-level workflow ontology can be further specialized into concrete ones of different scientific domains. With the workflow ontology, it is possible to automatically find alternatives resources (e.g. agents), thereby allowing the workflow execution to proceed even some resources are unavailable at runtime. Moreover, the declarative rules in the work are allowed to integrate external domain ontologies as typed rules to support scientific workflows to be semantically described.

The presented framework expands the range of workflow applications by combining the *orchestration* with a *choreography*-style execution, which describes

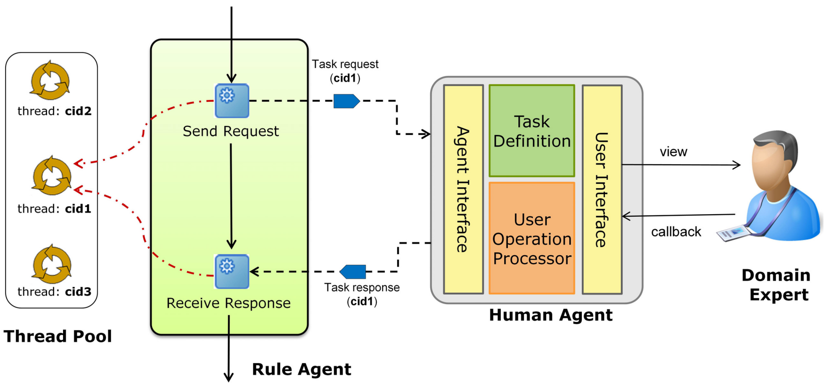


Fig. 2. Integrating Humans into Scientific Workflows

multiparty collaborations and focuses on the message exchange between agents. That is, during the execution of a workflow, a centralized workflow engine (also an agent) takes control a scientific workflow involving a group of tasks, and completes it via the composition of distributed *rule-based agents* (also known as *rule agents* in this work), which are not aware of the whole complex workflow (*orchestration*). For complex goal-oriented tasks, which require the coordination between different participants, the appointed agent is capable of spawning a *choreography* process, which supports the collaborations between agents sharing the same goals.

This work also allows the integration of human roles in weakly-structured scientific workflows. For the tasks that need to be performed by human actors, a human agent manages the life cycle of these human tasks and provides a user interface for scientists to operate on the task, as shown in Figure 2. This work designs a human task specification, which specifies a set of properties of human tasks, such as: task name, parameters, task initiator name, potential owner, actual owner. During the human interaction, the human task specification is sent as the payload of an event message (i.e. a human task request) to the human agent. After completing a human task, the task owners are allowed to call back the workflow engine (i.e. send the task result to the workflow engine) to resume the workflow execution, which is terminated after the human task is outsourced to the human agent. In a workflow, the event messages carrying both a human task request and its results are associated with a same conversation identifier. To support the asynchronous interaction, the activities of sending and receiving messages are processed by the threads taken from a conversation thread pool, which provides a thread according to the conversation identifier passed with a message. This means the workflow engine preserves the current rule context so that when the results of a matching human task arrive, the workflow resumes as though it had never been interrupted. Similarly, the human interaction also contributes to handling unexpected exceptions that might happen during workflow execution (e.g. no resource is available). Once an exception cannot be handled automatically by the agents, the exception will be sent the human agent to ask users to make a decision or complement with the required resources.

4 Research Methodology and Evaluation

This work strictly follows a general design science research methodology [2], which focuses on building and evaluating the utility of the Information System (IS) research artifacts. There are five phases in a design science research effort, i.e. *Awareness of Problem*, *Suggestion*, *Development*, *Evaluation*, and *Conclusion*. After identifying the problems in the scientific workflow domain (Section 1) and proposing an overall conceptual approach (Section 3), this work further implements a design artifact (prototype) to support the weakly-structured scientific workflow execution based on the combination of the declarative programming using rules and the agent technology.

[2] `http://www.desrist.org/desrist/`

The prototype employs a declarative Web rule language Prova[3] to represent the complex logic of weakly-structured scientific workflows as semantic rules. Prova is both a Semantic Web rule language and a high expressive distributed rule engine. On one hand, Prova rule language supports different rule types and provides a highly expressive, hybrid, declarative and compact rule programming language, which combines the declarative programming paradigm, the object-oriented programming and the Semantic Web approach [12]. On the other hand, Prova engine supports complex reaction rule-based workflows, rule-based complex event processing, distributed inference services, rule interchange, rule-based decision logic and dynamic access to external data sources, Web-based services, and Java APIs. For the purpose of seamlessly connecting distributed rule agents together quickly and easily, enabling them to exchange data, a tool for rule-based collaboration Rule Responder[4] is used as a communication middleware. Rule Responder is built on top of the Mule[5]-based Enterprise Service Bus (ESB) middleware for specifying virtual organizations and allows deploying the rule agents as Web-based endpoints in the Mule object broker. The broker object follows the Staged Event Driven Architecture (SEDA) pattern, which decomposes a complex, event-driven application into a set of stages connected by queues and avoids the high overhead associated with thread-based currency models [18]. Benefit from it, Rule Responder provides a highly scalable and efficient application messaging framework to communicate with external services and internal agents.

For the purpose of evaluating the level of solving different types of tasks, the rule-based workflow specification is compared with three scientific workflow management systems: Kepler [7], Taverna [10] and Triana [9] in terms of the workflow patterns[6] proposed by Wil van der Aalst et al. The evaluation is conducted from both control-flow and data perspectives. The initial evaluation results (including the results from [14]) reveal that, besides basic control-flow patterns directly supported by all the solutions, the rule-based workflow specification in this work has superiority in advanced branching and synchronization patterns, state-based patterns and trigger patterns over other three systems. For example, the *Structured Discriminator* (control-flow pattern 9), which requires the decision point to select one branch from two or more branches running in parallel and ignore others. None of Kepler, Taverna and Trianna supports it because that they are not able to reset the join construct when exactly one piece of data is received. Prova well supports this pattern via a reaction group @or, which requires *either* of the event channels to be successfully proved. In addition, with the benefits of the messaging reaction rules to define conversation-based workflow reactive and behavioral logic and the Semantic Web-based types integrated into declarative logic programs, the presented work supports an effective event-driven workflow execution and provides a higher expressiveness regarding to the data patterns than

[3] `https://prova.ws/`
[4] `http://responder.ruleml.org/`
[5] `http://www.mulesoft.org`
[6] `http://www.workflowpatterns.com/`

other systems, in particular the interaction patterns and the data-based routing patterns (e.g. data-based precondition and postcondition validation, event-based task trigger). Due to the space limitation, more details about the evaluation can be found on the project site at: `http://www.corporate-semantic-web.de/`.

The prototype is also evaluated by concrete use cases from different domains. One use case is taken from the EDIT (European Distributed Institute of Taxonomy)[7] to identify newly discovered ants. The process involves three actors: fieldworker, taxonomist and curator. The identification depends on the collaboration of these actors: a fieldworker who often works in the countryside firstly triggers the identification process by describing a newly discovered ant, then sends the description to a taxonomist, who has the experience and expertise to perform the identification. Finally, the taxonomist sends the identification result to the fieldworker and a curator (who is responsible for the ant administration) respectively. Ants differ widely in their food requirements and behaviors, some pests even can cause a serious impact on crops. The identification task involves complicated logic to distinguish an ant from its homogeneous groups. If a discovered ant is extraordinary, it might involve several taxonomists to identify it by their coordination. The other use case is taken from the applications of remote sensing and Geographic Information System (GIS) with a goal of building a snow depth model to predict the snow depth in pastoral areas without field measurements. The process of this experiment involves a task of selecting valid snow samples, which is decision-centric with many criteria. For example, in an experiment of in the pastoral area of Northern Xinjiang (in China) [20], the criteria of selecting valid snow samples are: the depth of the valid samples must be more than 3.0 centimeters, got from the same area as the satellite data at an elevation lower than 2000 meters, collected from an area with a local temperature less than 6 ℃, must be dry snow, collected in March, etc. The implementations of these use cases show that this work provides an expressive workflow description to weakly-structured scientific workflows, and it not only supports an adaptive execution of these workflows, but also supports the asynchronous user interaction at runtime.

5 Contributions

For the purpose of supporting the weakly-structured scientific workflow execution, this work proposes a rule-based agent framework, which exploits the benefits of both the declarative programming with rules and agent technology. In particular, the major contributions of this work are as follows:

1. *An upper-level workflow metamodel.* This work develops an upper-level ontological workflow metamodel and endows the typical concepts used in workflows with precise meanings. To carry out real concrete processes, the upper-level workflow ontology can be specialized into concrete domain ontologies, which facilitate agents to dynamically find appropriate or alternative resources at runtime.

[7] `http://www.e-taxonomy.eu/`

2. *A rule-based formalism for describing workflows.* From a technical perspective, this work provides a declarative rule-based workflow specification, which combines reaction and derivation rules to describe complex reactive and decision logic of weakly-structured scientific workflows. Based on message-driven conversations, it is possible to describe scientific processes that span several communicating agents.
3. *Domain knowledge-intensive decision expression combining logic programs with description logic.* This work gives a hybrid approach, which exploits the benefits of both Description Logic (DL) and Logic Programming (LP) to express complicated domain knowledge-intensive decisions. In other words, they are encoded as declarative rules and can incorporate external domain specific ontologies, which give domain semantics or even pragmatic meanings to used concepts, as typed rules.
4. *Adaptive workflow execution.* The declarative rules specify agent behavior and enable them to dynamically replace an exceptional service by reasoning the workflow ontology. The agent can choose an appropriate service or invoke a sub-process to perform a task according to the current circumstance. Moreover, a goal-oriented task can be allocated to a group of agents sharing the same goal, which complete the task via their coordination and collaboration.
5. *Detached user interaction.* This work supports two major activities which require user interactions: one is the human tasks which are performed manually. The other is the unexpected exceptions that cannot be handled by a workflow system. A human agent manages the life cycle of user interactions and provides a graphical user interface for users to operate on these activities.

Compared to the existing solutions, the presented work explicitly considers weakly-structured scientific workflows from a technical perspective, and highlights the following major aspects:

1. *Abstraction* via a distributed multi-agent model reflecting the semiotic structure of scientific teams in distributed choreography style workflow execution and distributed problem-solving.
2. *Complex decision logic* via derivation rules and logical inference deductions beyond the typical restricted expressiveness of simple gateways in process execution models.
3. *Situation-awareness and behavioral dynamic reactions* (workflow patterns) via reaction rules leading to dynamic and agile workflow reaction patterns.
4. *Semantic workflow execution* via domain models and information models represented as ontologies which are integrated into semantically typed rules.
5. *Decoupled* via event messages enabling asynchronous communication and parallel processing, non-deterministic execution branches of problem solving tasks in distributed agents.
6. *Expanding the range of workflow applications* via combining the benefits of both orchestration and choreography, i.e. maintaining the execution of a workflow in a centralized way, while at the same time scaling up to integrate

choreography sub-processes that are modeled by a group of collaborative agents sharing a same goal to perform goal-oriented tasks.
7. *Asynchronous user interaction* via the decoupled communication allowing users to operate on human tasks or unexpected exceptions and call back the workflow engine asynchronously.

6 Conclusion and Progress

The presented work explicitly focuses on the real-world weakly-structured workflows, and proposes a rule-based agent framework, which combines the declarative programming using rules with the agent technologies to support them. With the benefits of the declarative rules and the agent technology, the framework supports the dynamic and adaptive workflow execution from different aspects. In addition, this work also integrates human behavior into the workflow execution. In contrast to three prominent scientific workflow systems in terms of the workflow patterns, the rule-based workflow specification in this work provides a higher expressiveness to support advanced workflow patterns and fulfills the requirements of weakly-structured scientific workflows.

This PhD work is currently ongoing. The motivation and problem statement have been defined. The overall conceptual approach was also proposed as described in this paper. The thesis is currently in the process of a formal description of the conceptual approach. Afterwards, a thorough evaluation of the approach will be performed. The intended final submission of the thesis is in the spring of 2014 and the further progresses will be reported on the project site at: `http://www.corporate-semantic-web.de/`.

References

1. Barker, A., Mann, R.G.: Flexible Service Composition. In: Klusch, M., Rovatsos, M., Payne, T.R. (eds.) CIA 2006. LNCS (LNAI), vol. 4149, pp. 446–460. Springer, Heidelberg (2006)
2. Buhler, P.A., Vidal, J.M.: Adaptive Workflow = Web Services + Agents. In: Proceedings of the International Conference on Web Services, pp. 131–137. CSREA Press (2003)
3. Charfi, A., Mezini, M.: AO4BPEL: An Aspect-oriented Extension to BPEL. In: World Wide Web, pp. 309–344 (2007)
4. Chen, L., Li, M., Cao, J.: A Rule-Based Workflow Approach for Service Composition. In: Pan, Y., Chen, D.-x., Guo, M., Cao, J., Dongarra, J. (eds.) ISPA 2005. LNCS, vol. 3758, pp. 1036–1046. Springer, Heidelberg (2005)
5. Frincu, M., Craciun, C.: Dynamic and Adaptive Rule-Based Workflow Engine for Scientific Problems in Distributed Environments, ch. 10, pp. 227–251. CRC Press (2010)
6. Görlach, K., Sonntag, M., Karastoyanova, D., Leymann, F., Reiter, M.: Conventional Workflow Technology for Scientific Simulation. In: Guide to e-Science, pp. 323–352. Springer, März (2011)

7. Ludäscher, B., Altintas, I., Berkley, C., Higgins, D., Jaeger, E., Jones, M., Lee, E.A., Tao, J., Zhao, Y.: Scientific Workflow Management and The Kepler System, vol. 18, pp. 1039–1065 (2006)
8. Ludäscher, B., Weske, M., McPhillips, T., Bowers, S.: Scientific Workflows: Business as Usual? In: Dayal, U., Eder, J., Koehler, J., Reijers, H.A. (eds.) BPM 2009. LNCS, vol. 5701, pp. 31–47. Springer, Heidelberg (2009)
9. Majithia, S., Shields, M., Taylor, I., Wang, I.: Triana: A Graphical Web Service Composition and Execution Toolkit. In: Proceedings of the IEEE International Conference on Web Services, ICWS 2004, p. 514. IEEE Computer Society, Washington, DC (2004)
10. Oinn, T., Addis, M., Ferris, J., Marvin, D., Senger, M., Greenwood, M., Carver, T., Glover, K., Pocock, M.R., Wipat, A., Li, P.: Taverna: A Tool for The Composition and Enactment of Bioinformatics Workflows. Bioinformatics, 3045–3054 (November 2004)
11. Papavassiliou, G., Mentzas, G., Abecker, A.: Integrating Knowledge Modelling In Business Process Management. In: ECIS, pp. 851–861 (2002)
12. Paschke, A.: Rule Responder HCLS eScience Infrastructure. In: Proceedings of the 3rd International Conference on the Pragmatic Web: Innovating the Interactive Society, ICPW 2008, pp. 59–67. ACM, New York (2008)
13. Paschke, A., Boley, H., Zhao, Z., Teymourian, K., Athan, T.: Reaction RuleML 1.0: Standardized Semantic Reaction Rules. In: Bikakis, A., Giurca, A. (eds.) RuleML 2012. LNCS, vol. 7438, pp. 100–119. Springer, Heidelberg (2012)
14. Sara, M., Mauro, G., La Rosa, M., ter Hofstede, A.H.M.: Pattern-Based Evaluation of Scientific Workflow Management Systems. Technical report, University of Verona (2011)
15. Sonntag, M., Karastoyanova, D., Deelman, E.: Bridging The Gap Between Business And Scientific Workflows. In: Proceedings of the IEEE 6th International Conference on e-Science, Brisbane, Australia, December 7-10, pp. 206–213. IEEE Computer Society (December 2010)
16. van der Aalst, W.M.P., ter Hofstede, A.H.M.: YAWL: Yet Another Workflow Language. In: Information Systems, pp. 245–275 (2005)
17. Weigand, H., Van Den Heuvel, W.-J., Hiel, M.: Rule-Based Service Composition and Service-Oriented Business Rule Management. In: In Proceedings of the International Workshop on Regulations Modelling and Deployment, pp. 1–12. Citeseer (2008)
18. Welsh, M., Culler, D.E., Brewer, E.A.: SEDA: An Architecture for Well-Conditioned, Scalable Internet Services. In: SOSP, pp. 230–243
19. Woolridge, M., Wooldridge, M.J.: Introduction to Multiagent Systems. John Wiley & Sons, Inc., New York (2001)
20. Yu, H., Feng, Q., Zhang, X., Huang, X., Liang, T.: An Approach for Monitoring Snow Depth Based on AMSR-E Data in The Pastoral Area of Northern Xinjiang. Acta Prataculturae Sinica, 210–216 (2009)

Process Mining Techniques in Conformance Testing of Inventory Processes: An Industrial Application

Zbigniew Paszkiewicz

Department of Information Technology, Poznań University of Economics,
al. Niepodległości 10, 61-875 Poznań, Poland
zpasz@kti.ue.poznan.pl

Abstract. Contemporary business information systems record business events in event logs. Process mining techniques permit to explore these logs to understand and analyze processes. Although, there are a number of case studies demonstrating usefulness of process mining techniques in real-life case studies, process conformance check techniques miss such practical evaluation. Also the commercial tools have not been tested for complementarity with the widely used academic tool ProM Framework in context of conformance check. In this paper an application of process conformance check techniques in a production company is described. Using a variety of techniques, an inventory processes supported by Warehouse Management System (WMS) have been analyzed. In particular, special attention has been put on *Product Management* process. The goal of this paper is to demonstrate the applicability of conformance check techniques to quality management in a production company. Both academic and commercial tools are used and evaluated in terms of efficiency and complementarity with each other.

Keywords: process mining, business process intelligence, inventory management, quality management, warehouse management system.

1 Introduction

Modern organization is interested in information about conformance of its processes to rules that should be observed. Those rules follow from law, industry standards or internal regulations. Frequently, conformance of processes to rules is tested by analysis of organizational processes. Any identified deviations from assumed rules are a potential source of legal, financial or operational risk.

Conformance testing is performed periodically in organizations. Typically used methods are based on financial document analysis, workshops and employee interviews. Such analysis needs long time and requires a lot of resources. As a consequence, in big organizations only partial information concerning organization processes is analyzed. Conclusions derived from partial information are then generalized using statistical methods. If the analysis of partial information uncovers no deviations, it is assumed that all the operations performed in the whole organization conform to rules. Great examples of such analysis are audits [1].

W. Abramowicz (Ed.): BIS 2013 Workshop, LNBIP 160, pp. 302–313, 2013.

In many cases this approach is insufficient. In global economy, turbulent organization environment strongly influences organization's operation requiring fast adaptation to new operation rules. Modern approaches to management require new methods of conformance analysis that are faster, cheaper, free of human judgment, and based on possibly complete set of information.

Proliferation of information technology is followed by increased number of processes that are performed by electronic means. The concept of *Process Aware Information Systems (PAISs)* has been proposed in [2]. A PAIS is *"a software system that manages and executes operational processes involving people, applications, and/or information sources on the basis of process models"* [2]. Examples of PAISs are: workflow management systems, Customer Relationship Management (CRM) systems, Warehouse Management Systems (WMS). Modern PAISs log enormous numbers of events providing detailed information about the activities that have been executed. Events are stored in repositories called *event logs*. Event logs are a basis for process mining. The term process mining is used to describe *"techniques, tools, and methods to discover, monitor and improve real processes by extracting knowledge from event logs commonly available in today's information systems"* [3].

Exploration of event log aims at discovering process models describing actually executed process instances and facts associated with those models [7, 8]. Among others, process mining techniques include process conformance testing. *Conformance check* techniques aim at automatic comparison of envisioned rules or models with behavior recorded in an event log to identify discrepancies [2, 9].

Two types of process models and two types of data are distinguished and used in process mining methods. A *de jure* process model is normative, i.e., it specifies how things should be done or handled. For example, a process model used to configure a BPM system is normative and forces people to work in a particular way [3]. A *de facto* model is descriptive. Its goal is not to steer or control reality. Instead, *de facto* models aim to capture reality.

Application of basic process mining techniques to real-life data and business problems is well documented [1, 4, 5]. The study concentrating on conformance check techniques still miss wide practical evaluation. For example, they raise questions concerning their effectiveness and usability. Moreover, no study concerning tool support for conformance check is available. Therefore, it is important to confront existing techniques and tools supporting conformance checking with real-life questions and event logs taken from real-life systems.

In this paper a case study based on a log of inventory processes is presented. Inventory processed were performed in a Polish company producing mattresses. The company employs in all the departments about five hundred people. The event log was generated by WMS supporting company operation which is not a workflow system. The event log used in the study described in this paper includes 781 134 events recorded from January 1st, 2013 to May 25th, 2013. The log contains information about 126 832 process instances. The analysis of the event log aimed at answering the questions raised by warehouse managers concerning organization operation. In this paper, only questions concerning conformance testing of the *Product Management* process are discussed. Open source ProM Framework [6] and commercial Fluxicon Disco (htttp://www.fluxicon.com/disco) process mining tools have been used to conduct the analysis. ProM Framework was selected because it provides the functionality allowing the wide spectrum of analysis. Commercial tools in genera offer very similar functionality and Disco is their great representative well-known from its effective implementation of algorithms.

The remainder of this paper is organized as follows. The basic facts concerning the case study are presented in Section 2. In particular business goals of the analysis and basic facts concerning the event log are outlined. In Section 3, the case study is described in detail. In Section 4, process mining tools are discussed. Finally, Section 5 concludes the paper.

2 Understanding Inventory Processes

2.1 Process Scope

There is number of processes performed in a warehouse including: *Material Receiving* process, *Product Shipping* process, *Product Management* process, and *Material Management* process. Due to a limited scope of this paper and for the sake of simplicity, only *Product Management* process is analyzed. Only a part of information from the event log is associated with this process.

The *Product Management* process encompasses activities required to take the product, i.e., mattress, from the production line and ship it to company's client. Products waiting for shipment are stored in the warehouse. Products are organized in pallets which are the smallest shipment and storage units. The products are categorized into *families*, which are understood as mattress types. There are twenty various product families. The transport of pallets among production lines, storage areas and shipment areas is done by storekeepers. Storekeepers work 24 hours per day on three shifts except weekends.

The process involves the following types of activities that are performed:

- *Production* – refers to a storekeeper activity of taking a pallet from production line;
- *Rest* – refers to a storekeeper activity of putting a pallet in the storage area;
- *Shipment approved* – refers to a storekeeper activity of preparing a pallet to shipping by putting it in a shipment area;
- *On fork* – refers to a storekeeper activity of transporting a pallet within the warehouse;
- *Shipped* – refers to an activity of actually shipping a pallet form a warehouse; obviously the shipping of an article is a process itself but WMS records it as a single activity and it is analyzed this way;
- *Deleted* – refers to an activity of removing a pallet from evidence.

Each activity performed by a storekeeper is recorded in the WMS. Before performing any activity, a storekeeper is obliged to scan a bar code available on every pallet. The WMS keeps track of a pallet life cycle and associates each scanning with appropriate activity. For instance, once a pallet is scanned after production (*Production* activity), the next recorded activity must be *On fork* activity and *Rest* activity. By choosing an option *"Start shipment"* in the WMS user panel, a storekeeper has a possibility to perform *Shipment approved* and *Shipped* activities. *Deleted* activity is performed only in exceptional situations.

There is one *de jure* model describing the life cycle of each pallet. In Fig. 1 *de jure* model is presented as a Petri net. The model assumes sequential execution of activities. When a pallet is transported within a warehouse multiple times the subsequence *On fork* > *Rest* can be performed multiple times.

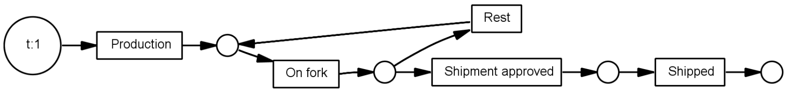


Fig. 1. Petri Net describing desired pallet life-cycle

2.2 Conformance Rules

The following aspects of conformance of the *Product Management* process have been identified by interviewing the *Product Management* process owner:

- *Conformance to model* – process instances must follow *de jure* model presented in Fig. 1;
- *First In – First Out policy* – products that were produced first must be shipped first; the FIFO rule must be satisfied for particular mattress family and size; to conform to FIFO rule, storekeepers must follow WMS recommendations concerning which pallets must be handled as the first; WMS does not recommend a particular pallet but a stand where the oldest pallet is located; the decision of which pallet to take from the stand is up to a storekeeper;
- *Quality assurance* – all the pallets shipped to a client must be checked by a quality department;
- *Process performance* – a particular pallet cannot be stored in a warehouse for more that fourteen days;
- *Pallet damage handling* – when a pallet is in disrepair, it must be transported to a special storage area; all the storekeeper are responsible for handling damaged pallets in this way;
- *Work distribution* – all the three shifts should perform equal amount of work; storekeepers are divided into two groups: (1) taking pallets from production lines, (2) shipping pallets from a warehouse; storekeepers from one group should not be involved in activities of the other.

2.3 Basic Facts

The WMS stores information concerning activities in a database. The following information is associated with each activity in the WMS event log: activity name, activity timestamp, name of a storekeeper executing the activity, identifier of a pallet being a subject of the activity, mattress family, warehouse name, optional stakeholder comment, optional pallet description. Information available only for the subset of activities include: storage area code (*Rest* activity), shipment area code (*Shipment approved*), recommended storage area code (*Rest*), information whether the recommendation was followed by a stakeholder (*Rest*). To validate all the rules defined in Section 3.2, the following attributes were derived from other data available in the WMS database: stakeholder shift, information if a pallet was damaged, information if a pallet was approved in terms of quality. The maximum number of attributes associated with a particular activity is sixteen. The pallet identifier is used to group activity instances into process instances.

Product Management process analysis is performed on basis of captured 554 745 events associated with 87 660 instances. This is a subset of information stored in the event log (*cf.* Section 1). 81% of process instances required less that 7 activities to be executed. The number of process instances performed for the most popular mattress family is 17 749. Execution of all the recorded *Product Management* process instances involved 55 persons.

3 Mining Process Aspects

In this section, conformance of process instances to rules defined in Section 3.2 is tested using process mining techniques.

3.1 Conformance to Model

In Fig 2, *de facto* model discovered from the event log generated by Fluxicon Disco is presented. Numbers assigned to activities and transitions indicate the number of process instances they appeared in the log.

De facto model presented in Fig 2 indicates that the execution of the process is far more complex than assumed in the *de jure* model (*cf.* Fig. 1). In particular, many additional transitions among various activities appear which are not included in the *de jure* model, e.g. transition from *Shipped* activity to *Rest* activity, self-loops for activities *On fork*, *Production*, *Shipment approved*, *Shipped* and *Rest*. *De facto* model indicates that not only *Shipped* activity, but also *Rest*, On *fork* and *Deleted* activities are closing the process. Finishing *Rest* and *On fork* activities correspond to process instances that are still running.

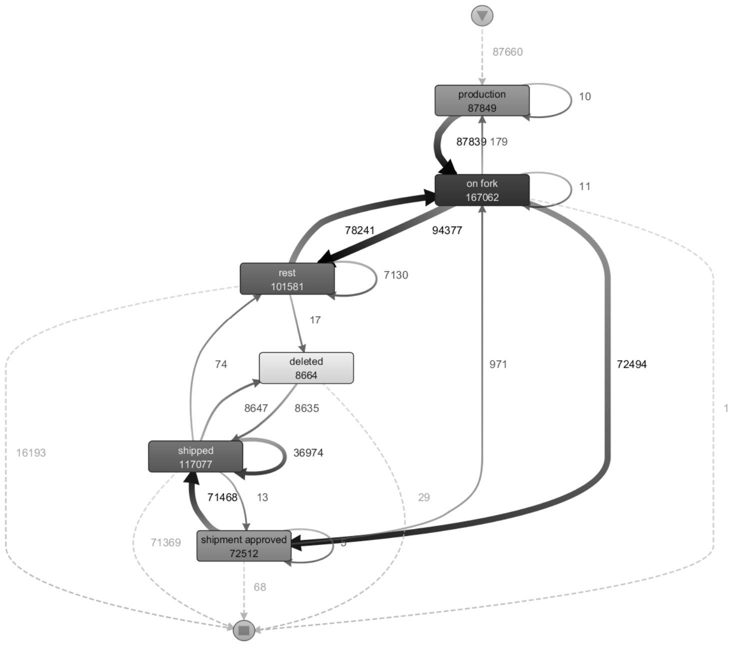


Fig. 2. Discovered model of the Product Management process

The most frequently executed activities and transitions marked with thick arrows and dark rectangles fit to the *de facto* model almost perfectly. Thus, in general, the process conforms to the model but some deviations must be investigated more closely. The result of automatic comparison of the event log behavior with the *de jure* model is presented in Fig 3. The validation was performed using *Conformance Checker* plugin [10] of the ProM Framework. The validation was performed for 1068 process instances executed in May 2013.

The quality of the *de facto* model is measured by fitness measure. *Fitness measure* estimates to which extent a model allows for the behaviour perceived in the event log. In this approach the *de jure* model is represented by the Petri net (*cf.* Fig. 1). In particular, fitness measure relates the number of missing tokens with the number of consumed ones and the number of remaining tokens with the number of produced ones [10]. If the model fully supports the behaviour seen in the event log, i.e., no tokens are missing nor remaining, the fitness is evaluated to 1. In the worst case, every produced and consumed token is remaining or missing, the metric evaluates to 0.

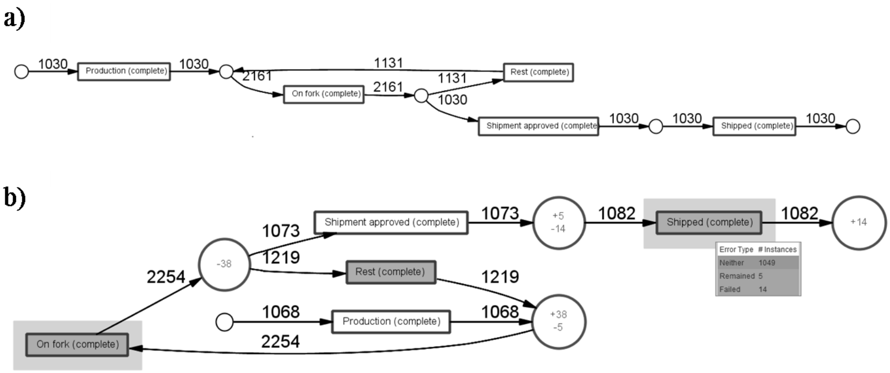


Fig. 3. Automatic conformance checker

The results of the analysis of the subset of 1030 process instances is presented in Fig 3a. Execution of these 1030 process instances is fully supported by the *de jure* model, so the fitness value is estimated to 1. Fig 3b includes the conformance testing of the full set of 1068 process instances performed in May. Fitness is evaluated to 0,9926. Visualization of the process model in Fig 3b provides some additional information:

- *Token Counter* (circles with numbers) – indicates parts in the model where the mismatch took place by visualization of the missing and remaining tokens;
- *Failed Tasks* (dark rectangles) – transitions that were not enabled and therefore could not be successfully executed;
- *Remaining Tasks* (gray bigger rectangles) – transitions that remained enabled, which indicates non-proper completion of the process; marked activities should have been executed;
- *Path Coverage* (thick borders) – transitions that were executed;
- *Passed Edges* (black numbers) – Indicates for each transition how often it was followed by various process instances.

The high value of the fitness measure indicates good conformance of a model with the event log behaviour. Problems appear mainly before the execution of *Shipped*, *Rest* and *On fork* activities. This observations are confirmed by complementary analysis performed in Fluxicon Disco. The analysis reveals ten variants of process execution. Each variant refers to aifferent sequence of activities performed in the process. Depending on the variant, the average duration of a process instances varies from 1 day 19 hours to 17 days 1 hour.

The most frequently executed variant fits the *de jure* model: *Production* > *On fork* > *Rest* > *On fork* > *Shipment approved* > *Shipped*. In May 2013, this variant was executed 946 times and accounts for 88,58% of all the process instances. Three other variants include multiple executions of *On fork* > *Rest* subsequence. Those variants also correspond to the *de jure* model. The variants not conforming to the *de jure* model include: execution of many *Rest* activities in a row, double execution of the *Shipped* activity at the end of the process, and execution of *On fork* activity after *Shipment approved* activity. Wrong variants were identifed in 38 process instances that account for 3,55% of all process instances perfomed in May. Those observations are in line with the results of analysis performed in *Conformance Check* plugin.

3.2 First In-First Out

The *Dotted Chart plugin* available in the ProM Framework is used for FIFO rule conformance testing. Generated charts are presented in Fig 4. Each row in the dotted chart corresponds to exactly one process instance and each dot corresponds to an activity instance. Process instances are sorted from the top according to start time. The color of the dot is associated with activity type: light grey color corresponds to *Production* activity while dark grey color corresponds to *Shipped* activity. Other activities were excluded from this analysis.

a) b)

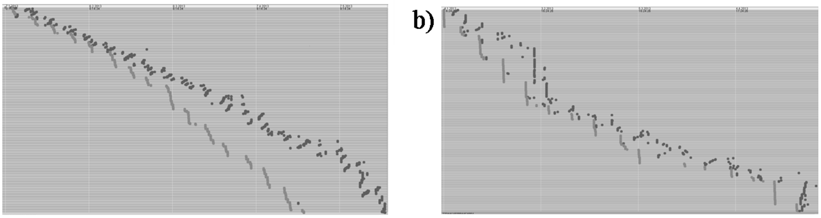


Fig. 4. FIFO analysis using Dotted Charts in ProM for two different mattresses

In Fig 4, the company efforts for FIFO assurance are clearly visible. Some deviations from the rule are visible, as some pallets are shipped significantly later than others. Conformance to the FIFO rule varies from product to product. This is visible when comparing diagram in Fig 4a and Fig 4b.

The non-conformance of storekeepers' behavior to recommendations generated by the WMS has been recorded for 5231 activities performed within 3665 process instances (4% of all the process instances). Only *Rest* and *Production* activities are affected by missing adherence to recommendations. In case of 644 process instances not following the recommendation is justified, as those instances involve damaged pallets. Such pallets must be transported to a separate storage area. The number of process instances affected by not recommended activities is not big but influences conformance to the FIFO rule.

It is worth to note that process mining helps in noticing some deviations or trends. Evaluation concerning positive or negative impact of identified problems on the organization must be done taking into account particular organizational context. For instance, the level of conformance to FIFO is strongly influenced by organization of stands in the warehouse. Current organization of stands forces early produced product pallets to be placed deeper on the stand. Access to such products requires removal of the later produced pallets which usually is not performed. Also the WMS provides recommendations concerning stands not pallets. This fact is visible especially in Fig 4a, where dark grey dots create small lines with direction opposite to light grey lines. Thus, it is worth to note that the perfect conformance of warehouse operation to FIFO will never be achieved by the company.

3.3 Quality Assurance

Using filtering based on activity attributes (*cf.* Section 3.3) it is possible to extract those process instances from the event log that were not accepted by the Quality Department. Exactly 12 such process instances were identified.

In Fig 5, the model describing the execution of such process instances is presented. All the 12 pallets were not shipped to a client. Ten of them were destroyed by performing *Deleted* activity. The two process instances were still running when the event log was created (*Rest* activity). This confirms high conformance to the quality assurance rule.

In Fig 5, performance characteristic of the process is also presented. The numbers correspond to the average time of transition execution. Typically, process instances are short and transitions are performed quite fast. Only transition from *Rest* to *Deleted* activity takes longer time. This is due to one process instance, where this transition took 2 days and 12 hours raising the average.

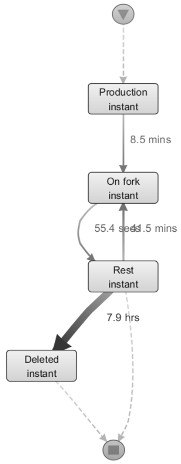


Fig. 5. Process instances missing quality acceptance

3.4 Process Performance

In evaluation of process performance, only process instances that finished with the *Shipped* activity are taken into consideration. The 14-day rule is satisfied for 15 800 of finished process instances. They account for 22% of all the process instances in the event log. 8 562 (11.99%) process instances were performed within 8 days, while 55 569 finished process instances exceeded the desired time.

Although, the rules are rarely satisfied, such behavior is justified in a particular context. During the analyzed time period, the company was creating inventory for an intensive selling period that always takes place in the second half of the year. Early produced pallets were waiting in the warehouse to be shipped in the second half of the year. This fact is confirmed by an increasing number of active process instances growing over time.

3.5 Pallet Damage Handling and Work Distribution

The issue of pallet damage handling appeared in 676 process instances. This number is a number of process instances in which the performed *Rest* activity is associated with a storage area designated for damaged mattress storage. The number of affected process instances is equally distributed among various product families. The number of damages for the three most frequently damaged mattresses was: 127, 126 and 115.

Not all the storekeepers are equally involved in placing the damaged pallets in required storage areas. There are two persons that together deal with 24% of process instances involving damaged pallet. Then the group of 24 employees is responsible for the remaining 76% of process instances.

Observation about an unequal distribution of work among storekeepers can be generalized to all the *Product Management* process instances. It is possible to distinguish 4 persons performing more than 30% of activities within the *Product Management* process. It is worth to note that the *Product Management* process is one of many processes performed in the warehouse. Thus, the work distribution should be analyzed concerning all the processes performed in the warehouse. This is, however, out of the scope of this paper.

Analyzing the table presented in Fig 6, it can be also easily noticed that storekeepers are divided into two separate groups. In Fig 6, rows correspond to different storekeepers while columns correspond to activity types. Numbers contained in the table indicate the number of times a particular type of activity was executed by a particular person. Storekeepers that perform *Production* activities are usually not involved in *Shipped* activities.

Finally, differences in the number of activities performed on different shifts are significant. While the first shift performs 37,54% of activities, the third shift handles only the 29,12% of activities.

On fork	Production	Rest	Shipment approved	Shipped
0	0	0	0	78
0	0	0	0	31
0	0	0	0	1
0	0	0	0	186
0	0	0	0	14
0	0	0	0	202
0	0	0	0	8
0	0	7	0	0
0	0	12	0	0
44	0	5	39	0
159	17	58	101	1
207	202	207	0	0
99	19	34	65	10
113	101	113	0	0
90	10	13	77	124
173	0	1	172	0
0	0	0	0	1
70	70	70	0	1
96	106	96	0	0
0	0	0	0	3
138	1	6	132	180
227	222	227	0	0
0	0	0	0	13
130	11	12	118	0
84	30	36	48	2
0	0	0	0	3
23	0	0	23	80
147	146	147	0	0
50	2	25	25	16
247	0	0	247	124
130	131	130	0	0
0	0	12	0	0
27	0	1	26	4
0	0	7	0	0

Fig. 6. Originator by task matric for Product management process instances executed in May

4 Tools

The set of activities with corresponding attributes was extracted from a database using standard SQL statements. Both Java and Python languages were used to clean the data and extract additional information.

Two tools for process mining were used in the analysis: Fluxicon Disco (version 1.3.6) and ProM Framework (version ProM 5). Disco is a commercial software developed by the Fluxicon company. The academic license allows analysis of event logs up to 1 million events. The process map generator (*cf.* Fig. 3), build-in filtering mechanism, import of CSV files and export to MXML and XES formats compatible with ProM application has been used in the analysis. Disco's built-in filtering algorithm allows including or excluding process instances and activities based on the appearance of one or more attributes. Although Disco does not provide support for process conformance checking, the filtering algorithms can be used to partially fill this gap. This is especially useful when testing various hypotheses. In this way process discovery techniques can be used for conformance testing.

ProM Framework complements the functionality of Disco. ProM has been used to perform conformance tests, generate dotted charts and calculate originator to task matrix. ProM provides refined functionality of conformance testing but general usability of this software is low, e.g., finding basic facts concerning a particular process requires a lot of effort. The most important drawback of the ProM Framework is low efficiency of algorithms composing it which prevents its users from analysis of big volumes of data.

5 Conclusions

Current methods of conformance testing based on process mining are applicable for real-life data. The presented analysis demonstrated business value coming from such analysis. The analysis was performed on data coming from the WMS as it is, without any modifications of the system. The analysis required cooperation with the process owner and occasional support from company's IT department during data preparation. No other resources were involved.

Among others, the analysis performed using process mining techniques lead to the following action undertaken in a company: training were scheduled for storekeepers not following de jure model, the company received a precise information on level of conformance to a FIFO policy for various families of products and was able to schedule action aiming to eliminate discrepancies. Many conformance problems are not always a consequence of storekeepers' behavior or wrong work organizations. Frequently, the wrong configuration of WMS is an issue, e.g., activities saved twice in the database. In such case, process mining methods contribute to information system testing.

The main conclusion is that conformance testing questions requiring the analysis of mutual impact of process instances and social relation structures could not be answered. For instance, does the presence of the two particular storekeepers on the same shift contribute to the increase of damaged pallets number? Currently, no automatic analysis concerning mutual impact of process execution and social relations on one another can be performed. Moreover, the existing conformance testing methods support only structured and repeatable processes like the ones analyzed in this paper. In future works, the methods contributing to conformance testing of unstructured process are planned to be developed. The developed methods must support conformance testing of both processes and social relations. The analysis of mutual impact of process execution on social relations and vice versa should demonstrate great practical value. A step towards such analysis was made in [11] where service protocols were proposed as an approach to modeling social relations as a part of process model was proposed. The next step would be to develop techniques automatically discovering and validating service protocols.

Pure conformance testing functionality is implemented only in the ProM Framework. ProM Framework requires very specialized knowledge concerning processes modeling and interpretation of generated results. ProM has low efficiency of analysis of big amounts of data. Events log with more then 700 000 events can hardly be imported into application. This prevents conformance testing based on data coming from wider time periods. For example, in Section 4.1 only one month data have been analyzed. To answer real-life conformance-related questions, techniques traditionally classified to *process conformance checking* group (*cf.* Section 2) are not sufficient. In practice, the use of those methods must be backed up by *process discovery* methods.

Acknowledgments. This work has been partially supported by the Polish National Science Center. Grant no. DEC-2011/01/N/ST6/04205.

References

1. Aalst, W.M.P., Hee, K.M., Werf, J.M., Verdonk, M.: Auditing 2.0: Using Process Mining to Support Tomorrow's Auditor. Journal Computer 43(3), 90–93 (2010)
2. Dumas, M., Aalst, W.M.P., Hofstede, A.H.: Process Aware Information Systems: Bridging People and Software Through Process Technology. John Wiley&Sons, Hoboken (2005)
3. IEEE Task Force on Process Mining: Process Mining Manifesto (2012), `http://www.win.tue.nl/ieeetfpm/doku.php?id=shared:process_mining_manifesto` (last access: July 18, 2013)
4. Mans, R.S., Schonenberg, M.H., Song, M., Aalst, W.M.P., Bakker, P.J.M.: Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital. In: Fred, A., Filipe, J., Gamboa, H. (eds.) Biomedical Engineering Systems and Technologies, pp. 425–438 (2008)
5. Caron, F., Vanthienen, J., Weerdt, J., Baesens, B.: Beyond X-Raying a Case-Flow: Adopting Different Focuses on Care-Flow Mining (2011), `http://www.win.tue.nl/bpi2011/doku.php?id=challenge` (last access: July 18, 2013)
6. Aalst, W.M.P., Dongen, B.F., Günther, C., Rozinat, A., Verbeek, H.M.W., Weijters, A.J.M.M.: ProM: The Process Mining Toolkit. In: BPM 2009 Demonstration Track, Ulm, Germany (2009)
7. Weijters, A.J.M.M., Aalst, W.M.P.: Process Mining: Discovering Workflow Models from Event-Based Data. In: 13th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2001), pp. 283–290. BNVKI, Maastricht (2001)
8. Aalst, W.M.P., Rubin, V., Dongen, B.F., Kindler, E., Günther, C.W.: Process Mining: A Two-Step Approach to Balance Between Underfitting and Overfitting. Software and Systems Modeling 9(1), 87–111 (2009)
9. Rozinat, A., Aalst, W.M.P.: Conformance Checking of Processes Based on Monitoring Real Behavior. Information Systems 33(1), 64–95 (2008)
10. Conformance Checker, `http://www.processmining.org/online/conformance_checker` (last access: July 18, 2013)
11. Picard, W.: A Formalization of Social Requirements for Human Interactions with Service Protocols. In: Information Sciences, pp. 1–21. Elsevier (2013)

Author Index