LNBIP 97

Witold Abramowicz
Leszek Maciaszek
Krzysztof Węcel (Eds.)

Business Information Systems Workshops

BIS 2011 International Workshops and BPSC International Conference
Poznań, Poland, June 2011
Revised Papers

Springer

Lecture Notes in Business Information Processing 97

Series Editors

Witold Abramowicz
Leszek Maciaszek
Krzysztof Węcel (Eds.)

Business Information Systems Workshops

BIS 2011 International Workshops
and BPSC International Conference
Poznań, Poland, June 15-17, 2011
Revised Papers

Springer

Volume Editors

Witold Abramowicz
Poznań University of Economics
Al. Niepodległości 10
61-875 Poznań, Poland
E-mail: w.abramowicz@kie.ue.poznan.pl

Leszek Maciaszek
Macquarie University
NSW 2109, Australia
and
University of Economics in Wrocław
ul. Komandorska 118/120
53-345 Wrocław, Poland
E-mail: leszek.maciaszek@mq.edu.au

Krzysztof Węcel
Poznań University of Economics
Al. Niepodległości 10
61-875 Poznań, Poland
E-mail: k.wecel@kie.ue.poznan.pl

ISSN 1865-1348 e-ISSN 1865-1356
ISBN 978-3-642-25369-0 e-ISBN 978-3-642-25370-6
DOI 10.1007/978-3-642-25370-6
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011940803

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

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 but have not yet found a place in canonical research. Workshops give researchers the possibility to share preliminary ideas, first experimental results, or to discuss research hypotheses. Discussions held during presentations strengthen the paper and prepare it for publication. From our experience, workshops are a perfect instrument with which to create a community around very specific research topics, thus offering the opportunity to promote it. Topics that do not find critical feedback in the main track of the International Conference on Business Information Systems (BIS) may experience fruitful discussion when confronted with a well-focused audience.

Over the last few decades, business information systems have been one of the most important factors of the transition toward a knowledge-based economy. At the same time they have been subject to continuous rapid development and innovation driven both by industry and by academia. These innovations were carefully observed as well as shaped by researchers attending BIS for the last 14 years.

From time to time a well-recognized conference is collocated with BIS. In 2011, the Poznań University of Economics also hosted the Business Process and Services Computing (BPSC) conference. Additionally, three workshops were successfully organized in conjunction with BIS 2011, covering topics of applications and economics of knowledge-based technologies (AKTB), business and IT alignment (BITA), and legal information systems (LIT). This volume contains papers that were accepted and presented at the BIS 2011 workshops.

On this occasion, we would like to express our thanks to everyone who made the BIS 2011 workshops a success: all Workshop Chairs, members of the Workshop Program Committees, authors of submitted papers, invited speakers, and finally all workshop participants. We cordially invite you to visit the BIS website at http://bis.kie.ue.poznan.pl/ and to join us at future BIS conferences.

August 2011

Witold Abramowicz
Leszek Maciaszek
Krzysztof Węcel

BPSC 2011 Conference Chairs' Message

The papers published in this volume were presented at the 4^{th} International Conference on Business Process and Services Computing (BPSC 2011) held in Poznań, Poland, in conjunction with the 14th International Conference on Business Information Systems (BIS 2011) during June 15–17, 2011.

The BPSC conferences present and publish research findings and IT industry experiences with relation to process-centric service-oriented development and integration of enterprise and e-business information systems. By looking at the convergence of business processes and services computing, BPSC conferences identify the most hopeful trends and propose new directions for consideration by researchers and practitioners involved in large-scale software development and integration. The background principle is to consider process management as a new paradigm with semantics and meta-models as radically different from applications that merely implement business processes.

The BPSC papers in this volume were selected after a rigorous review process. All submitted papers received about three/four reviews on average and some received as many as five reviews. The acceptance rate was 50%.

The Program Committee was chaired by the undersigned. The Organizing Committee was chaired by Krzysztof Węcel of Poznań University of Economics. As the originators of BPSC conference series and as the Chairs of BPSC 2011, we would like to thank the members of the Program and the Organizing Committees for making the 2011 edition of the conference a success.

Witold Abramowicz
Leszek Maciaszek

AKTB 2011 Workshop Chairs' Message

The Third Workshop on Applications of Knowledge-Based Technologies in Business (AKTB 2011), organized in conjunction with 14th International Conference on Business Information Systems (BIS 2011) in Poznań, continued the series of AKTB workshops held in Berlin (2010) and Poznań (2009).

AKTB 2011 kept the tradition of focusing on in-depth knowledge of various application areas, innovative and robust solutions supported by experimental research.

The main goal of this workshop was to bring together researchers and practitioners, specialists and market analysts to share their experience and to invoke scientific discussion in the areas of application of contemporary computational intelligence methods for modeling and implementation of business information systems. We invited papers which could provide advanced services for the information systems users as well as propose innovative solutions for systems and process modeling.

The main topics of the workshop were concentrated on serving the needs of and providing solutions to contemporary businesses by applying intelligent knowledge-based technologies:

- Advanced knowledge-based business information systems
- Computational intelligence for business (artificial neural networks, fuzzy systems, expert systems)
- Decision support systems in business enterprises, financial institutions and e-management
- Knowledge-based models of data mining in business
- Information technologies and software developments for business process modeling and information requirements analysis
- 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

Eighteen articles were submitted to the AKTB workshop. Each paper was evaluated in a double-blind review process by at least two independent reviewers of the Program Committee.

Each reviewer evaluated the quality of the article according to the following criteria: conformity of the article to the workshop topics, originality and novelty, methodological background, relevance of the article, adequacy of the article title and the content, substantiation and validity of the conclusions, and quality of presentation of the paper.

The 28 members of the Program Committee are outstanding researchers from 18 countries, as shown in the next table.

Country	PC members
Austria	2
Brazil	1
France	1
Germany	2
Greece	1
Hungary	2
Korea	1
Lithuania	5
Norway	1
Poland	3
Romania	1
Singapore	1
South Africa	1
Spain	1
UK	3
USA	2

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

The nine highest ranked articles were accepted for the conference proceedings and presentation during the conference. Eight articles were evaluated as not corresponding to the workshop themes or requirements.

The variety of countries represented by the authors and the article acceptance rate are as follows:

Country	Authors	Submitted papers	Accepted	Acceptance rate
Australia	1	0.50	0.50	1.00
Austria	1	1.00	1.00	1.00
Belgium	1	0.50	0.50	1.00
Denmark	1	0.33	0.33	1.00
Germany	10	3.50	1.50	0.43
Lithuania	16	8.67	4.67	0.54
Poland	1	0.50	0.50	1.00
Romania	5	3.00	1.00	0.33

We would like to express our gratitude to all authors of submitted papers, the members of the Program Committee, the Department of Information Systems of the Poznań University of Economics, Vilnius University and our acknowledgement of the outstanding efforts of the Organizing Committee of BIS 2011.

June 2011 Virgilijus Sakalauskas
Dalia Kriksciuniene

Second Workshop on Business and IT-Alignment (BITA 2011)

Preface

A contemporary challenge for enterprises is to keep up with the pace of changing business demands imposed on them in different ways. There is today an obvious demand for continuous improvement and alignment in enterprises, but unfortunately many organizations do not have proper instruments (methods, patterns, best practices etc.) to achieve this. Enterprise modeling and business process management are two areas belonging to a tradition where the mission is to improve business practice and business and IT alignment. In this tradition the alignment process usually is manifested in taking a business from one state into another improved state, i.e., a transformation of the business and its supporting IT into something that is regarded as better. A challenge in business and IT alignment is to move beyond this narrow focus on one tradition or technology. We need to be aware of and 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 2011 was the second workshop on this subject following an event in 2010, which was located at the 13$^{\text{th}}$ International Conference on Business Information Systems in Berlin. 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. Specific focus was on practices of business and IT alignment, i.e., on case study and experiences papers.

The workshop received 14 submissions. The Program Committee selected five submissions and agreed to invite one additional paper for presentation at the workshop.

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

July 2011

Ulf Seigerroth
Kurt Sandkuhl

Organization

Chairs

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

Program Committee

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

LIT 2011 Chairs' Message

The 4th Legal Informatics and Legal Information Technology Workshop was held on June 17, 2011 in Poznań, Poland. The event was organized together with the 14th International Conference on Business Information Systems. It was well received by all participants and should be remembered as a successful continuation of the LIT workshop series.

In terms of statistics LIT 2011 was smaller than previous edition. Six papers were submitted, out of which four works were accepted for publication. This resulted in the moderate acceptance rate of 66.7%.

Still, despite the higher acceptance rate, we managed to stick to the rigorous procedures in terms of accepting papers, similar to the ones applied to the previous edition[1]. We are certain that this factor inevitably led to the very good quality of the workshop and specifically the works contained in this publication.

At the opening of the session, Mirosław Kutyłowski of the Wroclaw University of Technology gave an inspiring keynote address on challenges for electronic identity documents. The resume of the speech is included in the book.

In the paper of Tantisripreecha et al. the problem of creating a procedure to build rules for a legal knowledge base is scrutinized. Their method uses abduction to select rules for the legal reasoning process. At the end of the article a case study is given to prove its practical usability.

Boer et al. propose a model-based diagnosis view on the complex social systems in which public administration organizations operate. Such a view targets multi-agent system and can be helpful in determining problematic agent role instances. The diagnosis itself can be used to explain the driving forces behind policy making in public administration.

Kutyłowski et al. in their study investigate the concept of "sole control" of the signature creation device which is fundamental for the framework of electronic signatures as defined in EU law. On the brink of the technology development the paper shows how the notion of this kind of legally declared idea may evolve in time. The outcome of this work is that recent approaches can potentially give the digital signer much more control and protection than with the technologies employed so far.

The work of Gray et al. discusses the quest for a superexpert system which will support negotiation. The authors classify such systems as ones using expert epistemology – a theory of knowledge. In the paper the system supporting the Harvard Principled Negotiation model is presented. The system manages

[1] Review notes were independently given by at least three 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.

hierarchical complexity of issues in conflict, advises on cumulative scoring of wins and losses, and, above all is able to maximize win – win options.

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 to support the development of knowledge in the IT and law area.

Finally, we would like to give our cordial thanks – especially—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 thanks to the Program Committee members who supported us with their knowledge and judgments. 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 2011). Their support allowed us to organize of the 4th Legal Informatics and Legal Information Technology Workshop.

Erich Schweighofer
Piotr Stolarski
John Zeleznikow

Organization

Program Committee

Emilia Bellucci	Deakin University, Australia
Daniele Bourcier	CNRS-CERSA, Germany
Tania Cristina D'Agostini Bueno	Presidente da Diretoria Executiva IJURIS, Brazil
Jerzy Handschke	Poznań University of Economics, Poland
Arno Lodder	Free University of Amsterdam, The Netherlands
Rafal Morek	University of Warsaw, Poland
Marta Poblet	Autonomous University of Barcelona, Spain
Erich Schweighofer	University of Vienna, Austria
Piotr Stolarski	Poznań University of Economics, Poland
Andrew Stranieri	University of Ballarat, Australia
John Zeleznikow	Victoria University, Australia

Table of Contents

BPSC Conference

AKTB Workshop

BITA Workshop

LIT Workshop

A Roadmap for Research in Business Process Compliance

Shazia Sadiq

School of Information Technology and Electrical Engineering,
The University of Queensland,
St. Lucia, QLD 4067,
Brisbane, Australia
shazia@itee.uq.edu.au

Abstract. We stipulate that for researchers interested in the computational and organizational aspects of research in compliance and risk management, there are two key aspects to be considered. Firstly, research should aim towards a sustainable methodology for compliance management. Secondly, research in this highly multi-disciplinary area must be aligned with industry demands in order to maximize potential for impact and relevance. We have presented a snapshot recommendation for the two aspects, namely a compliance by design methodology that has a fundamentally preventative focus, and an industry driven research agenda that is derived through expert opinion and practitioner feedback. We hope that this will assist researchers to better position their future research endeavors in the area of risk and compliance management.

1 Introduction

Developing strategies to manage inevitable regulatory shifts emerging from government and global reactions to the financial crisis is high on corporate agendas and will continue to be in the coming years. Finding resources to ensure compliance with regulations and managing compliance obligations in a cost effective way are key issues that organisations currently face. Regulatory pressures can be found across all industry sectors and affect all sizes of organization. In addition, pervasiveness of the globalised economy means that compliance may span across geographical boundaries. In these times, more than ever before, organisations seek tools, techniques and methods that allow them to design and implement an effective and efficient compliance regimen. In this project, we will address a particular aspect of this large and complex problem, which is to assist organisations in measuring the adequacy of their control portfolios against particular regulations.

Compliance essentially means ensuring business processes, operations and practice are in accordance with a prescribed and/or agreed set of norms. Compliance requirements may stem from legislature and regulatory bodies (e.g. Sarbanes-Oxley, Basel II, HIPAA), standards and codes of practice (e.g. SCOR, ISO9000) and also business partner contracts. Compliance directives can be complex, vague and often require interpretation. Business will typically deal with a number of regulations/standards at one time, which may have overlapping and even conflicting requirements.

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 1–4, 2011.

In general, a compliance regimen must include three interrelated but distinct perspectives on compliance, *viz.* corrective, detective and preventative perspective. *Corrective* measures can be undertaken due to a number of reasons, but, when undertaken in a proactive manner, position the organisationfavourably with regulators or other control authorities. *Detective* measures are typically based on reporting and traditional audits conducted 'after-the-fact'. A shortcoming of the above two approaches (in varying degrees of impact) is lack of sustainability. Even with an automated detection facility, the hard coded check repositories can quickly grow out of control, making it difficult to evolve and maintain them for changing legislatures and compliance requirements.

A sustainable approach for achieving compliance should fundamentally have a *preventative* focus, thus achieving 'compliance by design' (Sadiq et al. 2007). That is, compliance should be embedded into the business practice, rather than seen as a distinct activity. Intuitively, the role of business process platforms seem instrumental to this end, as they have the potential to contribute to all three perspectives, namely preventative (through compliance aware process design (Lu et al. 2007)), detective (through business process monitoring tools (van der Aalst et al. 2003) and corrective (through model-driven execution of business transactions (Giblin et al. 2007). However, the disparate life cycles, ownership and governance mechanisms of the business and compliance functions within organisations indicates that driving compliance through business processes management is not trivial. In (Sadiq et al. 2010), a methodological framework is presented to align business processes with regulatory compliance obligations.

The framework is based on six key interrelated tasks, namely: (1) Controls directory management in order to provide requisite content management functionality which may span several business functions within the organization as well as several regulations that impact on them; (2) Ontological (or conceptual) alignment between the controls, risk and business processes that are related; (3) Modeling of the controls in order to provide conceptually faithful and machine interpretable specifications for compliance obligations; (4) Tools, techniques and methods for process model enrichment, which may include various checking, annotation and enhancement approaches; (5) assistance with compliance enforcement through targeted tools and technologies such as control automation, simulation for enriched process models etc. and lastly (6) compliance monitoring to ensure continued surveillance where in technologies such as data mining, business intelligence etc. are instrumental.

The various tasks of the above framework hold a number of research challenges. In (Syed Abdullah et al. 2009) we embarked upon a large review of existing information systems research literature to identify contributions that may help fulfill the challenges of such a framework as well as to gain an understanding of the current research landscape.

Furthermore we conducted an empirical study that contrasts the review of literature with industry input derived from expert professionals in the Australian compliance industry. The study has uncovered insights into problematic areas within the compliance management domain and the results show a glaring gap with current research efforts. Below is a summary of the identified research agenda, which we hope can direct ICT and Information System researchers towards industry relevant research questions:

First and foremost, there is an urgent need for proper **benchmarking studies** to help address the challenge of high cost. Particularly for SMEs, there is high cost and great difficulty in **measuring the adequacy of controls** for principles based regulations where the onus is on the organization to design an appropriate compliance regimen. Benchmarking and best practice studies will allow improvement of controls effectiveness, a reduction of costs, and an improved potential to deal with resistance to change through demonstrating methods used by others. Such additional knowledge can further help alleviate the perception of legislation weaknessesin principles based regulations and consequently promote regulation acceptance.

There is also a need for investigation of **process reference models** relating to various regulations. A focus on the development of such reference models and the study of the impact of the use of such models in organizations (i.e. impact on compliance management spending, frequency of breaches, etc) is largely missing in Information Systems research. The development of proven reference models, however, may significantly lessen the cost of compliance management in organizations.

The culture of compliance is ingrained in the daily rituals of each of the firm's employees, including senior management, who must learn to lead by example. There is a clear lack of Information Systems research on **organisationalbehavior**in the context of compliance management. In particular we see a need for investigation of how IT and IS tools can be used to incentivize employees to 'do the right thing' and adapt their practices. There is also a need for the development of relevant IT and IS tools that can help facilitate employee training for compliance management, promote communication among staff and increase organizational capacity to manage its compliance knowledge base.

How the compliance (and risk) factor interrelates with the operations of business units is understudied, with only a small number of researchers working on the **conceptualisation of compliance and risk** requirements per se let alone their inter-relationships with business processes and business activities. A comprehensive and well-grounded conceptual model for compliance and risk is needed.

Further to the point above, tools and methods are needed to **annotate, enhance, analyse and simulate business models** with compliance and risk modeling elements. This will facilitate better coordination between an organization's compliance and business functions and help employees understand compliance value and business relevance.

Although reporting and monitoring tools of high sophistication are available, there is little development towards tools that provide **specialized solutions in monitoring and analysing** compliance related data (partly due the absence of any generic conceptual models for GRC), thus causing big problems for organisations required to create evidence of compliance. Accordingly, we see a need for affordable IT and IS tools that facilitate compliance management self-audits and compliance monitoring activities in general. Furthermore, there is also a clear need for tools that facilitate the identification of non-compliance processes with respect to a given regulation.

Lastly, although frequency of change, as well as inconsistency and overlaps in regulations is beyond the realm of IS research, studies to understand the **impact of regulation changes** (inconsistencies and overlaps) can promote better understanding of the cost of compliance and allow business to lobby for regulatory reform where needed. Multi disciplinary research is warranted in order to cover legal, business and IT aspects. From an Information Systems perspective, there is a need for solutions that can filter out updates that are not relevant to a given organization or industry sector, thus reducing the amount of information that the organization has to process in order to update or assess their compliance management initiatives.

In summary, we stipulate that for researchers interested in the computational and organizational aspects of research in compliance and risk management, there are two key aspects to be considered. Firstly, research should aim towards a sustainable methodology for compliance management. Secondly, research in this highly multi-disciplinary area must be aligned with industry demands in order to maximize potential for impact and relevance. Above, we have presented a snapshot recommendation for the two aspects with the hope that it will assist researchers to better position their future research endeavors.

Acknowledgements. I would like to acknowledge the following colleagues for their contributions to various aspects of this work: Syed Norris Abdullah, The University of Queensland; Guido Governatori, NICTA, Australia; Marta Indulska, The University of Queensland; Ruopeng Lu, SAP Australia; ZoranMilosovic, deontik.com; Ingo Weber, University of NSW; Michael zurMuehlen, Stevens Institute of Technology.

References

Giblin, C., Müller, S., Pfitzmann, B.: From Regulatory Policies to Event Monitoring Rules: Towards Model-Driven Compliance Automation. IBM Research Report RZ3662 (2006)

Lu, R., Sadiq, S., Governatori, G.: Compliance Aware Business Process Design. In: 3rd International Workshop on Business Process Design (BPD 2007) In Conjunction with the 5th International Conference on Business Process Management. LNCS. Springer, Heidelberg (2007)

Sadiq, S., Governatori, G.: A Methodological Framework for Aligning Business Processes and Regulatory Compliance. In: Brocke, J., Rosemann, M. (eds.) Handbook of Business Process Management. Introduction, Methods and Information Systems Series: International Handbooks on Information Systems. Springer, Heidelberg (2010)

Sadiq, S., Governatori, G., Naimiri, K.: Modeling Control Objectives for Business Process Compliance. In: Alonso, G., Dadam, P., Rosemann, M. (eds.) BPM 2007. LNCS, vol. 4714, pp. 149–164. Springer, Heidelberg (2007)

Syed Abdullah, N., Indulska, M., Sadiq, S.: A Study of Compliance Management in Information Systems Research. In: The 17th European Conference on Information Systems (ECIS 2009), Verona, Italy (2009)

Syed Abdullah, N., Sadiq, S., Indulska, M.: Emerging Challenges in Information Systems Research for Regulatory Compliance Management. In: Pernici, B. (ed.) CAiSE 2010. LNCS, vol. 6051, pp. 251–265. Springer, Heidelberg (2010)

van der Aalst, W.M.P., van Dongen, B.F., Herbst, J., Maruster, L., Schimm, G., Weijters, A.J.M.M.: Workflow Mining: A Survey of Issues and Approaches. Data & Knowledge Engineering 47(2), 237–267 (2003)

Beyond Roles: Prediction Model-Based Process Resource Management

Florian Niedermann, Alexandru Pavel, and Bernhard Mitschang

Institute of Parallel and Distributed Systems,
Universität Stuttgart,
Universitätsstraße 38,
D-70569 Stuttgart
{Florian.Niedermann,Alexandru.Pavel,
Bernhard.Mitschang}@ipvs.uni-stuttgart.de

Abstract. The outcome of a business process (e.g., duration, cost, success rate) depends significantly on how well the assigned resources perform at their respective tasks. Currently, this assignment is typically based on a static resource query that specifies the minimum requirements (e.g., role) a resource has to meet. This approach has the major downside that any resource whatsoever that meets the requirements can be retrieved, possibly selecting resources that do not perform well on the task. To address this challenge, we present and evaluate in this paper a model-based approach that uses data integration and mining techniques for selecting resources based on their likely performance for the task or sub-process at hand.

1 Introduction

In this introduction, we will first discuss the general role of resources in business processes. Then, we take a look at how *Business Process Management Systems (BPMS)* typically handle the assignment resources and some of the challenges associated with this approach. Finally, we briefly introduce our *deep Business Optimization Platform (dBOP)* which provides integrated process analysis and optimization facilities and in which the work presented in this paper is embedded.

1.1 Resources in Business Processes

Broadly speaking, resources are the (physical or virtual) entities that execute the activities that make up a business process. Resources are frequently - next to the process structure - the major factor when it comes to deciding the process outcomes such as process duration, cost, success rate or quality.

Consider for instance the mail order return process that will be the subject of the evaluation in Section 4. In the process, an administrative and a technical employee are tasked with assessing both the return information given by the customer as well as the returned item(s). Given that return rates can reach 30% or more [DKDBvdV02], this is a crucial process for any mail order company - the

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 5–17, 2011.

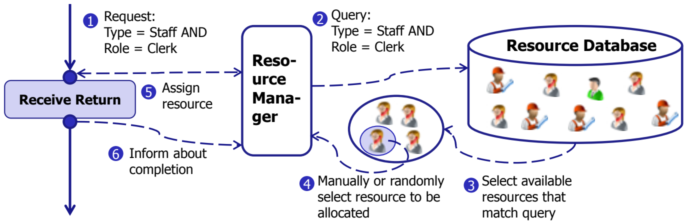


Fig. 1. Standard Resource Assignment

performance of which depends significantly on the employees executing it. As the number of available employees is limited, their employment is associated with considerable cost and their performance can greatly vary, depending on their own attributes and the nature of the returned articles, it is obvious that significant benefits can be realized through optimizing the task allocation in that and other scenarios. There is hence a considerable business need to assign resources in such a way that for a given task and a given process state, the resource that is most likely to create the best outcome (with regards to the process goals) is selected.

1.2 Resource Assignment in Business Process Management Systems

Many of today's business processes are described in a formal language and executed on a so-called *Business Process Management System (BPMS)*. As *BPMS* vendors have recognized the importance of managing resources in business processes, most major *BPMS* engines provide integrated support for the management and assignment of resources to activities and work items. While the individual implementations vary, the general process for resource assignment usually follows the one depicted in Figure 1.

The starting point is the resource (or organizational) database that contains and describes all the resources available to the business process. When a new resource is needed, the activity requiring that resource sends a query describing the resource's properties (typically, the required role) to the resource manager. The resource manager uses this query to fetch a set of resources that match the resource requirements. Finally, one of the resources contained in this set is more or less arbitrarily selected and allocated to the requesting activity (or alternatively, resources can actively claim a new task from the list of available requests).

While well-tested in practice, this approach cannot guarantee that the assigned resource will perform optimally. As the role requirements specified in the resource query are typically quite broad, the set of returned resources is potentially quite large. As the resource manager does not take the likely performance of the resources into account, it will routinely not select the best resource for the task at hand, causing sub-optimal process results.

1.3 Deep Business Optimization Platform

To overcome the limitations of standard resource managers discussed in the previous section, we need a resource manager that is able to predict the likely performance of resources and assign them to resource requests accordingly. For that, we need a platform that is able to gather all the required data, construct a prediction model based on this data and provide its results during process runtime to the resource manager.

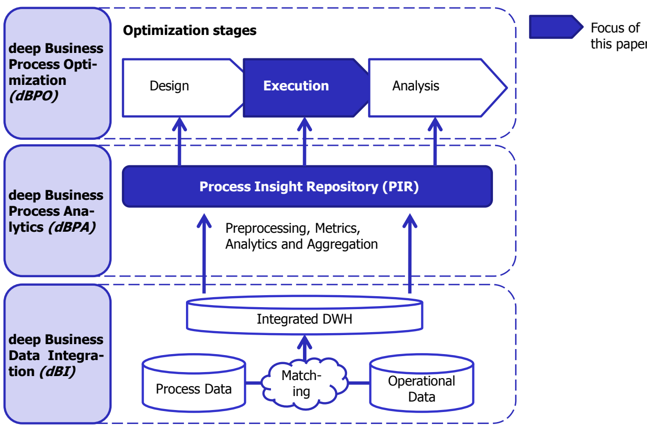


Fig. 2. dBOP Platform Overview

One example for a platform that provides such facilities is our *deep Business Optimization Platform (dBOP)* [NRM10]. The *dBOP*, as conceptually shown in Figure 2, consists of three architectural layers. The integration layer [RNB10] combines various heterogeneous data sources to ensure that all attributes that are relevant for selecting a resource are available to the analysis layer. In the analysis layer [NRM10], various process- and resource-related metrics are calculated based on the integrated data. Using the metrics and the integrated data as input, the analysis layer further builds prediction and classification models, as used e.g., for resource management or decision support [NMR+11]. The aggregated insights are then stored in the *Process Insight Repository (PIR)*. These insights also include the models used by the resource manager presented in this paper. In the optimization layer, the analysis results are used to ensure that the business process design and execution are optimal with respect to the given goal function(s) using a set of formalized optimization patterns [NRM11]. The model-based resource manager presented in this paper is an example for execution-stage optimization.

In this paper, we will show how the *dBOP* can be used to provide a model-based resource manager during process execution that assigns resources not arbitrarily, but based on their predicted performance. For that purpose, we will introduce the underlying process- and resource model as well as the basic functions of the resource manager in Section 2. Building on this, we discuss the detailed concepts and the design underlying our approach in Section 3, including the information required to specify a model-based query. In Section 4, we evaluate the performance of our resource management approach based on a sample case study. Finally, we discuss related work in Section 5 before concluding the paper in Section 6.

2 Fundamentals

In this section, some of the fundamental concepts underlying this paper are discussed. First, we take a look at the structure of resource models and resource queries. Then, we provide a process meta-model that is geared towards modeling resources usages and dependencies and that is used by the resource manager proposed in the next section.

2.1 Resource Models and Queries

While individual implementations vary, most resource (or organizational) models of *BPMS* characterize a resource with the following information: A resource role *Ro* and several resource attributes *Att* that additionally characterize the properties of the resource. In the resource model employed by the *dBOP* we further add a resource type T that determines the basic nature of the resource (typically, one of the following: staff, machine, service). For the purpose of unique identification in a resource model, a resource is given a unique resource name *Rn*. This resource name is resolved to a concrete resource identifier at runtime. For any given process, the set of all resources that can be retrieved using a resource query is called R.

Building on this resource model, we define a basic resource query Q_B as follows:

Definition. *Basic resource query: A basic resource query* Q_B *is a mapping* $T \times Ro \times Rn \times Rel[\times \wp(AC)] \rightarrow R$, *where*

1. T, *Ro, Rn and* R *are defined as above.*
2. $Rel \in \{KEEP, RELEASE\}$ *indicates whether to release the resource after execution (or reserve it for further activities of the process).*
3. $AC \in Att \times COMP \times DOMAIN(Att)$ *lists properties the resource attributes have to fulfill, with* $COMP \in \{=, <, >, \geq, \leq, \in\}$ *where* $\wp(AC)$ *denotes the power set and hence all possible combinations of these properties.*

2.2 Process Meta-model and Resource Dependency Graph

As the next section will show, improving the outcomes of resource selection requires the resource manager to have awareness of the process model and its execution behaviour. Hence, we present in this section a resource-aware process model that provides adequate expressiveness for the concepts discussed in this paper. It is based on a modified version of the graph-based process model introduced in [LR00]. For this paper, we are only introducing the elements essential for the resource management approach presented - see [NRM11] for a complete version.

Definition. *Simplified process graph: A simplified process graph G is a tuple $(V, N, R, \iota, \rho, E_C, E_R)$, $SUCC_R$ is a mapping and μ is a function, in which*

1. *V is the finite set of process data elements (also called variables).*
2. *N is the finite set of process nodes.*
3. *R is the finite set of available resources as defined above.*
4. *$\iota : N \cup \{G\} \to \wp(V)$ is the input data map, with $\wp(V)$ being the power set over V.*
5. *$\rho : N \to Rn$ is the resource usage map.*
6. *$E_C \subseteq N \times N \times C$ is the set of control connectors as explicitly modeled.*
7. *$E_R \subseteq N \times N \times R$ is the set of resource connectors which indicates resource dependencies between activities.*
8. *$SUCC_R : Rn \times N \to \wp(N)$ retrieves all successors of the given node that use the resource with the given resource name.*
9. *$\mu : \wp(V) \cup \wp(R) \to \wp(O)$ performs the matching, i.e., integrates operational data with process data, where O is the set of operational data pertaining to the process.*

In this graph, a resource dependency between two nodes exist, if they use a resource with the same unique name Rn. The direction of the resource dependency follows the direction indicated by the control dependencies.

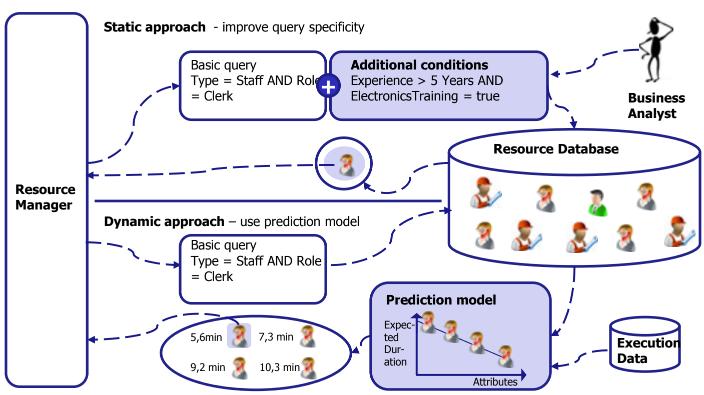


Fig. 3. Alternatives for Improving Resource Assignment Quality

3 Model-Based Resource Assignment

In this section, we first discuss two alternatives for improving the quality of process resource assignment and demonstrate, why a model-based approach will be preferable in most situations to its alternative. Making such a model-based resource manager work depends on the fulfillment of three principal requirements: First, the resource manager's capabilities need to be powerful enough to be readily adaptable to different usage scenarios. Second, the prediction model needs to be designed so that it provides a good selection performance and a high degree of flexibility. Finally, the resource manager needs to be able to offer its services to business processes at runtime. This section will discuss all of these aspects in detail.

3.1 Improving Resource Assignment Quality

In a typical application scenario, a resource query used by any given process activity will return multiple resources that match the given specification. The number of resources returned will thereby depend on the current resource utilization, as the lower the utilization the higher the number of available resources. As discussed in Section 1.2, all of the returned resources will fulfill the necessary attributes required to execute the activity - which, however, does not necessarily mean that all of them will perform equally well.

As Figure 3 shows, there are two basic approaches for improving the likely performance of the assigned resources. In the static approach, the process analyst determines (possibly aided by some analysis tool) the resource attributes that influence the activity performance positively. Then, he refines the query accordingly to make sure that the retrieved resources not only fulfill the basic requirements, but also are likely to perform well at the given activity.

While the static approach has the advantage of not requiring changes to the resource management application, there are several severe disadvantages. First, it requires considerable manual work to determine a well-fitting resource query - especially, since it might have to be updated frequently. Second, to provide an actual improvement, it likely produces enormous and excessively complex queries. Finally, it might actually worsen process performance, as it can significantly reduce the number of resources retrieved, potentially creating a situation where an activity has to wait for a resource to be available (i.e., a resource bottleneck).

In the dynamic (or model-based) approach, the specification contained in the resource query is not fundamentally changed. Instead, a prediction model is constructed based on the recorded process execution history. While this approach requires modification of the resource manager and the presence of execution data, it produces superior results. First, beyond the specification of the model's target parameters, little to no manual work is required. Second, the approach is highly flexible, as the model can include basically arbitrary data into its prediction. Further, it does not have the risk of creating a resource bottleneck through overtly specific queries.

In conclusion, while the model-based approach has higher requirements with regards to the capabilities of the resource manager than the static approach, it can also be expected to yield superior results.

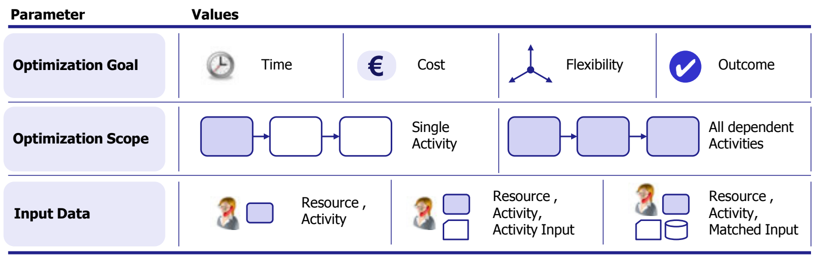


Fig. 4. Resource Manager Parameters

3.2 Resource Manager Capabilities

As the previous section has shown, a dynamic, model-based resource management approach is often likely to yield better results than (static) alternatives. As a monolithic model that can't be parameterized will only perform satisfyingly in a limited set of scenarios, the first step to making this approach work is determining the variability parameters that the resource manager should provide for. We distinguish between parameters that affect the model itself and those that are concerned with the final selection of the resources.

The model parameters need to ensure that the resource manager can provide a process analyst with the optimal model for a given situation. As Figure 4 shows, our resource manager offers three different model parameters. The first parameter concerns the optimization goal Go, such as process time or cost. This is important, as some of the goals are conflicting [RLM05] and it is unlikely that a resource is best along all dimensions (e.g., a highly trained employee might be quite fast, but also expensive). The optimization scope S is the second parameter, which is important when a resource is used by more than one activity. It indicates if the selection optimization only takes the performance of the resource for the current activity or for all activities that it executes into account. The third parameter is the amount of data D to be considered in the determination of the likely optimal resource - either only the resource attributes or also the work item attributes (either as modeled in the process or matched with operational data [RNB10]).

Next to the model parameters, the resource manager further supports additional selection parameters Con, e.g., to specify a maximum resource utilization threshold.

Using the parameters discussed above as well as the foundations introduced in Section 2, we can now define a model-based resource query Q_M as follows:

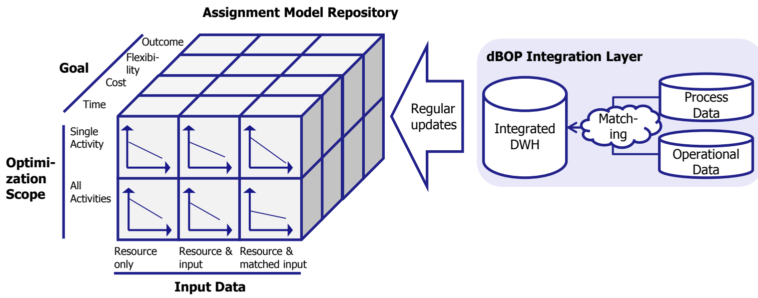


Fig. 5. Multidimensional Resource Assignment Model Repository

Definition. *Model-based resource query: A model-based resource query* Q_M *is a map* $a \times T \times Ro \times Rn \times Rel[\times \wp(AC)] \times Go \times S \times D[\times \wp(CC)] \rightarrow R$, *where*

1. $a \in N$ *is the activity initially requesting the resource.*
2. T, Ro, Rn, Rel, $\wp(AC)$ *are defined as in the definition of the basic resource query.*
3. $Go \in \{TIME, COST, QUALITY, FLEXIBILITY\}$ *determines the optimization goal.*
4. $S \in \{a, SUCC_R(a, Rn)\}$ *determines the optimization scope.*
5. $D \in \{\mu(R) \cup a, \mu(R) \cup \iota(a) \cup a, \mu(R) \cup \mu(\iota(a)) \cup a\}$ *indicates which data to use for the model.*
6. $CC \in Con \times COMP \times DOMAIN(Con)$ *lists further constraints to be considered, with* $COMP$ *defines as in* Q_B.

3.3 Prediction Model Design and Management

To be able to effectively deal with the parameters discussed in the previous section, the resource manager needs to ensure that the prediction model performs well, that the prediction can be done quickly even on large data sets and that the prediction model is current with respect to the recent execution history.

The first choice to make to satisfy these requirements is selecting an appropriate model type. After evaluating neural networks, multilinear regression models, regression trees [HK06] and M5 model trees [Qui92] we have decided to use the latter. M5 model trees are similar to "classical" classification trees [HK06], however, instead of a class label, each leaf node contains a regression model. It offers similar prediction performance to neural networks in most scenarios and superior performance to the other model types. Further, compared to neural networks, the model construction has proven to be considerably faster and more robust in our experiments. Our resource manager hence predicts the resource performance with a M5 model tree with the selected optimization goal as a target variable.

The second choice is how to store and when to update the model. The two extremes for these choice would be to either dynamically construct the model

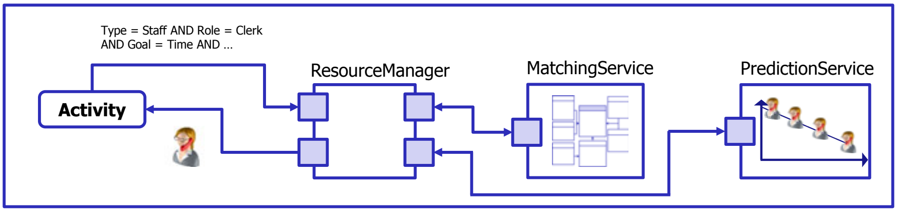


Fig. 6. Resource Manager Implementation

anew whenever a new resource request is or to initially construct a model for all possible model parameters, which is then used to process all future requests. However, the first extreme is impractical due to its high overhead for constructing the models and the second extreme is likely to become outdate. Our resource manager hence uses the third option, which caches and updates the models based on some user-defined criterion (e.g., time elapsed, number of activity executions etc.). To achieve optimal performance, a model is cached for every parameter combination in the multidimensional assignment model repository, as shown in Figure 5. In the context of the architecture shown in Figure 2, this model cache is part of the *Process Insight Repository*.

3.4 Resource Manager Implementation

The final step for making model-based resource assignment work is providing the resource manager's services at run time to business processes. For that, the resource manager is embedded in our *dBOP* platform and implemented as shown in Figure 6. First, the resource manager receives the resource request. Depending on the required input data, the request is further enriched with additional data according to the available mappings to, e.g., a data warehouse [RNB10]. Then, the appropriate model is selected from the model repository (which is implemented based on the WEKA [HFH+09] library). Finally, the performance of the resources that match the query and the constraints is computed and the best resource is returned.

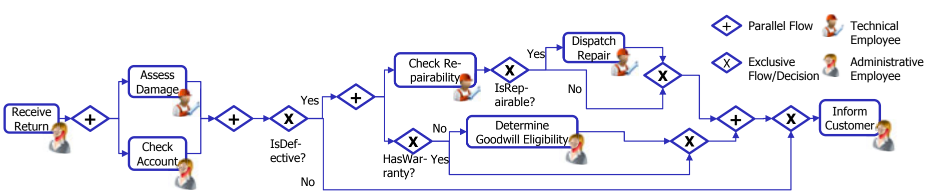


Fig. 7. Mail Order Return Process

4 Evaluation

Using a simplified sample process fragment taken from the retail business domain, we will demonstrate in this section quantitatively the benefit of a prediction model-based resource manager as well as the effects of the different model settings. First, we will briefly introduce the sample scenario and the evaluation setup. Then, we will discuss the evaluation results and their impact on the work presented in this paper. Finally, we discuss further opportunities and limitations of the presented approach.

4.1 Evaluation Design

The evaluation design is based on the sample fragment of the mail order return handling process introduced in Section 1 which is shown in Figure 7. The performance of an employee at each of the depicted tasks can vary greatly, depending on, e.g., experience, (formal) education, familiarity with the particular item being requested or specific received. Hence, this scenario provides a good testing environment for giving a first indication of the benefits that can be realized through a model-based resource assignment approach. As the benefit of using integrated process data has been shown in related work by the authors [NMR+11], this evaluation focuses on two other aspects. First, it measures the impact of different optimization scopes S on the process performance. Second, the impact of different levels of resource utilization are assessed to validate the hypothesis that the benefit of the model-based approach increases with the number of available resources. For that purpose, the evaluation is set up as follows:

- **Scenarios:** Nine different scenarios are measured by combining two dimensions. The first dimension is the level of optimization used, with either no optimization (i.e., a basic resource query Q_B) or with "single activity" or "all activities" optimization scope. The second dimension is the share of resources available during the evaluation, ranging from low (utilization $> 80\%$) to high (utilization $< 20\%$).
- **Data:** In total, we use a set of 1.800 sample processes in the evaluation. The models are built using 1.200 of these sample processes executed using a resource manager without enabled optimization. The evaluation (testing) itself is performed using 600 samples in each of the scenarios listed above.
- **Success measurement:** The success of the respective approach is measured by the performance of the process fragment with respect to the selected goal function (process duration).

4.2 Results

The first observation that can be made from the results shown in Figure 8 is that there is, in this scenario, a clear benefit from using the model-based approach. When using model-based resource allocation, the process performance is on average significantly better than in the non-optimized case. Further, using the

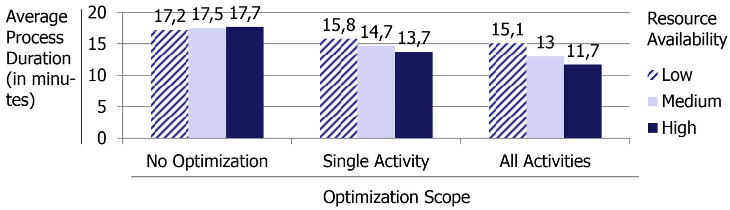


Fig. 8. Evaluation Results

"all activities" optimization scope yields better results than the "single activity" optimization scope.

The second observation is concerned with the impact of different resource utilization rates. The evaluation shows that the lower the utilization rate, the higher the benefit of the model-based approach. This can be attributed to the larger number of resources that are available at any given time, which also increases the likelihood that a resource with a good predicted performance is available.

4.3 Current Limitations and Further Development

Through the case study discussed in this section, we have demonstrated that a model-based resource assignment can be of significant value for improving process performance in certain scenarios. Specifically, the case study has illustrated the usefulness in a scenario where the work to be done can be freely distributed, each process instance or work item is assumed to be of equal value to the process stakeholders and there are no specific resource setup times, e.g., when the type of work item being processed changes. These properties are typical in a "classical" business process management scenario, however, not in a manufacturing scenario.

In a manufacturing context, the work item is typically physical, the priority of work items can vary significantly and there are considerable setup times. To deal with these additional complexities, the presented resource management approach needs to be further extended. Some of these extensions include a cost function for moving work items to a certain resource, a value function for work items and the inclusion of setup times when deciding on which resource to use. All of these extensions are considered in our current work on the topic, which deals with applying the *dBOP* platform to a manufacturing context, as is also discussed in Section 6.

5 Related Work

This paper is part of our work on the *dBOP* platform [NRM10]. The data integration layer is discussed in [RNB10]. The methods employed in the analysis layer are adapted from standard data mining and machine learning literature [HK06] [Qui92]. Examples for their application can be found in [NMR$^+$11]. The optimization layer builds heavily on existing research into business process optimization techniques, such as [RLM05]. Its role within the *dBOP* is the subject of [NRM11].

The approach closest to the one presented in this paper is the classifier-based approach to resource assignment presented in [LWYS08]. However, our approach extends the presented approach in several dimensions. First, the resource manager seems to offer only a single optimization mode. Second, it lacks data integration capabilities, which - as the authors themselves argue - significantly improves classifier performance. Finally, it is unclear how the authors envision the resource manager to be used during an actual process execution. [YJJ07] discusses and validates the feasibility of using prediction models for resource assignment by comparing the results of the model to human judgment in a manufacturing setting. Similarly, [AHM06] show the feasibility of that approach in a software engineering scenario. [JT09] presents an approach that uses association rule mining to assist in the creation of more sophisticated resource assignment rules - basically, a specific variant of the static approach discussed in Section 3.1. [LRDR06] use the mining of workflow data to determine staff assignment rules from the workflow's execution data.

6 Conclusion and Outlook

This contribution introduces a model-based approach to resource assignment that is embedded in the *dBOP*, an integrated platform for analytical process improvement. We have discussed the benefits of such a dynamic approach over the static approach focused on improving the assignment rules themselves. After discussing the details of the model-based resource manager's capabilities and its design, we have illustrated the usefulness of our approach using an order management case study.

As both the core components of the *dBOP* in general and the resource manager in particular have been successfully implemented, our current work focusses on two tasks. First, we are currently working on transferring the tools and concepts of the *dBOP* to the manufacturing application domain. Within this effort, we are also addressing the limitations of the approach discussed in Section 4.3. Second, we are conducting a thorough evaluation of the platform, both from an effectiveness and from a user's perspective. For that purpose, we are currently conducting user studies on the topic and collaborating with several companies for determining additional application scenarios.

References

[AHM06] Anvik, J., Hiew, L., Murphy, G.C.: Who should fix this bug? In: Proceedings of the 28th International Conference on Software Engineering, pp. 361–370. ACM (2006)

[DKDBvdV02] De Koster, R.B.M., De Brito, M.P., van de Vendel, M.A.: How to organise return handling: an exploratory study with nine retailer warehouses. International Journal of Retail & Distribution Management 30(8/9), 407–421 (2002)

[HFH+09] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: An update. ACM SIGKDD Explorations Newsletter 11(1), 10–18 (2009)

[HK06] Han, J., Kamber, M.: Data mining: concepts and techniques. Morgan Kaufmann (2006)

[JT09] Jablonski, S., Talib, R.: Agent assignment for process management: agent performance evaluation. In: Proceedigns FIT 2009 (2009)

[LR00] Leyman, F., Roller, D.: Production Workflow. Prentice-Hall, Englewood Cliffs (2000)

[LRDR06] Ly, L., Rinderle, S., Dadam, P., Reichert, M.: Mining Staff Assignment Rules from Event-Based Data. In: Bussler, C.J., Haller, A. (eds.) BPM 2005. LNCS, vol. 3812, pp. 177–190. Springer, Heidelberg (2006)

[LWYS08] Liu, Y., Wang, J., Yang, Y., Sun, J.: A semi-automatic approach for workflow staff assignment. Computers in Industry 59(5), 463–476 (2008)

[NMR+11] Niedermann, F., Maier, B., Radeschütz, S., Schwarz, H., Mitschang, B.: Automated Process Decision Making based on Integrated Source Data. In: Abramowicz, W. (ed.) BIS 2011. LNBIP, vol. 87, pp. 160–171. Springer, Heidelberg (2011)

[NRM10] Niedermann, F., Radeschütz, S., Mitschang, B.: Deep Business Optimization: A Platform for Automated Process Optimization. In: Proceedings BPSC 2010 (2010)

[NRM11] Niedermann, F., Radeschütz, S., Mitschang, B.: Business Process Optimization Using Formalized Optimization Patterns. In: Abramowicz, W. (ed.) BIS 2011. LNBIP, vol. 87, pp. 123–135. Springer, Heidelberg (2011)

[Qui92] Quinlan, J.R.: Learning with continuous classes. In: 5th Australian Joint Conference on Artificial Intelligence (1992)

[RLM05] Reijers, H.A., Liman Mansar, S.: Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics. Omega 33(4), 283–306 (2005)

[RNB10] Radeschütz, S., Niedermann, F., Bischoff, W.: BIAEditor - Matching Process and Operational Data for a Business Impact Analysis. In: Proceedings EDBT (2010)

[YJJ07] Yingbo, L., Jianmin, W., Jiaguang, S.: A machine learning approach to semi-automating workflow staff assignment. In: Proceedings of the 2007 ACM Symposium on Applied Computing (2007)

Analyzing Separation of Duties Constraints with a Probabilistic Model Checker*

Tamara Mendt[1,2,3], Carsten Sinz[1], and Olga Tveretina[1]

[1] Institute for Theoretical Computer Science,
Karlsruhe Institute of Technology, Germany
[2] SAP Research Center, Karlsruhe, Germany
[3] Universidad Simón Bolívar, Venezuela
05-38546@usb.ve, carsten.sinz@kit.edu, olga@ira.uka.de

Abstract. Separation of Duties (SoD) is the concept that conflicting activities cannot be assigned to the same individual. A goal of SoD is to separate roles and responsibilities to reduce the risk of fraud or error. We consider the problem of verifying SoD constraints in the presence of uncertain information. We demonsrate the feasibility of implementing probabilistic model checking in a business process design with a case study. Modeling and verification is done with the probabilistic model checker PRISM.

1 Introduction

A business process workflow is a sequence of operations, declared as work of a person or a group of persons. Workflow management systems provide the facilities to define, manage and execute business processes. Hence, designing a business process workflow is essential for any business or organization.

Role Based Access Control (RBAC) is a widely used access control model. In a workflow environment, tasks are not linked directly to subjects. This means that within an organization roles are created for various job functions. The permissions to perform certain operations are assigned to specific roles. Roles assign the execution of a task to a group of persons with certain responsibilities. When a representative of a role leaves the company, task definitions can remain unchanged. Just the role has to be 'untied' from that person [1,2].

Many criminal activities are performed by insiders of companies, and, therefore, it is of great importance to implement mechanisms to prevent such illegal activities [1]. Separation of Duties (SoD), alternatively called segregation of duties, is the concept that ensures that conflicting activities cannot be assigned to the same individual and, as a result, they are appropriately segregated. A main goal of SoD is to separate roles and responsibilities to reduce the risk of fraud or error to an acceptable level. SoD policies are used for internal control and require some critical tasks to be executed by different persons.

* This work was supported in part by the "Concept for the Future" of Karlsruhe Institute of Technology within the framework of the German Excellence Initiative.

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 18–29, 2011.

Additionally, the legal regulations about the security and data privacy of businesses are constantly growing and changing. This means companies need to put a lot of effort into ensuring that their businesses are compliant with the regulations that govern them. Therefore, it is necessary to provide business process modeling software with tools that allow the business experts designing a process to specify constraints within their process model.

The use of model checking to verify business processes with respect to defined constraints has been studied before (see the next section for references). However, these approaches generally assume that the behavior of the business process in runtime is predictable. We consider the problem of verifying properties of business processes that contain uncertain information.

One of the first studies on the use of a probabilistic model checker for the verification of business process constraints is reported in [3]. This work continues along this line of research, and analyses SoD constraints with PRISM, a tool for the modeling and analysis of systems which exhibit probabilistic behavior.

The rest of the paper is organized as follows. Section 2 gives an overview of related work. In Section 3 we present the basic (probabilistic) model checking notions continuing with discussion of the probabilistic model checking tool PRISM which is used throughout our study. Section 4 presents our case study which is based in the context of SoD. Section 5 shows the modeling of our case study using the tool PRISM. Finally, we present the results of our experiments in Section 6 and conclusions as well as future research directions in Section 7.

2 Related Work

Since the mid nineties, verification techniques applied to business related workflow have been studied intensively [4]. One of the first model checking approaches in workflow models was a part of the Testbed Project [5,6]. This work focuses on making model checking accessible to users who lack training in formal methods.

The use of model checking to verify business processes with respect to defined constraints has been studied in [6,7]. Model checking in the area of business process management has been used mostly to ensure soundness and consistency in business process specifications [8]. Thao et. al. [9] identify a set of fundamental requirements for supporting semantic constraints in process management systems and analyze to which extent existing model checking approaches meet these requirements.

Another important application of model checking in the business process context is ensuring compliance with security requirements such as authorization constraints [10]. Model checking approaches applied to business process security can be found, among others, in the verification of compliance with SoD properties. Schaad et.al. [10] present a model checking approach classifying SoD policies into five main categories according to the level of restriction they specify over the business process. For each type of SoD, an LTL specification is defined. This study demonstrates that LTL model checking is a viable approach for verifying the consistency of SoD properties on workflow systems with role based access control policies.

The approach for model checking SoD properties taken in [11] is quite similar to the previous, as it also proposes the use of LTL to represent security policies in an access-controlled workflow system. The verification tool that is used in this study is the SAT based model checker for planning systems: SATMC. In this approach, the specification of the workflow model and the access control policies occur separately and a more formal specification of the security properties as LTL formulae is presented.

3 Probabilistic Model Checking and PRISM

Model checking is a technique to verify formally finite state systems [12]. The essential idea behind model checking: A model-checking tool accepts system requirements or design (called models) and a property (called specification) that the final system is expected to satisfy. The tool then outputs yes if the given model satisfies given specifications and generates a counterexample otherwise.

Whereas model-checking techniques focus on the absolute guarantee of correctness, in practice such rigid notions are difficult to guarantee. Instead, systems are subject to various phenomena of a stochastic nature making the correctness become less absolute.

The key point of probabilistic model checking is the ability to combine probabilistic analysis and model checking in a single tool [13].

In the scope of this work the probabilistic systems that will be used are Discrete Time Markov Chains (DTMCs) and Markov Decision Processes (MDPs). Formally, a DTMC can be defined by a tuple $M = (S, s_0, T, AP, L)$ where S is a set of states, $s_0 \subseteq S$ corresponds to the initial set of states, $T : S \times S \to [0, 1]$ is a transition relation such that $\forall s \in S$, $\sum_{s' \in S} T(s, s') = 1$, AP is a set of atomic propositions, and finally $L : S \to 2^{AP}$ is a a labeling function.

Markov decision processes are generalizations of Markov chains which allow a system to present non-determinism. For each state there may exist more than one probability distribution for transitions to next states.

PRISM (PRobabilistIc and Symbolic Model checker) is a tool for the modeling and analysis of systems which exhibit probabilistic behavior [14]. PRISM was chosen among several probabilistic model checking tools because of its user friendly interface and overall benefits related to model checking times and expressiveness.

In PRISM, system modeling is accomplished using a module-based language which basically consists of state variable declarations and transitions between states. This tool only allows the use of boolean and integer variables and they can be declared as follows:

```
<guard> -> <command>;
```

where guard is a predicate over system variables (representing a state or set of states) and `<command>` is the transition executed by the system if the guard command evaluates to true (if the current state matches a state of the set determined by the guard). If the transition must be chosen probabilistically, the discrete probability distribution is specified as follows:

```
<guard> ->  p1:<command1> + p2:<command2> + ... + pN:<commandN>;
```

Transition represented by `commandi` is executed with probability `pi`, and the sum of the probabilities must equal 1.

The PRISM property specification is based on temporal logics PCTL and PCTL*. PCTL and PCTL* include a probabilistic operator P, used to reason about the probability of a system property. This operator can be used to determine whether the probability of a path property holding is between some bounds (qualitative approach), or to calculate the numerical value of the probability of the property holding (quantitative approach).

The first statement declares a boolean type variable which initially takes the value false. The second statement declares an integer type variable with initial value 1, and which can be assigned values between 0 and 10. The global state of a system is determined by the current value of the variables.

4 Case Study

In this section we analyze separation of duties constraints in a role based access control system in presence of uncertain information. The case study models the planning and decision making process to determine the supply chain route for the transaction of a order made from a retailer. This case study is used to reason over separation of duties (SoD) policies since these are the most common application level security properties defined over business processes. Probabilities are present in the workflow of this process, particularly in exclusive choices where each possible choice has a probability value associated to it.

The case study consists of an entirely procedural business process model which represents the process that occurs when a new product order is submitted to company X. The process includes a subprocess which must be executed within a certain time bound. Each task in the subprocess has a probability distribution associated to the time it requires to execute, making it possible to reason over the probability of the subprocess ending correctly or ending due to time expiration.

A business process can be represented by linking tasks in a directed graph, where the arcs of a graph are associated with some predicate conditions. Figure 1 shows the business process model which will be explained in detail throughout this section.

The actors involved in the process are the following:

- **R1: Supply chain director:** Manages organizations manufacturing and logistics operations so that approved products are manufactured on schedule, within quality standards and budget objectives.
- **R2: Logistics manager:** controls the logistics function to ensure availability of raw materials

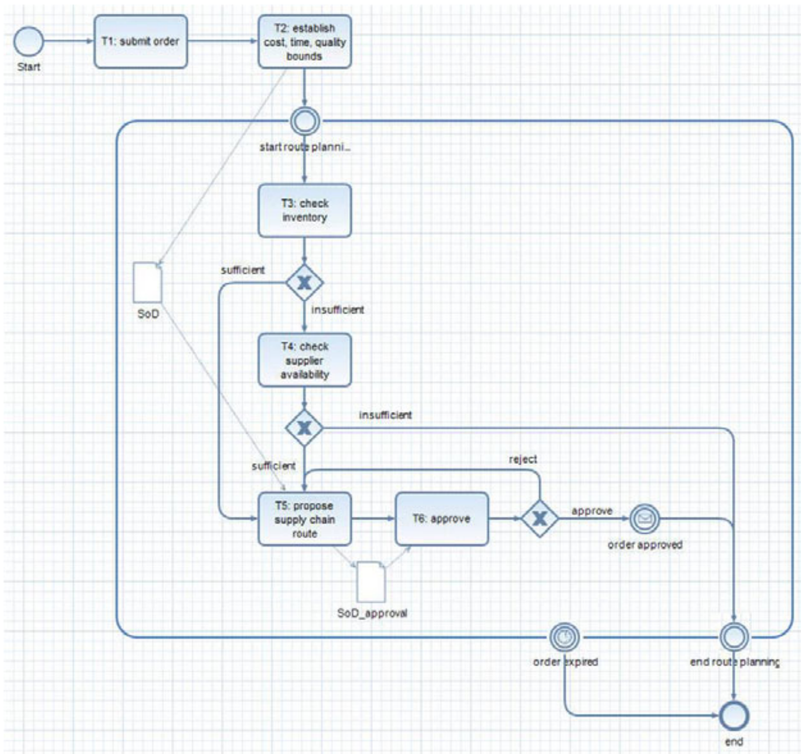


Fig. 1. Process *New Order*

– **R3: Route planning engineer:** designs optimal routes throughout the supply chain to meet defined standards.

The process consists of six tasks:

– **T1: submit order**. When an order is requested from a retailer, a system user must submit the order in the system. The order must specify how many of each type of product is being requested along with other requirements.
– **T2: establish cost, time and quality bounds**. Once the order has been submitted, and based on the specified client requirements, maximum cost, time and quality bounds are determined. Other supply chain constraints such as preference for one packaging option over another can also be specified in this activity.
– **T3: check inventory**. The available inventory (finished products and purchased, non used raw materials) is checked to verify if it is enough to cover the order. If it is, there is no need to check supplier availability and the process continues directly with route proposal.
– **T4: check supplier availability**. If it is the case that there is not enough stock in inventory to cover the order, suppliers are contacted to check if it

is possible to verify if they can provide enough goods. If this is not possible, the order is rejected and the process ends. On the other hand, if the supplier can provide the required amount of goods, the process continues the flow with the next task.

- **T5: propose supply chain route**. Once it has been verified that the company can cover the requested order, the supply route planning process takes place. Taken into account the set of restrictions established in T2, a route is designed for the products of the order. All routed related information is covered in this step, from budget, approximate time required, etc.
- **T6: approve route**. Once a route is proposed, it is necessary to approve the route. If it is approved, a message is sent to the retailer confirming availability and including budget and time expectance information. If the route is not approved, it must be reworked, the process flow returns to the route proposal task.

Each task of the process can be executed by a set of roles and company X has only hired one user for each role.

$$\begin{aligned}
&\mathrm{T1} \leftarrow \{\mathrm{R1, R2, R3}\}\\
&\mathrm{T2} \leftarrow \{\mathrm{R1}\}\\
&\mathrm{T3} \leftarrow \{\mathrm{R1, R2, R3}\}\\
&\mathrm{T4} \leftarrow \{\mathrm{R1, R2}\}\\
&\mathrm{T5} \leftarrow \{\mathrm{R1, R3}\}\\
&\mathrm{T6} \leftarrow \{\mathrm{R1, R2, R3}\}
\end{aligned}$$

Separation of duties properties must hold between task T2 and task T5: the person who establishes the route requirements cannot be the same person to propose the route. SoD must also hold between task T5 and T6: the user who approves a route must be different than the one who proposes the route.

Probability distributions in this case study are present in two different areas. The first is in the exclusive choices of the workflow. It is assumed that the probability of taking one path or another, can be determined, and is independent of the path taken before reaching the exclusive choice. In the case study, each exclusive choice follows only two options, the probability of taking one path is p for some rational value p between 0 and 1, and the probability of taking the other is $1 - p$.

The process *New Order*, contains a sub process *SC Route Planning*. If the sub process *SC Route Planning* does not finish within a time limit, the process ends without having approved a SC route, hence the order which was submitted, expires. Every task in subprocess *SC Route planning*, requires a determined amount of time to execute. This time cannot be exactly predicted, it follows a probabilistic distribution. Table 1 shows the probabilistic distribution for the time required for each task in the subprocess.

Using this information, it is possible to predict how likely it is that an order expires. From the customer relationship point of view, it is in the interest of company X to avoid expired orders. Also, the likeliness of SoD properties between tasks of the subprocess *SC Route Planning* holding, may increase if the

Table 1. Time distribution for tasks in subprocess *SC Route Planning* of the process *New Order*

Time	T3:check inventory	T4:check supplier	T5:prop. route	T6:app. route
10	45%	0.01%	0.01%	88.13%
20	20%	0.23%	0.02%	9.25%
30	15%	0.46%	0.04%	1.3%
40	15%	0.81%	0.06%	0.8%
50	4%	1%	0.08%	0.45%
60	0.5%	25%	0.13%	0.07%
70	0.217%	30%	0.25%	0%
80	0.12%	27%	0.41%	0%
90	0.057%	10%	0.7%	0%
100	0.046%	2%	1.3%	0%
110	0.035%	1.1%	3.01%	0%
120	0.025%	0.85%	9.08%	0%
130	0%	0.62%	15.34%	0%
140	0%	0.55%	21.72%	0%
150	0%	0.37%	47.85%	0%

subprocess ends before an order is approved. This is because the activities of the subprocess become less likely to repeat themselves.

In this use case the applicability of probabilistic model checking is observed in the verification of SoD properties over a business process including randomly distributed control flow. Model checking can also be applied to predict the probability of an order expiring in a process where each activity has a randomly distributed time duration.

5 Modeling of the System

This case study is entirely procedural. Still, the process is represented as a MDP instead of a DTMC. Non-determinism in the case study is not present in the workflow but in the user to task assignment. Since for each task there is a set of users which have permission to execute it, the choice of the specific user who will execute it, is non-deterministic.

The MDP is constructed through three modules. As for the MDP in the previous case study, a main module or workflow module is defined to represent the workflow of the process. Two integer variables are defined: `nxt_task` and `end_task`. These range from cero to the number of tasks of the process. The value of `nxt_task` determines the next task to be carried out in the process. The value of `end_task` indicates the last task to be executed in the process. At the beginning of the process `nxt_task` has value 1 and `end_task` takes the value cero. When `nxt_task` takes the value cero, the process has ended.

The second module of the MDP represents the user to task assignment of the process. An integer variable `user` is defined which ranges over the number of users of the process. The transitions in this module occur concurrently with those of the workflow module. For each task which is executed, the variable user takes the value of a user which is allowed to execute that task. This way, the system state `(end_task=i & user=j)` represents that task number `i` was executed by user number `j`.

The third module is related to the time distribution of the tasks in subprocess *SC Route Planning*. An integer variable `time` is defined which ranges between cero and a the maximum time which can elapse before the subprocess ends abruptly. With each task execution of the subprocess, `time` is updated concurrently. If the time related variable reaches its maximum value, the subprocess has expired and will end.

Table 2. Variables used for MDP

Identifier	**Type / Range**	**Description**
`nxt_task`	int:[0...#_tasks]	Controls workflow
`end_task`	int:[0...#_tasks]	Indicates which task has just been executed. This is necessary for determining the user to task assignments made throughout the process.
`user`	int:[0...#_users]	Indicates the user which executed the last task. Each user is assigned an integer between 1 and the number of users.
`time`	int:[0...max_time]	Indicates how much time has elapsed in the subprocess *SC Route Planning*. If time=max_time, the subprocess ends.

In case study, model checking is used for the verification of separation of duties (SoD) properties. Two properties of interest have been identified and are listed below, along with the PCTL formulae corresponding to each property. For the definition of the constraints, labels `"Ti_Uk"` were defined and represent the system states where `(end_task=i & user=k)`.

- **SoD(Ti,Tj):** This property states that the two task parameters must be executed by different users. The constraint template used for this property is "responded_absence" which specifies that if a state in the first label set parameter is reached, then no state in the second label set can be reached during the entire process. The PCTL formula for this property is:

$$\texttt{Pmax=?}\left[\bigwedge_{\texttt{Uk}\in users} responded\_absence(\texttt{"Ti\_Uk","Tj\_Uk"})\right] \equiv$$

$$\texttt{Pmax=?}\left[\bigwedge_{\texttt{Uk}\in users} (\texttt{(F "Ti\_Uk")} \Rightarrow \texttt{(G ! "Tj\_Uk")})\right]$$

– **at_most_N(Ti):** states that task `Ti` can be executed by the same user a maximum of N times within a process. The constraint template used for this property is ”absence_N” which specifies that states in a specific label set cannot be reached N times. If a task involved in an SoD policy cannot always be executed by the same user, the SoD policy is less likely to hold throughout the process. The PCTL formula for this property is:

$$\texttt{Pmax=?}\,[\bigwedge_{\texttt{Uk}\in users} absence\_(N+1)\texttt{("Ti\_Uk")}] \equiv$$

$$\texttt{Pmax=?}\,[\bigwedge_{\texttt{Uk}\in users} \neg\, existence\_(N+1)\texttt{("Ti\_Uk")}] \equiv$$

$$\texttt{Pmax=?}\,[\bigwedge_{\texttt{Uk}\in users} \neg\,(\texttt{F ("Ti\_Uk" \& (}existence\_N\texttt{("Ti\_Uk"))))}]$$

Notice that, once again only `Pmax` is considered, rather than `Pmin`. This is because model checking is used to validate properties under the best case scenarios, rather than the worst case scenarios.

Binding of duties policies, stating that two tasks must be executed by the same users can be defined similarly. However these were not implemented in the use case. For this case study it can also be in the interest of the user to calculate the probability of the time related to the subprocess running out. This can be checked simply with an ”existence” constraint template stating that eventually a state will be reached where the time variable take the maximum time value (`Pmax=?[F time=max_time]`). The results obtained from running PRISM over the case study can be found in the following section.

6 Experimental Results

The dimensions of the model of the case study depend of the maximum time value allowed. Table 3 compares the size of five models with different maximum time values. It is visible that the range of variables has an important weight upon the size of the model. Though this case study presents non-determinism, the size of the model remains manageable since the number of users is small. Table 3 shows the size comparison between models using different values for maximum time before the subprocess expires. The column model size is the product of the number of states and number of transitions of the model. Relation refers to the division of the size of the model of a given row and the size of model of the first row. This number is included as a comparative measure. Notice that the relation between model sizes grows proportionally to the relation between the max_time values.

Table 4 shows the model size comparison between models including different number of users. The maximum time related to the time variable for the models in this table is fixed at 300. Notice that as the number of users grow, the size of the model grows exponentially, making model checking unfeasible. When

Table 3. Case study: comparison of model sizes with different values for max_time

max_time	# states	# transitions	model size	relation
100	172	1,654	284,488	
200	301	5,598	1,684,998	5.92
300	409	8,350	3,415,150	12.01
400	489	10,300	5,036,700	17.71
500	569	11,962	6,806,378	23.93

increasing the number of users, it was assumed that the added users had permission to execute every task. Restricting the users which can execute a task to only a small proportion from the whole set of users can significantly reduce the level of non-determinism. Still, keeping the size of the model manageable when the amount of users involved in the process is large, is challenging, or impossible.

PRISM was run on one model with max_time = 300. Four different SoD related properties were verified. The results of the model checking process can be found in Table 5.

The number of states and number of transitions in this table correspond to the states and transitions of the MDP-DRA (Markov Decision Process - Deterministic Rabin Automata) product (see LTL model checking algorithm in [15]). For the resulting times, the model checking process was run 10 times for each

Table 4. Comparison of model sizes with different number of users for max_time=300

# users	# states	# transitions	model size	relation
3	409	8350	3,415,150	
4	573	16971	9,724,383	2.85
5	737	28558	21,047,246	6.16
10	1557	130983	203,940,531	59.72
20	3197	558283	1,784,830,751	522.62

Table 5. Model checking times for different properties

property	# states (MDP-DRA)	# transitions (MDP-DRA)	time
SoD(T2,T5)	1209	22106	0.16 ± 0.02
SoD(T2,T5) & Sod(T5,T6)	1575	27843	0.29 ± 0.05
at_most_3(T5)	3269	49270	0.45 ± 0.01
SoD(T2,T5) & SoD(T5,T6) & at_most_3(T5)	3529	60403	0.98 ± 0.03

property and the confidence level used for the error bounds is 0.99. Notice that although the time requirements increase with the complexity of the PCTL formula, they still maintain low enough (less than 1 second) to consider the model checking process efficient.

7 Conclusions and Future Research

Our work follows along the lines of the study presented in [3], and demonstrates the strength of applying probabilistic model checking techniques on business process models. With a PRISM-based representation of the use cases, we could demonstrate that it is possible to model business processes through DTMCs and MDPs. In this sense, we could show that a broad class of business processes dealing with uncertain information can be modeled manually through DTMCs and MDPs using PRISM.

A major limitation we encountered was the restriction that formulas cannot be defined recursively, resulting in a large number of variables in the model of the system. This is less of a PRISM deficiency, but more of a general limitation basing modeling on Markov chains.

Future investigations in our application area could include the design of new case studies of business processes including stochastic behavior. Future research could focus on integrating PRISM into a Business Process Modeling (BPM) engine, too. This would also include the study of software requirements and user interface requirements to allow BPM designers to include stochastic behavior in business processes and verify properties based thereon.

In order to represent business processes through DTMCs or MDPs, a formal semantics must be determined to be able to handle stochastic information. Otherwise automating the translation process of a business process to a stochastic model will be quite hard or even impossible. Since the two case studies were modeled manually, any process related variable could be considered as randomly distributed. However, future studies focused on the automation of the modeling process should define probability templates or some technique to enable process designers to include certain aspects of randomness in processes following predefined parameters.

As for regular model checking techniques, we could envisage a business process modeling framework which combines procedural and declarative approaches. Although this modeling combination has been partially included in BPM notation, it is lacking a formal constraint semantics and model checking techniques applicable to ad hoc subprocesses. We could also imagine a model checking framework where the designer is provided with a highly flexible business process modeling tool, in which the underlying model checking engine does not model the system separately from the properties, but rather optimizes the model checking process by adapting the model for each property.

Acknowledgements. We thank Dr. Andreas Schaad and Dr. Phillip Miseldine from the Security and Trust Group at the SAP Research Center Karlsruhe for their support.

References

1. Knorr, K., Weidner, H.: Analyzing separation of duties in petri net workflows. In: MMM-ACNS, pp. 102–114 (2001)
2. Lawrence, L.G.: The role of roles. Computers and Security 12(1), 15–21 (1993)
3. Mendt, T., Sinz, C., Tveretina, O.: Probabilistic Model Checking of Constraints in a Supply Chain Business Process. In: Abramowicz, W. (ed.) BIS 2011. LNBIP, vol. 87, pp. 1–12. Springer, Heidelberg (2011)
4. Wynn, M.T., Verbeek, H.M.W., Aalst, W.M.P., Ter Hofstede, A.H.M., Edmond, D.: Business process verification - finally a reality! Business Process Management Journal 15(1), 74–92 (2007)
5. Janssen, W., Mateescu, R., Mauw, S., Springintveld, J.: Verifying business processes using spin. In: Proceedings of the 4th International SPIN Workshop, pp. 21–36 (1998)
6. Janssen, W., Mateescu, R., Mauw, S., Fennema, P., van der Stappen, P.: Model Checking for Managers. In: Dams, D.R., Gerth, R., Leue, S., Massink, M. (eds.) SPIN 1999. LNCS, vol. 1680, pp. 92–107. Springer, Heidelberg (1999)
7. El Kharbili, M., de Medeiros, A.K.A., Stein, S., van der Aalst, W.M.P.: Business process compliance checking: Current state and future challenges. In: Modellierung Betrieblicher Informationssysteme (MobIS 2008). LNI, vol. 141, pp. 107–113 (2008)
8. Lu, R., Sadiq, S., Governatori, G., Yang, X.: Defining Adaptation Constraints for Business Process Variants. In: Abramowicz, W. (ed.) Business Information Systems. LNBIP, vol. 21, pp. 145–156. Springer, Heidelberg (2009)
9. Ly, L.T., Göser, K., Rinderle-Ma, S., Dadam, P.: Compliance of semantic constraints - a requirements analysis for process management systems. In: 1st Int'l Workshop on Governance, Risk and Compliance - Applications in Information Systems, Montpellier, France (2008)
10. Schaad, A., Lotz, V., Sohr, K.: A model-checking approach to analysing organisational controls in a loan origination process. In: SACMAT 2006: Proceedings of the Eleventh ACM Symposium on Access Control Models and Technologies, pp. 139–149. ACM, New York (2006)
11. Armando, A., Ponta, S.: Model checking of security-sensitive business processes (2010)
12. Clarke, E.M., Emerson, E.A., Sistla, A.P.: Automatic verification of finite-state concurrent systems using temporal logic specifications. ACM Trans. Program. Lang. Syst. 8(2), 244–263 (1986)
13. Kwiatkowska, M., Norman, G., Parker, D.: Stochastic Model Checking. In: Bernardo, M., Hillston, J. (eds.) SFM 2007. LNCS, vol. 4486, pp. 220–270. Springer, Heidelberg (2007)
14. PRISM website (2010), http://www.prismmodelchecker.org
15. Baier, C., Katoen, J.-P.: Principles of model checking. MIT Press (2008)

Autocompletion for Business Process Modelling

Karol Wieloch, Agata Filipowska, and Monika Kaczmarek

Department of Information Systems,
Faculty of Informatics and Electronic Economy,
Poznan University of Economics
{a.filipowska,k.wieloch,m.kaczmarek}@kie.ue.poznan.pl

Abstract. This paper presents an idea and prototype of the semantic-based autocompletion mechanism supporting development of business process models. Currently available process modelling tools support business analysts by suggesting elements that may be incorporated in the process, validating modelled processes, providing additional descriptions easing automation, etc. However, these solutions based mainly on syntactic data, disregard proper identification and usage of previously modelled process fragments. The mechanism described in this paper analyses context and annotations of process tasks (also on the semantic level) in order to deliver a list of suggestions for possible successor tasks: process fragments that may complete the model being developed.

We argue that the proposed autocompletion mechanism has an ability to improve the efficiency of the modelling process by among others reducing modelling errors and shortening the duration of the modelling process.

1 Introduction and Problem Description

Business process management (BPM) encompasses methods, techniques, and tools to design, enact, control, and analyse operational business processes involving humans, organizations, applications, documents, and other sources of information [1]. Typically, BPM follows a life cycle that consists of four phases, namely: design (modelling), implementation, enactment, and analysis [2].

Business process modelling being an introductory phase of the whole BPM lifecycle is a way of collecting, documenting and analysing processes. This is usually done by business analysts: experts being able to describe every detail of a business process taking place in a company or public administration.

There is a number of tools supporting the process modelling, some of which enable also process execution, monitoring and further analysis (supporting the whole process lifecycle). Among these tools the most popular ones are ARIS Platform, iGrafix and Proforma.

These tools, being very usable taking into account different aspects, however, provide limited support for users when it comes to intelligent process modelling. Typical process modelling resembles a scratchboard rather than a technical approach. The guidance offered concerns mainly process syntactic data and the

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 30–40, 2011.

process semantics is omitted. But in a company, some of the processes interact, some of them include similar process fragments. If an expert is unaware that a fragment was previously modelled, he spends time on re-modelling it. Often also a significant training is required to teach people how to model their processes using a given tool or notation.

Additional issues concern the fact, that business models are usually transformed manually into executable models and the new, executable process models are neither understandable nor available for changes to business experts. These process models stay out of reach of process experts and are not used while modelling new processes.

In this paper we address the problem of supporting business analysts while modelling processes by providing them with a set of process fragments that semantically match the process they are working on. This list is prepared based on the description of the process being modelled. The only requirement is that the process needs to be (at least partially) semantically annotated. We call this functionality an autocompletion mechanism similarly to what is available in the Integrated Development Environments (IDEs).

The work presented in the paper was a part of the approach developed in the FP6 EU SUPER[1] project regarding the Semantic Business Process Management. The SBPM is to close the Business-IT gap by using semantic technologies [3]. Similarly to how Semantic Web services achieve more automation in discovery and mediation as compared to conventional Web services, in SBPM more automation should be achieved in process modelling, implementation, execution and monitoring phases by using ontologies and Semantic Web services technologies.

The autocompletion mechanism and component that was developed is integrated with the BPMO editor (version of the WSMO Studio[2]).

The paper is structured as follows. Next section presents the overview of the related work concerning the issue of process autocompletion. Then a description of the scenario follows. Next, we present the completion strategies and the mechanism overview. The article concludes with an overview of the architecture of the solution and a brief summary.

2 Related Work

Our approach is related to the work of [4], which also concerns semantic business process modelling. In this paper the author describes methods and techniques to support modelling of semantically annotated business processes in Petri Nets focusing on measuring the similarity of process task labels to suggest matching process fragments. We developed our approach for Business Process Modeling Notation (BPMN) and its underlying metamodel BPMO, being less restrictive when it comes to process modelling. Here, we use the notion of the process fragment and their decomposition suggested by [5].

[1] http://www.ip-super.org

[2] http://www.wsmostudio.org/

[6] present an approach for supporting the modelling of business processes using semi-automated Web service composition techniques. They take into account the functional part of service descriptions when making suggestions during modeling. Similarly to [6], [7] in their work focus on implementation aspects considering also the process context and non-functional properties in addition to the functional properties, thus increasing the level of precision of the suggestions.

We also apply the ontology stack described in detail in [8]. In our work we focus on enhancing user experience while modelling processes by making the improving the effectiveness.

3 Scenario and User Interface

In this section we describe the principles of the autocompletion mechanism based on naming and functional annotations of process tasks and discuss its implementation within the semantic modelling tool developed within the SUPER project.

3.1 BPMO Editor and Available Tasks' Descriptions

The proposed autocompletion mechanism is envisioned to constitute an improvement within process modelling activities. Therefore, it needed to be implemented as a prototype within one of the available business process modelling tools. However, within our work we wanted to focus on the tool supporting the semantic annotation of business process models being the on-going trend in the area of BPM. Therefore, the decision has been reached to implement the proposed mechanism within the Business Process Modelling Ontology (BPMO) Editor for modelling semantic business processes developed within the EU 6FP SUPER project. The BPMO modelling environment [9] provides a basic functionality for adding semantic annotations to the process models. It operates on the Business Process Modelling Ontology [10] being an abstraction over the BPMN and EPC notations. Within the BPMO a task is defined as a Business Activity (Figure 1) and

```
Concept BusinessActivity subConceptOf upo#BusinessActivity
    hasName ofType (0 1) _string
    hasDescription ofType (0 1) _string
    hasNonFunctionalProperties ofType (0 1) BusinessActivityNonFunctionalProperties
    hasBusinessDomain ofType upo#BusinessDomain
    hasBusinessFunction ofType upo#BusinessFunction
    hasBusinessStrategy ofType upo#BusinessStrategy
    hasBusinessPolicy ofType upo#BusinessPolicy
    hasBusinessProcessMetrics ofType upo#BusinessProcessMetrics
    hasBusinessProcessGoal ofType upo#BusinessProcessGoal
    hasBusinessResource ofType upo#Resource
```

Fig. 1. An excerpt from the BPMO ontology

thus, can also refer to business attributes such as a business policy or a business process goal. Of course, tasks themselves have additional attributes to represent information about interactions with a partner process.

From our perspective the most important implications from using BPMO are as follows:

- All processes, process fragments and tasks are instances of BPMO concepts and as such are identified using IRIs (Internationalized Resource Identifiers);
- Some attribute types in Tasks are defined at a syntactic level, e.g. hasName ofType String.
- Some attribute types in Tasks are defined at a semantic level i.e. the specific values of task's attributes refer to the IRI of a concept within an ontology. The attribute of the special interest is the hasBusinessFunction pointing to the instances from the Business Functions ontology [8]. This allows for taking advantage of reasoning during the autocompletion procedure.

The scenario supported within the BPMO editor follows.

3.2 Supported Application Scenario

The following scenario is supported by the proposed solution. A business user is modelling a process and wants the system to suggest a task or a process fragment that would follow the one just added to the diagram (Figure 2).

As there might be many unfinished branches within the process flow, the user must explicitly select the task to be auto-completed. Before the auto-complete function is fully activated a user needs to select autocompletion strategy he wants to follow. Once the auto-completer returns a list of matching process fragments (Figure 3) the user selects the most appropriate fragment.

The chosen fragment replaces the selected task and possibly some more preceding ones depending on the strategy selected (Figure 4).

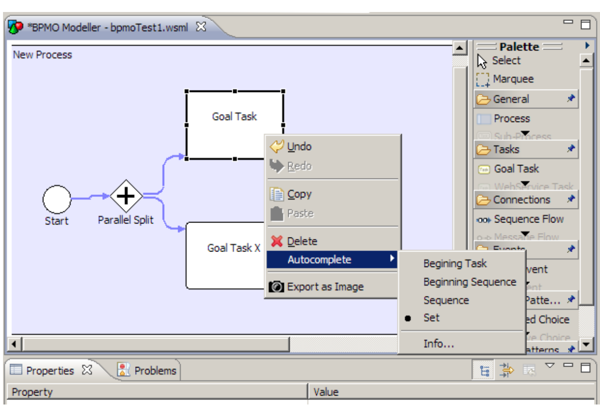


Fig. 2. An initial state of the modelling environment for the Auto-completer

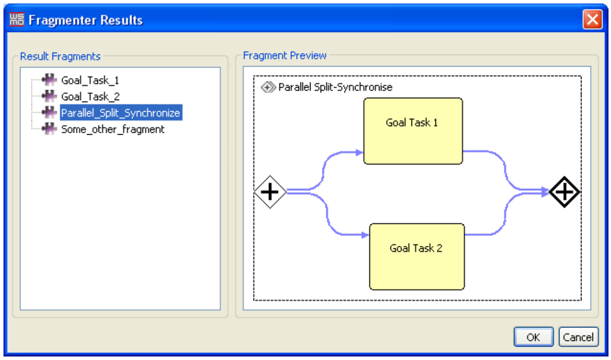


Fig. 3. A list of suggested fragments

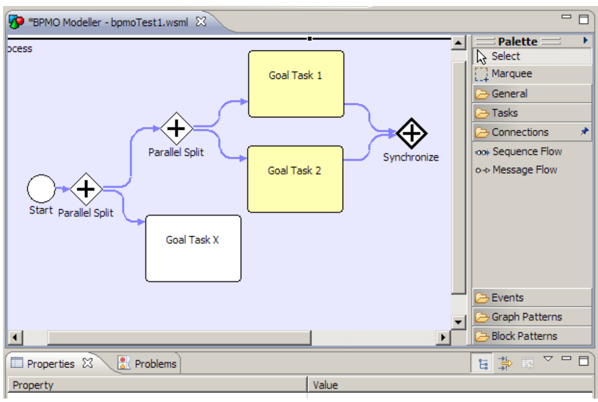


Fig. 4. The result of a single auto-complete action

The autocompletion strategy defines which part of an already modelled process is used as a query for finding potentially matching fragments and how the matching is performed. Five different autocompletion strategies have been defined and implemented within the tool. Their short overview follows.

3.3 Autocompletion Strategies

Beginning Task Match Strategy. The Beginning Task Match is the simplest autocompletion strategy. A user selects a task to be auto-completed. The autocompleter retrieves all tasks whose activities and names match the selected task (i.e. the similarity is above specific threshold) and ranks fragments they belong to according to their similarity to the user-provided task. After the user selects one of the suggested tasks, the task he initially selected is replaced with the chosen fragment. Effectively the selected task is auto-completed with a remaining part of the matched fragment.

An example follows:

a process modelled by a user: *StartEvent-Task0-Gateway0-Task1-Task2*
a problem to be solved by the auto-completer: *find all the fragments that begins with the task similar to Task2*
returned solution (suggested fragment): *Task2a-Task3*
a process model once a user accepts the suggestion: *StartEvent-Task0-Gateway0-Task1-Task2a-Task3*

Beginning Sequence Match. A user selects a task to be auto-completed. The auto-completer finds the longest sequence of tasks that ends with the task selected by the user (i.e. it "moves back" until a gateway or process beginning is reached). This pattern sequence is used to find matching fragments that start with the sequence of task matching the user-provided sequence of tasks. After the user decides to use one of the suggested fragments, the initially found pattern sequence of task will be replaced with the chosen fragment. Effectively the task selected by the user is auto-completed with a remaining part of the matched fragment.

An example follows:

a process modelled by a user: *StartEvent-Task0-Gateway0-Task1-Task2*
a problem to be solved by the auto-completer: *find all the fragments that begins with the sequence similar to Task1-Task2*
returned solution: *Task1a-Task2a-Task3-Task4*
a process model once a user accepts the suggestion: *StartEvent-Task0-Gateway0-Task1a-Task2a-Task3-Task4*

Sequence Match. The strategy is a more general version of the "Beginning Sequence Match". In this strategy the query is constructed in the same way as in the previous strategy. The difference is that the auto-completer discovers all fragments that contain sequence of tasks matching sequence of tasks provided by user at any location within them. As the result the task selected by the user is followed by a remaining part of the matched fragment and is preceded by the initial part of the selected fragment.

An example follows:

a process modelled by a user:
StartEvent-Task0-Gateway0-Task1-Task2

a problem to be solved by the auto-completer:
find all the fragments that contains the sequence similar to Task1-Task2

returned solution:
Task5-Task1a-Task2a-Task4-Task3

a process model once a user accepts the suggestion:
StartEvent-Task0-Gateway0- Task5-Task1a-Task2a-Task4-Task3

Set Match. This strategy generalises the "Sequence Match" strategy in a sense, that the auto-completer does not pose a sequence restriction on potentially matching fragments. It returns all fragments that contain tasks matching all user-provided tasks, regardless of their positions, order and existence of additional tasks between them.

An example follows:

a process modelled by a user:
StartEvent-Task0-Gateway0-Task1-Task2

a problem to be solved by the auto-completer:
find all the fragments that contain tasks similar to Task1 and Task2

returned solution:
Task5-Task2a-Task3-Task1a-Task3

a process model once a user accepts the suggestion:
StartEvent-Task0-Gateway0- Task5-Task2a-Task3-Task1a-Task3

Auto-complete Name. This additional functionality provides a user with suggestions of the full name of a task whose name is only partially typed.

4 Matching Procedure

This section describes the matching procedure we have applied within the proposed solution. Matching consists of calculating distance value on four levels between following objects:

- task's attribute vs. task's attribute (attributes level),
- task vs. task (tasks level),
- set or sequence of tasks vs. other set or sequence of tasks (sequence/set level),
- one or more tasks (depends on strategy) vs. process fragment (fragment level).

Each level requires different procedure for calculating distance value as discussed within the following subsections.

4.1 Attributes Level

As already mentioned in the previous section, in the editor the tasks are described according to the BPMO ontology. At the level of task's characteristics we decided to consider two attributes: hasName and hasBusinessFunction.

Distance between task names is calculated using one of the standard string distance measures [11].

Distance between business functions assigned to a two given tasks is calculated using the following formula:

$$d(i,j) = \frac{CPL(i,j)}{\frac{H(i)+H(j)}{2}} \tag{1}$$

where $d(i,j)$ denotes the distance value, $CPL(i,j)$ is the length of the common path in the Business Function Ontology subsumption hierarchy (starting from the root) of the two concepts i and j; $H(i)$ is the length of the path in subsumption hierarchy from the root to the concept i.

4.2 Tasks Level

The distance between tasks is calculated as the average of the attribute level distances.

4.3 Sequence/Set Level

The similarity of two sequences is calculated as the arithmetic sum of similarity between respective pairs of tasks.

The similarity of two sets is calculated as the maximum of similarity of user-provided set and all possible subsets of the same size.

4.4 Fragment Level

The similarity of a user query to a specific process fragment is calculated as the maximum of similarities between set or sequence provided as user query and all possible sets or sequences belonging to specific fragment. For example in case of Beginning Task Match a single task is compared against first task of each fragment, while in case of Set Match set of n tasks is matched against each n-element subset of tasks of each fragment and the maximum of obtained similarity measures is used as query-to-fragment similarity.

The similarity measure is used while providing a user with a ranking of process fragments matching the process a user is working on. The more relevant fragments, the higher in the ranking they are presented. In case the similarity measure is 0, the process fragment is not included in the list presented to the user. When, the number of process fragments will increase, certain limits on the similarity measure (show only top 10% of matching processes) may be introduced.

5 Solution Architecture

This section provides some insights into the architecture of the proposed solution which is presented in the Figure 5.

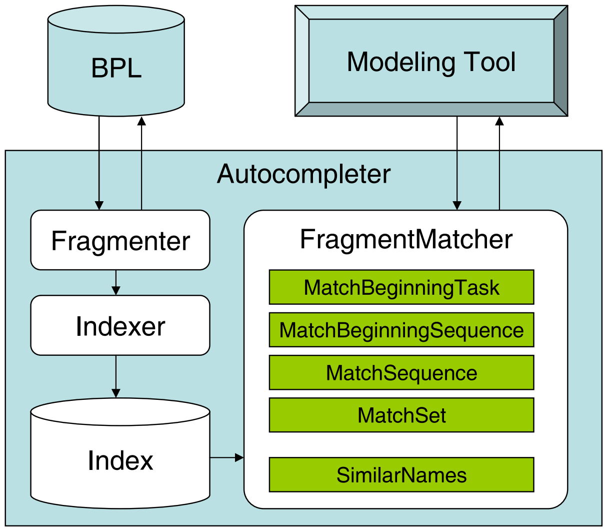


Fig. 5. Data flow within the Auto-completer and between other components

The Autocompleter is a separate component that communicates with the Business Process Library (BPL) and needs to have a graphical user interface (modelling editor). As such it can be used within any modelling tool.

The following components can be distinguished within the autocompleter itself:

- Fragmenter,
- Indexer along with the Index repository,
- Discoverer.

Discoverer is the main subcomponent of the auto-completer. It is responsible for implementing various autocompletion strategies. Each of matching strategies is implemented as a plug-in using the same index structure and implementing the same interface. Thus, a number of additional matching strategies may be implemented.

Fragmenter is responsible for identifying SESE fragments (Single Entry Single Exit fragments) taking advantage of the algorithm used e.g. in [5], being able to identify all SESE fragments in linear time, within processes stored in the

Business Process Library. Identified fragments are stored in the BPL again and send to the Indexer.

The role of the Indexer is to retrieve all relevant information regarding process fragments that is required during the discovery phase. The retrieved information is stored within an internal data structure (index).

Two structures are used to index process fragments and utilised in answering user queries. First and the simplest structure consists of a list of all tasks present (with IRIs of corresponding process fragments), identified by an IRI. The second structure contains all n-grams (including 1-grams) of tasks, specifying also if a specific sequence corresponds to the beginning of specific fragment.

The index is used basically to select candidate sequences or tasks that may be similar to a user query. In case of Beginning Task Match and Beginning Sequence Match strategies n-grams (n being the number of tasks in user query; n=1 in case of Beginning Task Match) that correspond to a start of any fragment are selected. In case of Sequence Match all n-grams are selected (whatever their position in the task is). Finally, in case of the Set Match, first the search space is limited to all fragments that contain at least n elements (i.e. that contain a sequence of at least n elements); next all subsets of n elements of all these fragments are considered for similarity measurement.

The index is implemented as an SQL database (HSQLDB Java library is used). Each of match strategies is responsible for construction of an SQL query performing selection of IRIs of process fragments. The similarity measure for two tasks in accessible directly in a query (as a stored procedure, implemented in Java). The results of such query are handled by the Discoverer module.

6 Conclusions

This paper presents a novel approach to improve existing tools that support business process modelling. Our solution supports the business analysts actively during their work helping them to find process fragments completing the process they are working on. This may lead to increasing effectiveness of the process modelling by avoiding the duplication of work (working on process models that were already developed).

The presented approach consists of two distinct phases: process fragments identification and their description and application of previously identified process fragments in processes being modelled. Although our algorithm operates on the semantic annotations of process models, we may also operate also on the syntactic level what increases the application possibilities of our approach.

The approach was tested by business users, that underlined its usefulness while process modelling. Some of the major benefits offered by our approach are i) identification of process elements that may enrich the process being modelled, ii) suggestions based not only on syntactic, but also semantic information, iii) improved model quality and faster modelling process.

References

1. Van Der Aalst, W.M.P., Hofstede, A.H.M.T., Weske, M.: Business Process Management: A Survey. In: van der Aalst, W.M.P., ter Hofstede, A.H.M., Weske, M. (eds.) BPM 2003. LNCS, vol. 2678, pp. 1–12. Springer, Heidelberg (2003)
2. Muehlen, M.Z., Rosemann, M.: Multi-paradigm process management. Development (Bpmds), 169–175 (2004)
3. Hepp, M., Leymann, F., Bussler, C., Domingue, J., Wahler, A., Fensel, D.: Semantic business process management: a vision towards using semantic web services for business process management. In: IEEE International Conference on eBusiness Engineering, ICEBE 2005, vol. 5(1), pp. 535–540 (2005)
4. Koschmider, A.: Ähnlichkeitsbasierte Modellierungsunterstützung für Geschäftsprozesse. Universitätsverl., Karlsruhe (2007)
5. Vanhatalo, J., Völzer, H., Leymann, F.: Faster and More Focused Control-Flow Analysis for Business Process Models Through SESE Decomposition. In: Krämer, B.J., Lin, K.-J., Narasimhan, P. (eds.) ICSOC 2007. LNCS, vol. 4749, pp. 43–55. Springer, Heidelberg (2007)
6. Schaffner, J., Meyer, H., Weske, M.: A Formal Model for Mixed Initiative Service Composition. In: IEEE International Conference on Services Computing, pp. 443–450. IEEE Computer Society (2007)
7. Born, M., Brelage, C., Markovic, I., Pfeiffer, D., Weber, I.: Auto-completion for Executable Business Process Models. In: Ardagna, D., Mecella, M., Yang, J. (eds.) BPM 2009. LNBIP, vol. 17, pp. 510–515. Springer, Heidelberg (2009)
8. Born, M., Markovic, I., Filipowska, A., Starzecka, M., Kaczmarek, M., Walczak, A.: Business Functions Ontology and its Application in Semantic Business Process Modelling, pp. 136–145 (2008)
9. Dimitrov, M., Simov, A., Stein, S., Konstantinov, M.: A BPMO Based Semantic Business Process Modelling Environment. In: Hepp, M., Hinkelmann, K., Karagiannis, D., Klein, R., Stojanovic, N. (eds.) SBPM. CEUR Workshop Proceedings, vol. 251, CEUR-WS.org (2007)
10. Cabral, L., Norton, B., Domingue, J.: The Business Process Modelling Ontology. In: Proceedings of the 4th International Workshop on Semantic Business Process Management, SBPM 2009, pp. 9–16. ACM, New York (2009)
11. Markovic, I., Pereira, A.C.: Towards a formal framework for reuse in business process modeling. In: ter Hofstede, A.H.M., Benatallah, B., Paik, H.-Y. (eds.) BPM Workshops 2007. LNCS, vol. 4928, pp. 484–495. Springer, Heidelberg (2008)

Extending BPMN 2.0 to Enable Links between Process Models and ARIS Views Modeled with Linked Data*

Feng Gao, Wassim Derguech, and Maciej Zaremba

National University of Ireland, Galway,
Digital Enterprise Research Institute
firstname.lastname@deri.org
www.deri.org

Abstract. Business Process Modeling Notation (BPMN) is an emerging standard for modeling business processes. In its 2.0 version it defined formal semantics to its elements which allows a process execution engine to understand how processes should be integrated and executed. However, BPMN2.0 elements use shallow String datatypes for their identification (e.g., process participants, process resources) which does not explicitly identify entities that a given element pertains to. In this paper, we propose to extend BPMN 2.0 in order to allow for linking its elements to external entities following the Linked Data principle. Our proposal leverage the extension mechanism provided by BPMN 2.0 which does not result in the alteration of the language specification. Our extensions consider the ARIS views to support better integration and collaboration from different perspectives of the enterprise systems.

Keywords: BPMN 2.0, ARIS, Linked Data, RDF.

1 Introduction

Business processes are capable to provide automated support for the collaboration between people, departments, organizations and corporations by coordinating their resources and behaviors when integrated with the information systems. The term *Business Process Management* (BPM) is sometimes referred to as the "business process optimization process". The BPM life cycle is generally divided into four phases: design, deploy, execution and analysis. Similar to the principles of developing *supportable* enterprise systems proposed in [8], BPM iterates these phases to improve the supportability.

Process modeling in the design phase plays a central role in the BPM life cycle. A recent advancement in the standardized business process modeling language of BPMN 2.0 [10] defines the formal semantics of its elements thus the process execution engines have explicit knowledge about their behaviors. Moreover, BPMN 2.0 provides means for its extensibility. However, business processes

* This work is funded by the Lion II project supported by Science Foundation Ireland under grant number SFI/08/CE/I1380 Lion-2.

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 41–52, 2011.

models defined in BPMN 2.0 do not follow Linked Data principles and are defined without linking to external information. BPMN 2.0 elements use shallow *String* datatypes for their identification (e.g., process participants, process resources) what does not explicitly identify entity that the given element pertains to. According to [1,11] both private and collaborative process models can benefit from semantic technologies, while the work in [14] elaborate the necessity of associating process models with the other views in the enterprise systems. Our aim is to improve the BPMN 2.0 process models in these aspects.

We leverage the BPMN 2.0 extension mechanism to make the process models interlinked with relevant knowledge of the enterprise systems, i.e., with the functional view, data view, organizational view and control view. We model the information in these views with our RDF based descriptions. Therefore, the extended process models can benefit from the machine processable knowledge and make improvements during the phases of the BPM life cycle. Current research efforts concentrate much on utilizing semantic technologies to provide automated execution support (i.e., ontology-backed service composition) for process models and bridging the gap between the business perspective and technical perspective, however we have a distinct interest and emphasize on using these technologies to improve the other phases in the BPM life cycle as well.

The remainder of this paper is organized as follows. We give account of the technologies and building blocks that we use in Section 2. We introduce our approach and an example scenario in Section 3. We describe related work in Section 4. Finally, in Section 5 we conclude the paper and discuss future works.

2 Background

In this paper, we leverage the **BPMN** 2.0 extensibility in order to achieve better integration between the external knowledge and business processes. The **ARIS** enterprise system modeling methodology is a guideline for our work. **Linked Data** principles are followed while we build our framework.

2.1 Business Process Modeling Notation

The BPMN language provides a graphical representation for the business managers and process designers to organize and manage their business process models. It has recently evolved to its 2.0 version which defines the formal semantics of its elements. It also provides a mapping from BPMN to WS-BPEL, which is an block-structured execution language for Web services. The vision of the BPMN 2.0 is to make the processes executable in a top-down manner where process designers can deploy, test and run developed processes without having to deal with the low level process execution details, which makes the execution of the language more "friendly" to upper layer users. Another interesting feature introduced in the latest release of the language specification is its extensibility. Both BPMN 2.0 *Flow Objects* and *Artifacts* can be extended to allow the process designers to express additional features of BPMN 2.0 process models.

2.2 ARIS Architecture

The Architecture of Information System (ARIS) provides a widely accepted enterprise modeling methodology. The ARIS framework defines five different views: organization view, functional view, data view, control view and product/service view. ARIS House shown in Figure 1 shows these five views and relations between them. As depicted in the figure, the process models are the controlling dispatcher of resources and actions, they act as the central role in the ARIS house, associating with all the other views and integrate them in a uniformed way. A process model respecting these associations between different views is understandable among involved parties despite of their different work perspectives and is therefore capable of providing better understandability, maintainability and scalability.

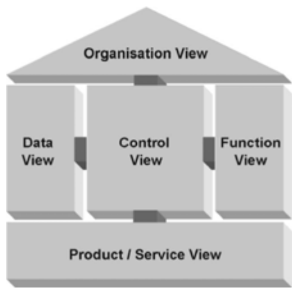


Fig. 1. The ARIS House (from [14])

2.3 Linked Data

Linked Data [3] is a set of best practices for publishing and interconnecting structured data on the Web. Linked Data provides explicit links between data from diverse domains such as social networks, organizational structures, government data and many others. The ultimate benefit of the Linked Data paradigm is the increased machine-readability of published and interconnected data. Linked Data is published using RDF where URIs are the means for connecting and referring between various entities on the Web.

Over the last years we have observed an increasing adoption of Linked Data principles and an explosion of datasets specified in RDF. Early adopters included mainly academic researchers and developers. However, more recently we have observed a considerable interest from organizations in publishing their data in RDF. Some of the most prominent examples include BBC music data[1], British government data[2] or Library of Congress data[3]. At the same time, we have observed an increasing number of public vocabularies (ontologies) and their interconnectedness. Data published in RDF uses and refers to these public vocabularies what further improves understanding of published data by data consumers.

[1] http://www.bbc.co.uk
[2] http://data.gov.uk
[3] http://id.loc.gov

3 Approach

The Business Process Management life cycle consists of four phases: Design, Deploy, Execution and Analysis as illustrated in Figure 2. In our approach, the extended BPMN processes are typically created and annotated in the design phase. However, we do not exclude the possibility that other phases may also produce the extensions for the BPMN documents, e.g.: some dynamic process meta-data like event triggering time, process completing time and other control information may be annotated during the process execution. BPMN 2.0 elements provide a shallow information i.e.: string types for element names. In order to improve the interoperability of the knowledge on the business function, organization and data etc, BPMN 2.0 elements should be described in more details with respects to all ARIS house views. With such extended annotations we establish the links between the process model and the different views in the ARIS house, while these views are modeled with linked data, they are interlinked with each other and can be further linked to the external open data to obtain better data integration using linked data mechanism. Thereafter, process management environments will be able to provide meaningful answers for the queries on the process models and instances with specific interests based on the common machine-understandable knowledge.

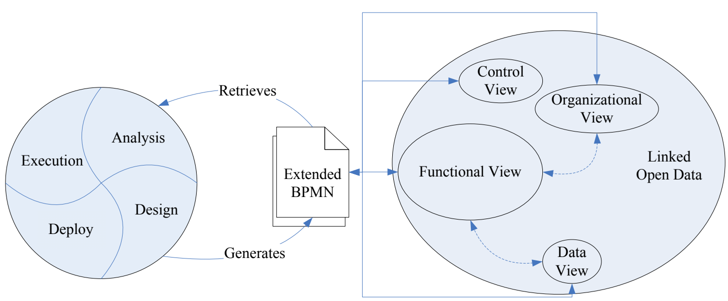


Fig. 2. The Overview of The Framework

In the following, we will detail our concept of *Structured Web Resource* for describing the information model in the ARIS views, the extension mechanism for the BPMN 2.0 and a use case scenario to demonstrate our methodology.

3.1 Structured Web Resource

We propose the Structured Web Resource (SWR) model to describe the information model from the ARIS views. SWRs are described with a series of attribute-value pairs. The *attribute* is a subclass of *rdf:Property* and its value is equal to *owl:Thing*. The basic ontology for SWR is very simple and intuitive, as shown in Table 1:

Table 1. Basic Concept Definition

:BusinessFunction	a	rdfs:Class, owl:Class.
bf:Attribute	rdfs:subClassOf	rdfs:Property;
	rdf:domain	:BusinessFunction;
	rdf:range	bf:AttributeValue.
bf:AttributeValue	owl:equal	owl:Thing.

On top of a simple framework like that, we define further rules and relations to promote the process management in the following aspects:

– **Reusability.** The *subClassOf* relation in RDFS[4] and OWL[5] does not imply the similar semantics as the inheritance relation in Object-Oriented Programming, since all the attributes and properties are defined at instance level and will not be retained through the *subClassOf* relation. We argue that this will lead to limitations on reusability. We propose to define relationships between SWR concepts and maintain a hierarchy of them. In order to achieve this, we firstly define a basic relation named *variantOf* on the resources. A resource r1 is *variantOf* a resource r2 if r1 is an instance *(rdf:type)* or a subclass *(rdfs:subClassOf)* r2, formally, for all resources r1 and r2:

$$r1\ swr: variantOf\ r2 \iff r1\ rdfs: subClassOf\ r2\ ||\ r1\ rdf: type\ r2 \quad (1)$$

We define extended relations, i.e., *extends, specifies* based on the mentioned basic relation.

Given two SWR S_1 and S_2, we denote A_1 and A_2 as their set of attributes, respectively; and for any attribute a in SWR S, we denote $S(a)$ as the attribue-value of a in S. We say S_1 *specifies* S_2 *iff*:

1. all the attributes of S_2 are also attributes of S_1,
2. for every shared attribute a_i, $S_1(a_i)$ is either equal or *variantOf* $S_2(a_i)$,
3. there exists at least one shared attribute a_j such that $S_1(a_j)$ *variantOf* $S_2(a_j)$.

Or more formally:

$$\begin{aligned}&S_1\ specifies\ S_2 \iff\\ &\{\forall a_i \in A_1 | ((a_i \in A_2) \& (S_1(a_i) = S_2(a_i)) || (S_1(a_i)\ variantOf\ S_2(a_i))\} \& \\ &\{\exists a_j \in A_1 | S_1(a_j)\ variantOf\ S_2(a_j)\}.\end{aligned} \quad (2)$$

Similarly, we say S_1 *extends* S_2 *iff*:

1. all the attributes of S_2 are also attributes of S_1,
2. for every shared attribute a_i, $S_1(a_i)$ is either equal or *variantOf* $S_2(a_i)$,
3. there exists at least one attribute a_j in S_1 that is not an attribute of S_2.

---

[4] http://www.w3.org/TR/rdf-schema/

[5] http://www.w3.org/TR/owl-features/

Or more formally:

$$\begin{aligned}&S_1\ extends\ S_2 \Longleftrightarrow\\&\{\forall a_i \in A_1 | ((a_i \in A_2) \& (S_1(a_i) = S_2(a_i)) || (S_1(a_i)\ variantOf\ S_2(a_i))\} \&\\&\{\exists a_j \in A_1 | a_j \notin A_2\}.\end{aligned} \tag{3}$$

Then, a process designer is able to reengineer on a process model rapidly by leveraging these relations and inherit from a proper parent in the hierarchy of SWR classes.

– **Granular-diversity.** Unlike the mechanism used in the SUPER[6] project and sBPMN [2] that treats all process activities/objects as the instances from the ontology, we distinguish between abstract process elements and concrete process elements, since we can inherit attributes from their "super" classes. For instance, from the functional view in the ARIS house, SWRs representing the Business Function (BF) can be either abstract categories or concrete offers, which are distinguished by the *Consumability* attribute. Concrete consumable BF offers have their value of *Consumability* attribute set to *true*.

– **Dynamicity.** Instances of *swr:AttributeValue* can be statically specified. However, in realistic and dynamic scenarios we cannot assume that this is the case. *AttributeValue*s are often: (1) context-dependent (e.g., availability depends on the customer location), (2) sensitive from the business perspective (e.g., insurance price is available after multi-step interaction), (3) dynamic (e.g., currency exchange rates). Therefore, in many cases instances of *swr:AttributeValue* have to be obtained on-the-fly. We support this scenario in the twofold manner.

  We describe the features of an attribute-value in the *desc* namespace. Attributes can be derived using a *desc:Specification* that describes the rules of computing an *swr:AttributeValue* based on inputs. On the other hand, some *swr:AttributeValues* have to be dynamically obtained by interacting with an external entity without explicitly defined rules, i.e., they are either unintended to be exposed (e.g., insurance calculating formulas) or unable to be defined statically (e.g., concurrency exchange rates). We make use of the data-fetching mechanism [15] to retrieve detailed and real-time information (after analyzing consumer parameters, constraints, etc.) about a given resource. Notice that these dynamic parts should not cause any real world state changes, as invocation of these data-fetching interfaces may happen outside of the execution phase and are used for analytical or discovery purposes only. Details in the *desc:DataFetching* rely on the available technical implementation, e.g. service interface for the Web services and URIs for the RESTful services.

[6] http://www.ip-super.org

3.2 Extension Mechanism

Two extension mechanisms are defined in the BPMN 2.0 specification. A process element can be either extended by *Extension* element or external *Relation.* Both methods can be used for our purpose. Their suitability varies in the category of the extended process element: *Flow objects*, e.g., *Subprocesses*, *Activities* and *Tasks*, can only be extended using *Extension* element according to the BPMN 2.0 specification, on the other hand, *Artifacts* like *Group*, *Data Object* and *Text Annotation* can only be extended with the *Relation* element. To define a BPMN *Extension* element, one must specify a boolean-valued attribute named *mustUnderstand* for it. The attribute indicates the essentiality for a process engine to correctly parse this extension. A *definition* attribute of the extension tells the URI of the extension definition, where the schema of the extended information is specified. Finally these extended elements and attributes can be added to the BPMN document by referencing to their paths/query names.

For the *Artifacts*, we can use the *Relationship* element to link them to their extended information model. Adding an external *Relationship* is straightforward. A relationship between *Artifacts* exists in a non-intrusive manner (i.e., without affecting the nature of the artifacts), it can be realized simply by indicating a *Relationship* element with attributes like *type*, *id* and *direction*, sub-elements named *source* and *target* will give the relative references.

We propose to extend the process models in BPMN 2.0 in the following views:

- **Organisation View** defines the roles, the participants, the organizational entities and the relationships between them. BPMN 2.0 *Activities* should be annotated with organizational information like: who is responsible with executing the activity (individual or group), which organization/group does he belong to and what role dose he have. Similarly, the meaning of BPMN 2.0 *PartnerRole*, *Participant* and *PartnereEntity* should be leveraged from a basic string label to that given by concepts from dedicated organizational ontologies for roles, participants and organizational entities respectively.
- **Data View** describes the information objects, the concepts in the messages exchanged by process tasks. Many BPMN 2.0 elements are using data which is encapsulated in the BPMN *ItemAwareElement.* There are also many BPMN elements related to *Events* and message flows that carry data with them. They should be annotated with concepts from domain specific ontologies.
- **Control View** describes the control flow of BPs that carry out the business operations inside the enterprise and it is an integration view between the other views. During business process execution, information like the occurrence time of an activity or event are very important for monitoring and auditing. BPMN *Activity*, *Event* and *EventDefinition* instances should carry with them such information.
- **Function View** defines the functions required to satisfy the objective of the enterprise. BPMN *Activity* should be defined as a domain specific business function instead of just a label.

3.3 Use Case Scenario

We demonstrate our approach with a use case in a pizza ordering scenario[7]. The process flow is shown in Figure 3. According to what we proposed in 3.2, Pool and Roles in the Lanes can be extended to link with the Organizational View, Tasks can link to the Functional View, Events can link to the Control View and finally Messages to the Data View. Due to the limit of space, we will take only the extensions for the Functional View as an example to detail the methodology we use. Extending to the other views will not have a major difference. More specifically, we will elaborate how the informational model of the *Provide ingredients* functionality is built and how it is associated with the task. We will also detail the extension mechanism through *Relation* elements without the functional description of the group element.

Building the Functional Model. To provide customers with correct information about pizza ingredients, the clerk will use an internal service to query on the available servings. We build a hierarchy for the query functionality with a top level category (coarse-grained) and a concrete querying service (fine-grained) for the specific pizza shop. The top level category for the ingredient querying function is a *QueryService*. Table 2 shows a part of the *QueryService* Business Function.

Table 2. Query Service Functional Model

SWR	Attribute	AttributeValue
strg:QueryService	a	ser:ServiceVariant;
	swr:extends	ser:ServiceVariantRoot;
	ser:consumability	xsd:Boolean;
	ser:hasCoParam	strg:QueryRequest;
	ser:hasProParam	strg:ListOfGoods;

The "*swr*" namespace refers to the SWR basic ontology, "*strg*" describes the ontology for the storage of goods, "*ser*" refers to the general service ontology. The *swr:extends* attribute indicates its inheritance from an upper level service category, the *root* service. *bf:consumability* is a boolean value to indicate that the service is abstract and is agnostic to the concrete consumability, its children in the class hierarchy that specifying this attribute will tell if it is directly consumable. The *CoParam* and the *ProParam* refer to the consumer parameter and provider parameters, respectively. The provider of the business functionality is expecting a query request from the consumer, and will provide a list of goods based on the request. The request message and response message can also be a part of the information model in the data view, and hence we have a information model interlinked from different views. The abstract business function category can be further inherited to model a detailed category, namely *IngredientQueryService*, with which we use to model the *Provide ingredients* task. Table 3 shows a part of the functional model.

[7] Adapted from a similar one available at
`http://www.omg.org/spec/BPMN/2.0/examples/PDF/`

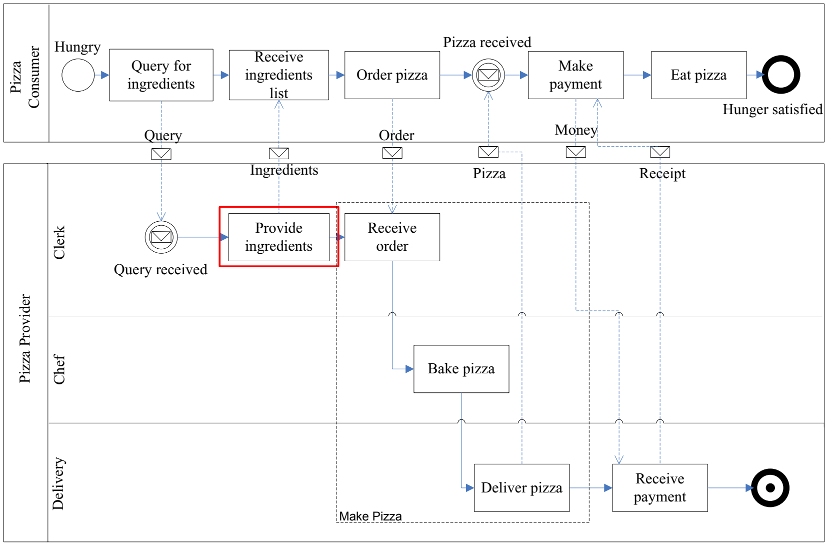


Fig. 3. The Pizza Purchasing Ordering Process Model. The process begins when a customer feels hungry, then he will pick up the phone and make an pizza order, a clerk answering the telephone in the shop will check the ingredients for the customer. According to the list of ingredients provided, the customer will order his pizza. When the clerk receives the order from the customer, he will inform the baker to bake to pizza according to the customer's demand. The delivery boy takes the pizza when ready and delivers it to the customer. The customer will pay for the pizza on receiving it, and the receipt will be provided by the delivery boy. After consuming the pizza, the process ends.

Concrete service offers should indicate the information about the service providers, and such information can also be part of the organizational view. In our case, the service provider can be specified to a person, department, or pizza company etc. According to our defined semantics for the *extends* and *specifies* relationship, only the different part needs to be described comparing from the abstract functional model. The *consumability* is set to "true" to indicate that this concrete service is consumable. The *pizza:Ingredients* resource is an instance of the *ProParam* in the abstract model, its dynamicity is declared with the type of *desc:DynamicParameter*, in which the namespace "*desc*" is an ontology for describing the details of the parameters used in SWRs as we mentioned before. Dynamic parameters will have a binding element whose value is a segment of *SPARQL* construction query. The segment is expected to be utilized by the application to build the concrete instance of the resource at run-time, based on inputs either directly from user (using *desc:specification* we described before) or from the result of the data-fetching mechanism.

The last element in the table is the data-fetching element that specify the ways of retrieving the data, e.g., service endpoints, interfaces etc. Data-fetching is used here to provide real-time and on-demand details for queries from process analyzer

Table 3. Pizza Ingredient Query Functional Model

SWR	Attribute	AttributeValue
strg:IngredientQueryService	a	ser:ServiceVariant;
	ser:hasProvider	strg:QueryServiceProvider;
	ser:extends	strg:QueryService;
	ser:consumability	xsd:true;
	ser:hasProParam	pizza:Ingredients.
pizza:Ingredients	a	strg:ListOfGoods, desc:DynamicParameter;
	desc:hasBinding	"some_sparql_construct_pattern"
	desc:hasDataFetching	:IngredientQueryDataFetching.

or discovery agent, without invoking the process and having post conditions. The technical information contained in *IngredientQueryDataFetching* for executing the data-fetching will not be detailed due to the space limit.

Referencing from BPMN. Referencing to the external business function models with an *Extension* requires a extension definition schema for an XML attribute, which can be specified as:

```
<xsd:attribute name='hasBusinessFunction' type='xsd:anyURI'/>.
```

We can attach this extended attribute to a *task* while having indicated the extension element for the process, as shown below:

```
<extension mustUnderstand="true" definition="ext:hasBusinessFunction" />
<process>
<task id='id-t3' name='Provide ingredients'
ext:hasBusinessFunction='strg:IngredientQueryService'>
...
</task>
...
</process>.
```

As the readers may have noticed, three tasks in the use case are grouped together with a dotted box, indicating they serve as sub-functionalities of the "make pizza" business function. Logically this may be a bad example, however, the intent is to demonstrate how we could extend the *Artifacts* using a *Relationship* element. We can define a relationship with the id "er-1" and relate the group "group1" to the business function "MakePizza" as follow:

```
<relationship type='swr-reference' id='er-1' direction='both'>
<documentation>
Reference from a group of activities to their information model from the
functional view
</documentation>
<source ref='src:group1'/>
<target ref='strg:MakePizza'/>
</relationship>
```

4 Related Work

We propose SWR to be a lightweight advancement for describing concepts with RDFS and OWL ontologies, respecting the concreteness and dynamicity of the concepts. SWR is similar to WSML *concept* [13] which also introduces inheritance between classes. However, WSML focus on the service domain and is considered "heavy weight" as it is equipped with explicitly defined axioms, relations and different mediators. However it's not standardized and does not benefit from linked data.

A relevant research work to ours is the sBPMN [2], it has defined semantic variants of BPMN language. It used a WSML ontology language [4] to semantically describe business processes. It aimed to provide the Semantic Business Process Management (SBPM) [6]) framework. However, the approach taken by the project resulted in the alteration of language specification and has not been standardized. In our work, we use an extension mechanism of BPMN 2.0 which is compatible. Another work of BPEL4SWS [9] follows the similar principles we use to extend a "hosting" language with existing mechanisms to enable the semantic relations non-intrusively. However it is mainly concerned with the semantical process executions, including semantical service discovery and composition. While we are also interested in providing dynamic information about the process model, thus more concrete and up-to-date information can be realized during process analysis and discovery.

Other approaches such as those proposed in [7,12,5] proposed BPMN extensions while considering some particular perspectives. [7] proposed a BPMN extension with business process goals and performance measures, [12] proposed a BPMN extension for modeling security requirements and [5] proposed a BPMN extension with a resource information layer. However, our work is aligned with the ARIS views and categorize all the essential information in the enterprise applications accordingly. It supports better integrations and collaborations from different perspectives of the enterprise systems.

5 Conclusion and Future Work

In this paper we justify the need of providing semantics to the process models from different perspectives according to ARIS architecture. We elaborate our novel approach of leveraging the BPMN 2.0 extension mechanism to introduce a compatible way of establishing the semantic links for process models in BPMN 2.0. We also propose a *StructuredWebResource* framework to model the information related in the enterprise applications and provide means to enhance the reusability, granular-diversity and dynamicity. We argue that we can make improvements with semantic technologies not only in the execution phase but also during other phases in the BPM life cycle. As a future work we plan to expand the relations in SWRs for more complex resource clouds beyond class hierarchies.

References

1. Abramowicz, W., Haniewicz, K., Kaczmarek, M., Zyskowski, D.: Semantic modelling of collaborative business processes. In: eKNOW (February 2009)
2. Abramowicz, W., Filipowska, A., Kaczmarek, M., Kaczmarek, T.: Semantically enhanced business process modeling notation. In: SBPM (2007)
3. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. Int. J. Semantic Web Inf. Syst. 5(3), 1–22 (2009)
4. de Bruijn, J., Lausen, H., Polleres, A., Fensel, D.: The Web Service Modeling Language WSML: An Overview. In: Sure, Y., Domingue, J. (eds.) ESWC 2006. LNCS, vol. 4011, pp. 590–604. Springer, Heidelberg (2006)
5. Großkopf, A.: An extended resource information layer for bpmn. Systems Engineering, 1–17 (2007)
6. Hepp, M., Leymann, F., Domingue, J., Wahler, A., Fensel, D.: Semantic business process management: a vision towards using semantic web services for business process management. In: ICEBE (October 2005)
7. Korherr, B., List, B.: Extending the epc and the bpmn with business process goals and performance measures. In: ICEIS (3), pp. 287–294 (2007)
8. Maciaszek, L.A.: Roundtrip architectural modeling. In: Proceedings of the 2nd Asia-Pacific Conference on Conceptual Modelling, vol. 43. APCCM (2005)
9. Nitzsche, J., van Lessen, T., Karastoyanova, D., Leymann, F.: BPEL for Semantic Web Services (BPEL4SWS). In: Chung, S., Herrero, P. (eds.) OTM-WS 2007, Part I. LNCS, vol. 4805, pp. 179–188. Springer, Heidelberg (2007)
10. OMG: Business Process Model and Notation (BPMN) Version 2.0 Beta 2 (2010), http://www.omg.org/spec/BPMN/2.0
11. Pedrinaci, C., Domingue, J., Brelage, C., van Lessen, T., Karastoyanova, D., Leymann, F.: Semantic business process management: Scaling up the management of business processes. In: ICSC (2008)
12. Rodríguez, A., Fernández-Medina, E., Piattini, M.: A bpmn extension for the modeling of security requirements in business processes. IEICE Transactions 90-D(4) (2007)
13. Roman, D., Keller, U., Lausen, H., de Bruijn, J., Lara, R., Stollberg, M., Polleres, A., Feier, C., Bussler, C., Fensel, D.: Web service modeling ontology. Applied Ontology 1(1) (2005)
14. Scheer, A.-W., Schneider, K.: Aris architecture of integrated information systems. In: Handbook on Architectures of Information Systems (1998)
15. Zaremba, M., Vitvar, T., Moran, M.: Towards optimized data fetching for service discovery. In: ECOWS (November 2007)

Integration of UI Services into SOA Based BPM Applications

Jörg Hohwiller and Diethelm Schlegel

Capgemini, CSD Research, Berliner Str. 76, 63065 Offenbach, Germany
{joerg.hohwiller,diethelm.schlegel}@capgemini.com
http://www.de.capgemini.com

Abstract. A service oriented architecture (SOA) combined with business process management (BPM) puts enterprises in a position to build applications in a very flexible and agile way. The SOA makes services available and BPM suites (BPMS) combine these services with user tasks to complex business processes. Integration of data and application services is supported by common technical standards and a wide range of products that are applicable out of the box. In contrast, the number of standardized user interface (UI) components that can be composed in SOA based BPM applications in a straightforward way is quite low. This is because the demands on such components are much higher. They must be embeddable into graphical front ends and interact with end users as well as the underlying business models. Unfortunately, there is no generally accepted approach and BPM tools provide their own proprietary mechanisms. The main drawback is the lack of re-usability resulting in higher cost and longer development phases. This paper derives basic requirements for UI components in order to make them generally applicable for SOA based BPM applications. It defines UI services meeting these demands in a platform-independent manner and presents one possible implementation based on JSR-286 Portlets and WSRP 2.0. This solution allows the development and embedding of system-independent and re-usable UI services by slightly extending existing BPM suites and enterprise portals.

Keywords: BPM, BPMS, BPMN, SOA, UI-Service, UI-Integration, Portal, Portlet.

1 Introduction

Market conditions are changing increasingly fast. This leads to growing demands on enterprise application landscapes. To be able to react faster on new requirements, function oriented systems are replaced by process related approaches resulting in a move from isolated legacy applications to service oriented architectures (SOA). In doing so, monolithic systems are divided into services ([1]) and grouped into components and domains.

It should be noted that according to [2] the definition of service landscapes is done independent of the underlying technologies. Nevertheless, the following

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 53–64, 2011.

realization will be based on some available standards and products like web services (WS), enterprise service bus (ESB), or enterprise application integration (EAI) frameworks.

Enterprises introduce business process management (BPM) in order to leverage high flexibility and agility. The use of BPM suites (BPMS) allows the integration of services and user tasks to business applications. Thereby, business processes are modelled and simulated in the BPMS. They can be executed as long as the service orchestration is based on the same technology stacks as used by the SOA realizations.

Data and application services are very well supported by a large number of products following several standards like REST ([3]), SOAP ([4]), WSDL ([5]), UDDI ([6]), enterprise service bus (ESB), enterprise application integration (EAI) and so on. This is reflected in a growing range of business related web services. They are even provided in the cloud, e.g. Amazon Web Services (`http://aws.amazon.com`) and Salesforce (`http://www.salesforce.com`).

In contrast, the integration of user interfaces (UI) in the area of SOA and BPM is more difficult by far. [7] describes the main challenges concerning user interfaces based on SOA and BPM including the heterogeneity of existing application landscapes as well as the absence of approved design patterns. [7] proposes the use of so-called dialogue services but gives no explicit definition. So, there is hardly any standardized solution based on them.

As a result, user interaction is often exclusively realized by the BPM suites themselves. The most common technique is the design of proprietary forms on the basis of business objects. Only a few products support more far-reaching standards like portlets or gadgets. Nevertheless, business process integration as well as the construction of comprehensive user interfaces still highly depend on the selected BPM product.

Graphical interface components can rarely be re-used and must be individually implemented leading to longer development time and higher cost. Based on a survey, [8] expects that the development of user interfaces takes about 50 percent of time. Later on, their maintenance still requires about 40 percent.

This just goes to show the importance of standardized and re-usable UI services that can easily be integrated into BPM suites. They are the prerequisite for product vendors and cloud providers to offer a growing number of user interface components in addition to data and application services. The availability of such UI components, in turn, will accelerate the implementation of user interfaces. To achieve this vision, UI services should satisfy the following conditions:

C1 In the SOA, UI services are accessible by common standards (e.g. SOAP, REST, WSDL and UDDI).

C2 In the BPMS, UI services can easily be integrated in the same way as data or application services.

C3 UI services can be integrated into enterprise portals as well as individual dialogues in a simple and straightforward way.

C4 There is an synchronization mechanism between UI services and business processes (UI$\Rightarrow$Process and Process$\Rightarrow$UI).

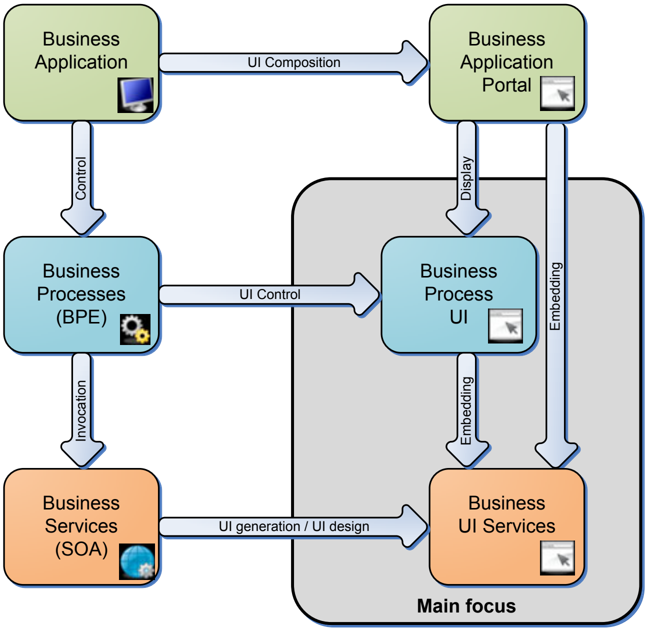


Fig. 1. UI Service Integration

The reason for these conditions with respect to the combination of UI services and SOA based BPM applications is illustrated in figure 1. The left column corresponds to components that are well supported by data and application services.

At the lowest level, the SOA will be supplemented by UI services according to condition C1. Mostly, they relate to existing services and are derived by UI generation or by manual UI design.

Building on this, the business process models are executed by the business process engine (BPE). The BPE invokes the services and controls the UI flow according to the executed process. As demanded by C4, this includes opening and closing modal dialogues as well as event handling and more sophisticated changes in existing user interfaces. In accordance with C2, the business process UI may embed one or more UI services.

At the highest level, there is the business application with its enterprise portal that is implemented by UI composition. It is possible to include UI services and to display business process UI if needed (condition C3). Thereby, the application controls the BPE, e.g. by starting processes or by accessing the currently running tasks.

The grey box shows the main focus of this paper concerning UI service and their integration. The following section 2 presents existing approaches in the area of service orientation and user interfaces. Thereafter, UI services are defined in a technology-independent way in section 3 followed by a possible implementation based on portlets (section 4). At the end, there is a conclusion and an outlook (section 5).

2 Related Work

A lot of research has been done in the area of graphical user interface development. Thereby, attempts have been made to reduce the complexity by applying component based methods and, at the same time, increase the re-usability of existing UI components. For example, JSF ([9]) represents a step in this direction.

There are some other proposals for **component based UI development**. WebRatio uses a domain specific language called WebML ([10]) to integrate web services whereas VisualWADE is based on an object-oriented notation ([11]). [12] realizes UI composition via annotations and semantic event exchange but has no service oriented background. The CRUISe integration system ([13]) is a SOA approach. It defines UI services using a description language. These UI services are made available by a registry and are dynamically combined to UI mash-ups.

In the area of **web mash-ups**, integration is done in two ways. At first, the connection is established on the data layer. In most cases, this is achieved by distributing events though pipes like RSS feeds ([14], [15]). [16] additionally defines an universal component model that contains all available services (data, application, and UI). Other proposals use the data federation pattern ([17]). Additionally, UI integration is done by combining several web pages to a composite user interface ([18], [15]).

The ServFace Builder uses **UI generation** for the design of web applications ([19]). It automatically produces UI components on the basis of annotated web services. The runtime engine realizes the interaction between individual components. A similar solution is described in [20]. Hereby, the UI component orchestration and behaviour is specified by the business process execution language (BPEL, see [21]).

The so-called **distributed user interface orchestration** extends BPEL for defining user interface specific details ([22] and [23]). The UI components are constructed by compiling the enriched BPEL models.

In 2003, **portlets** were specified by extending servlets ([24]). They produce mark-up code parts that can be inserted into complete HTML pages. Integration is accomplished by so-called portal servers. [25] compares selected frameworks. Unfortunately, there were many restrictions such as proper communication between portlets and integration of JavaScript/AJAX. This prevented a wide-spread application of this standard. In 2008, version 2 of the specification ([26]) was released addressing these shortcomings by adding events, proper AJAX integration and an overall UI component model.

A main disadvantage of portlets is the restriction to Java based platforms. **Web services for remote portlets** (WSRP) ([27] and [28]) make portlets suitable for non-Java environments. As WSRP is based on WSDL it can easily be integrated into service oriented landscapes. While the first version of WSRP includes only basic functionality, the second one adds important features like event handling.

All the proposals listed above have their focus on building more or less self-contained web applications. They do not offer generic interfaces for embedding UI services into BPMS or enterprise portals in such a way that the UI can be controlled by a business process engine (BPE). All the same, several of them can be used as the technological foundation for the integration of UI services into SOA based BPM applications. This will be shown for portlets (JSR-286 or WSRP) in section 4.

3 UI Services

Section 1 listed four conditions UI services should satisfy when they are integrated into BPMS or enterprise portals. The objective of these requirements is to make sure that the resulting interface provides a seamless user experience.

In this section, UI services are defined in a platform-independent and technology-neutral way. Based on this, compliance with the conditions is deduced. UI services are characterized by their interface that must support a number of functions needed for embedding and controlling of the UI service component.

Definition 1. *A* UI service *denotes a service in the sense of a SOA (see [1]) that supports the following interface functionality:*

- ***Life cycle (LC)*** *it is possible to handle the complete life cycle of UI services with respect to its presentation (e.g. created, closed, normal, modal, minimized, maximized, hidden).*
- ***Rendering (RE)*** *the UI service is able to render itself depending on its life cycle status (for example by producing mark-up code to be integrated in portal web pages). This method highly depends on the technology stack.*
- ***Event handling (EV)*** *events can be sent as well as received by the UI-service in order to communicate with other components (for example, when the user clicks on buttons or changes the selection). Events usually contain additional information.*

It is noted that the precise design of the interface functionality depends on the selected technical basis. The actual realization is assumed to be compliant with the corresponding requirements.

It must be shown that UI services fulfil the four conditions (see section 1). Because they belong to the services of a SOA, the first two conditions C1 and C2 apply as soon as they are implemented using suitable technical foundations. The life cycle and rendering functionalities allow the integration into enterprise portals (condition C3). The synchronization required by condition C4 can be implemented based on the event handling mechanism.

Some portal frameworks and few BPM suites use technical standards that already provide the relevant features with regard to life cycle, rendering, and event handling. Others support additional features like access to shared variables (states) that will simplify UI service implementation in certain cases. However, it is not crucial because it can be realized by events. For example, reading shared variables can be accomplished by sending READ-events that are answered by a VALUE-events. Writing variables works in a similar way.

3.1 Architecture Overview

Figure 2 shows an overview of an architecture that embeds UI services into a SOA based BPM application landscape. At top level, the user interacts with the enterprise portal that integrates BPMS control and one or more individual UI services. The BPMS provides services for process management and task handling as well as manually designed forms. The business process engine (BPE) in turn calls services that are available in the SOA. This includes UI services because they meet the four conditions so that they can be accessed by common standards and integrated into BPM suites and enterprise portals similar to application services.

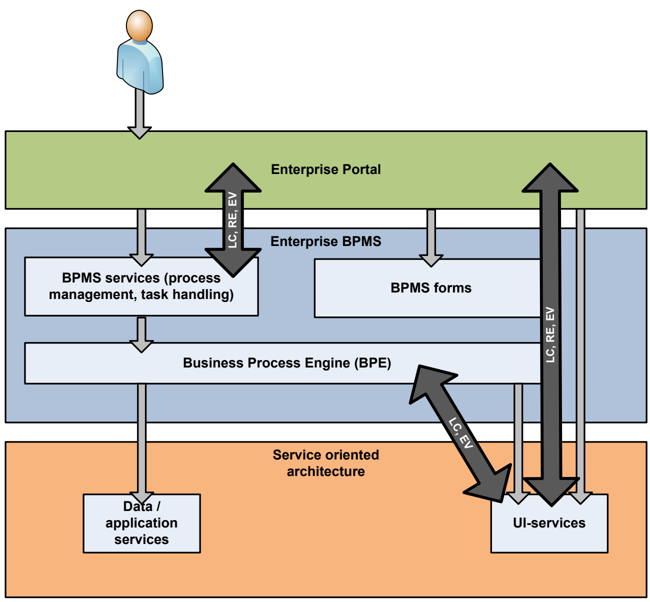


Fig. 2. Architecture Overview

Embedding of UI services into the enterprise portal needs the additional functionality described in the definition above. It relies on life cycle management (LC) to create UI components in the corresponding status as well as on rendering (RE) to display the component inside portal pages. Furthermore, the event handling (EV) must be supported by both UI services and enterprise portal. The portal framework is responsible for event distribution.

UI services synchronization with the BPE is realized by two mechanisms. First, it is possible to influence the UI behaviour by using life cycle (LC) functionality. For example, the business process will be able to open a modal dialogue or to bring certain UI components to the foreground. On the other hand, the BPE is able to synchronize with UI services using event handling (EV). A possible application is a process step that waits on user interaction. In this case, the UI service generates an event that is received by the BPE.

Many enterprise scope BPM suites already have integrated portals. As a consequence, there is no separated enterprise portal layer. The user interacts with the BPMS that controls the business processes and coordinates UI integration. With respect to the UI services, this makes no difference and the statements remain valid. The functionalities described above allow the seamless integration in this situation, too.

3.2 Synchronization

This section defines approaches for synchronizing business processes and UI services by using an appropriate process modelling notation. While many BPMS bring their own, an increasing number of tools support standardized notations. This paper uses BPMN 2.0 ([29]) because it is a very popular, well-known standard and, on the other hand, already contains elements needed for the integration of user tasks and services. It is even possible to express more technical aspects like escalations, timers, conditional branching, parallelism, and synchronization.

Figure 3 illustrates how business process models initiate actions in the user interface. The complete process model is represented by a pool that is divided into

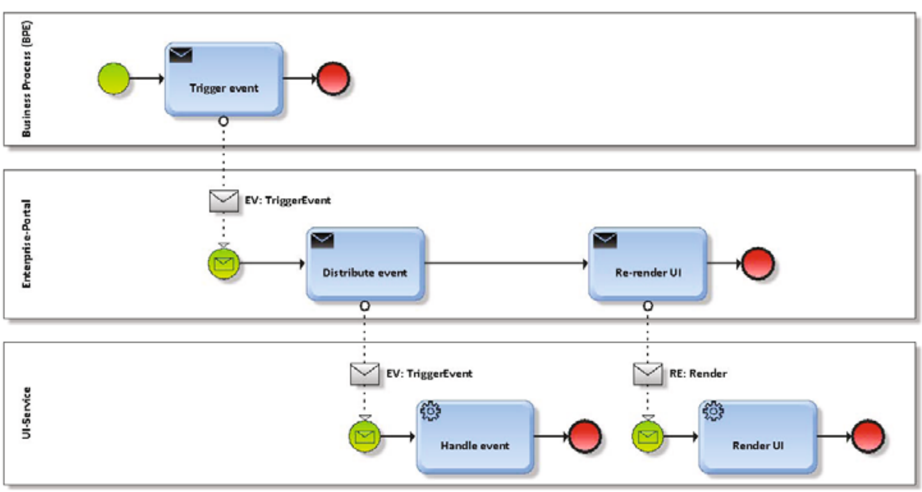


Fig. 3. Business Process Event

three lanes. The upper one corresponds to the actual business process whereas the two other have a technical background showing the operations within the enterprise portal and the UI service. Usually, the notation for sending the event message is sufficient from a business perspective and the technical lanes can be simplified or even completely omitted. Here, they are added for explaining the detailed sequence of interactions.

The business process triggers one or more UI services by sending an event to the enterprise portal. The enterprise portal distributes this event to the appropriate receivers and finally re-renders the user interface to let possible changes take effect.

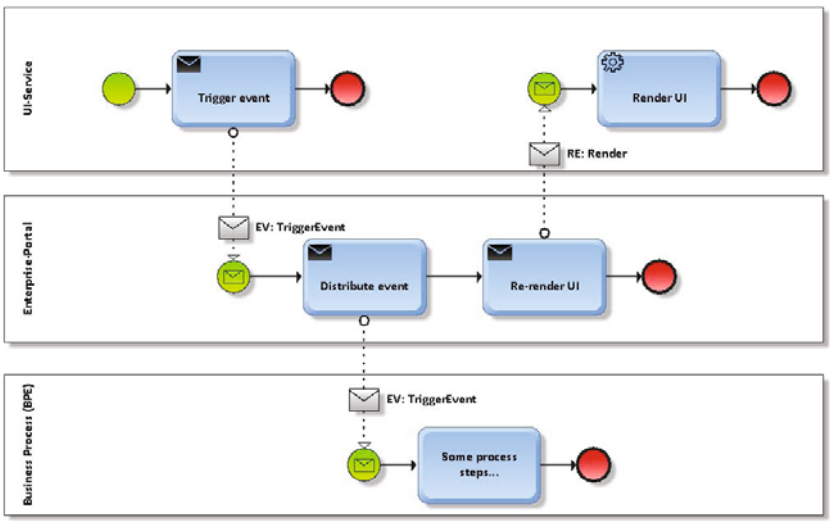


Fig. 4. UI-Service Triggers BPMS

Synchronization works in reverse as well (see figure 4). The UI service sends the event to the portal server that distributes it to other components including the BPE. So, the business process is able to receive it and proceed with any further steps. In this case too, the portal re-renders the user interface because the event could have influence on other UI services. Normally, the technical lanes may be omitted as proposed above.

Figure 5 illustrates the possibility to open UI services as modal dialogues. The business process sends a life cycle message to the appropriate UI service and waits until the close event is received. Finally, the UI service is cleaned up by using another life cycle message. Please note that the model is simplified for two reasons. First, the enterprise portal lane is omitted as suggested, and secondly, there is no error handling in the case of failed operations.

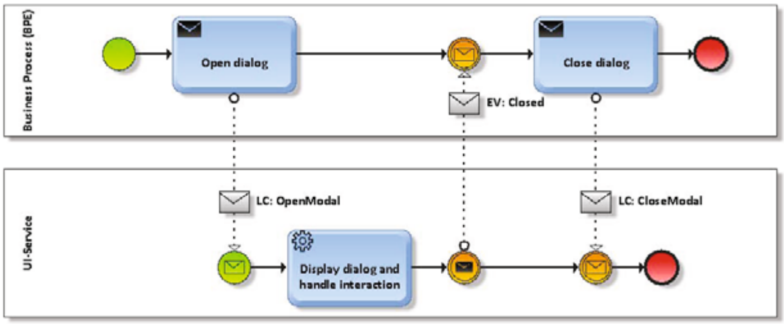


Fig. 5. BPMN Model for Modal Dialogues

4 UI Services with JSR-286 and WSRP 2.0

In section 3, the concept of UI services is introduced in a platform-independent manner. Thereby, some approaches that were pointed out in section 2 could serve as basis for a technical realization. This paper focusses on Java portlets in accordance with JSR-286 or WSRP 2.0. The reason for doing so is that portlets and WSRP are standardized and well-known mechanisms. They are supported by many products in the area of SOA and BPM. First of all, it must be examined whether Java portlets and WSRP correspond to the definition of UI services. This is achieved by inspecting their specifications:

- **Life cycle:** WSRP 2.0 provides life cycle functionality by defining operations of the following categories: service description, registration, and portlet management. They include initialization, cloning, configuration, and destroying of portlets. JSR-286 specifies several life cycle methods, especially *init()*, *processAction()*, and *destroy()*.
- **Rendering:** WSRP 2.0 implements rendering support by the operation *getMarkup* while JSR-286 contains the method *render()*. The return values are code fragments that can be included into portal web pages.
- **Event handling:** WSRP 2.0 defines event handling using two arrays (*publishedEvents* and *handledEvents*) together with the operation *handleEvents*. JSR-286 declares the relevant events in the *portlet.xml* deployment descriptor and supplies the *processEvent()* method. In both cases, new events are generated within life cycle and rendering functions.

Altogether, it is obvious that WSRP 2.0 and JSR-286 implement the interface functionality of UI services. This means that the four conditions for a seamless UI service integration are fulfilled. Moreover, they even offer further concepts like shared states, resource handling, and AJAX support.

Figure 6 gives an overview of the resulting SOA based BPM application. The basis consists of a landscape of web service and WSRP 2.0 or JSR-286

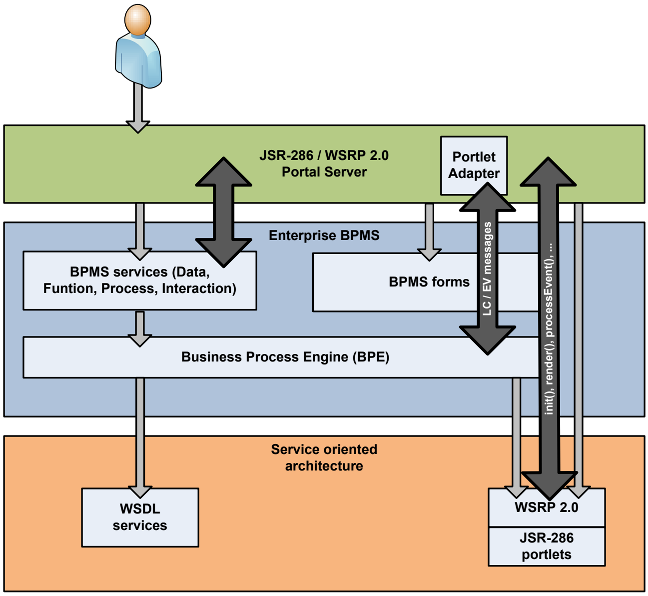


Fig. 6. Implementation Overview

portlets (SOA). The enterprise scope BPMS allows the business processes to invoke the relevant web services.

The portal server is based on WSRP 2.0 or JSR-286 and integrates BPMS control and SOA portlets. This assumes that the BPMS provides appropriate functionalities based on web service and portlets. The interaction between BPE and UI services is realized by a so-called *portlet adapter* that translates the portlet functionality for use in the BPE.

Most components are available as standard solutions. There are two important elements that may be absent: (1) the BPMS service interface and (2) the portlet adapter. In this case, they have to be developed as an extension.

5 Conclusion and Further Work

This paper has presented a solution for making UI components available as services that can be integrated into SOA based BPM applications. At first, some requirements concerning the integration of user interfaces have been explained. After providing an overview of existing approaches, UI services have been defined in a platform-independent way so that the demands are met. An architecture overview and synchronization methods between business processes and UI services have been derived. So, the suitability of implementations can easily be evaluated by verifying the compliance to the UI service definition.

Finally, the paper has proposed a possible realization on the basis of JSR-286 or WSRP 2.0 portlets. The benefit of this solution is that most of the components are already available in the form of standard products. Thereby, it has been shown that portlets are compliant to UI services.

As a next step, the theoretical results will be applied in practice. It is planned to implement sample systems using several BPM suites and portal servers. Besides that, more research is needed in the area of alternative technical platforms. Investigations must be carried out into how the concept of UI services can be realized by mobile platforms like smart phones or tablet computer.

References

1. Engels, G., Hess, A., Humm, B., Juwig, O., Lohmann, M., Richter, J.P., Voß, M., Willkomm, J.: A method for engineering a true service-oriented architecture. In: Cordeiro, J., Filipe, J. (eds.) ICEIS, vol. (3-2), pp. 272–281 (2008)
2. Engels, G., Hess, A., Humm, B., Juwig, O., Lohmann, M., Richter, j.P., Voß, M., Willkomm, J.: Quasar Enterprise. dpunkt.verlag (2008)
3. Fielding, R.T.: Architectural Styles and the Design of Network-based Software Architectures. PhD thesis, University of California, Irvine, Irvine, California (2000)
4. Gudgin, M., Hadley, M., Mendelsohn, N., Moreau, J.J., Nielsen, H.F., Karmarkar, A., Lafon, Y.: Soap version 1.2 (2007), http://www.w3.org/TR/soap12
5. Chinnici, R., Gudgin, M., Moreau, J.J., Schlimmer, J., Weerawarana, S.: Web services description language (wsdl) version 2.0 (2007), http://www.w3.org/TR/wsdl20/
6. Clement, L., Hately, A., von Riegen, C., Rogers, T., et al.: Oasis uddi specification (2005), http://uddi.org/pubs/uddi-v3.0.2-20041019.pdf
7. Nandico, O.F.: Integration für den benutzer: Einheitliche benutzungsoberfläche mit interaktionsservices. OBJEKTspektrum Ausgabe 02/2011 (Febuary 2011)
8. Myers, B.A., Rosson, M.B.: Survey on user interface programming. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1992, pp. 195–202. ACM, New York (1992)
9. Burns, E., Kitain, R.: Javaserver faces technology, http://java.sun.com/j2ee/javaserverfaces/
10. Gómez, J., Bia, A., Parraga, A.: Tool Support for Model-Driven Development of Web Applications. In: Ngu, A.H.H., Kitsuregawa, M., Neuhold, E.J., Chung, J.-Y., Sheng, Q.Z. (eds.) WISE 2005. LNCS, vol. 3806, pp. 721–730. Springer, Heidelberg (2005)
11. Acerbis, R., Bongio, A., Brambilla, M., Butti, S., Ceri, S., Fraternali, P.: Web Applications Design and Development with Webml and Webratio 5.0. In: Paige, R.F., Meyer, B. (eds.) TOOLS (46). LNBIP, vol. 11, pp. 392–411. Springer, Heidelberg (2008)
12. Paulheim, H.: Seamlessly integrated, but loosely coupled: building user interfaces from heterogeneous components. In: Proceedings of the IEEE/ACM International Conference on Automated Software Engineering, ASE 2010, pp. 123–126. ACM, New York (2010)
13. Pietschmann, S., Voigt, M., Meiner, K.: Dynamic composition of service-oriented web user interfaces. In: Perry, M., Sasaki, H., Ehmann, M., Bellot, G.O., Dini, O. (eds.) ICIW, pp. 217–222. IEEE Computer Society (2009)

14. Yahoo! Inc., et al.: Yahoo pipes, http://pipes.yahoo.com/pipes/
15. JackBe Corporation, et al.: Jackbe presto, http://www.jackbe.com/products/
16. Daniel, F., Casati, F., Benatallah, B., Shan, M.C.: Hosted Universal Composition: Models, Languages and Infrastructure in Mashart. In: Laender, A.H.F., Castano, S., Dayal, U., Casati, F., de Oliveira, J.P.M. (eds.) ER 2009. LNCS, vol. 5829, pp. 428–443. Springer, Heidelberg (2009)
17. López, J., Bellas, F., Pan, A., Montoto, P.: A component-based approach for engineering enterprise mashups. In: Gaedke, M., Grossniklaus, M., Díaz, O. (eds.) ICWE 2009. LNCS, vol. 5648, pp. 30–44. Springer, Heidelberg (2009)
18. IBM Corporation, et al.: Ibm mashup center, http://www.ibm.com/software/products/de/de/mashupcenter
19. Dannecker, L., Feldmann, M., Nestler, T., Hbsch, G., Jugel, U., Muthmann, K.: Rapid Development of Composite Applications Using Annotated Web Services. In: Daniel, F., Facca, F.M. (eds.) ICWE 2010. LNCS, vol. 6385, pp. 1–12. Springer, Heidelberg (2010)
20. Menge, F.: Generation of user interfaces for service compositions. Master's thesis, Hasso Plattner Institut (2009)
21. Jordan, D., Evdemon, J., et al.: http://docs.oasis-open.org/wsbpel/2.0/wsbpel-v2.0.pdf
22. Lee, J., Lin, Y.Y., Ma, S.P., Lee, S.J.: Bpel extensions to user-interactive service delivery. J. Inf. Sci. Eng. 25(5), 1427–1445 (2009)
23. Daniel, F., Soi, S., Tranquillini, S., Casati, F., Heng, C., Yan, L.: From People to Services to UI: Distributed Orchestration of User Interfaces. In: Hull, R., Mendling, J., Tai, S. (eds.) BPM 2010. LNCS, vol. 6336, pp. 310–326. Springer, Heidelberg (2010)
24. Abdelnur, A., Hepper, S.: Jsr 168: Portlet specification (2003), http://www.jcp.org/en/jsr/detail?id=168
25. Akram, A., Chohan, D., Dong Wang, X., Yang, X., Allan, R.: A service oriented architecture for portals using portlets. In: UK e-Science All Hands Meeting, Nottingham (2005)
26. Hepper, S.: Jsr 286: Portlet specification 2.0 (2008), http://www.jcp.org/en/jsr/detail?id=286
27. Kropp, A., Leue, C., Thompson, R., Braun, C., Broberg, J., Cassidy, M., Freedman, M., Jones, T.N., Schaeck, T., Tayar, G.: Web services for remote portlets specification (2003), http://www.oasis-open.org/committees/wsrp
28. Thompson, R., Allamaraju, S., Freedman, M., Jacob, R., Kramer, A.: Web services for remote portlets specification v2.0 (2008), http://docs.oasis-open.org/wsrp/v2/wsrp-2.0-spec.html
29. OMG, et al.: Business process model and notation (bpmn) version 2.0 (2011), http://www.omg.org/spec/BPMN/2.0

A Novel Generic Clinical Reference Process Model for Event-Based Process Times Measurement

Eva Gattnar[1,2], Okan Ekinci[1], and Vesselin Detschew[2]

[1] Clinical Competence Center Cardiology, Siemens Healthcare,
Allee am Röthelheimpark 3a, 91052 Erlangen, Germany
{eva.gattnar,okan.ekinci}@siemens.com
[2] Institute of Biomedical Engineering and Informatics, Ilmenau Technical University,
POB 100565, 98684 Ilmenau, Germany
vesselin.detschew@tu-ilmenau.de

Abstract. In recent years, performance measurement has become an important element of efficiency improvement projects in many organizations. Thereby, process-based measures are used to evaluate the process efficiency and quality. In health care, such measures are often neither commonly defined nor standardized. Therefore we present a novel approach for standardized clinical quality metrics measurement by means of a newly developed clinical reference process model and generic Key Performance Indicators (KPIs). We use both for a comprehensive description of the complete clinical patient-centered process in hospital. Our approach fulfils performance measurement requirements particularly in the field of time-critical diseases like heart attack and stroke.

Keywords: performance measurement, process monitoring, key performance indicators, clinical process model, event-driven process chain.

1 Introduction

In recent years, performance measurement and in particular process-based quality measures have become important elements of efficiency improvement projects in many organizations. In health care, such process quality measures are often not commonly defined and therefore not standardized. Especially in hospitals, there is a lack of approaches which provide feasible, relevant, meaningful, understandable, non-ambiguous and disease-oriented process quality measures in a standardized manner. Thereby, clinical process quality means the way in which care is delivered and specific clinical requirements are fulfilled [1]. Because of the complexity, which is characteristic in clinical domain, a comprehensive performance measurement approach is required to enable quality controlling and sustainable care quality improvement as well as to achieve a comparative quality reporting [2].

Nowadays, performance measures are very numerous and common in health care. They are defined e.g. by the national indicator library of the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), the National Health Services (NHS) or the Agency for Healthcare Research and Quality (AHRQ). They have all in common that the measures are not commonly defined and therefore not standardized.

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 65–76, 2011.

Furthermore the measures are not disease-oriented which is important regarding a differentiated performance evaluation.

Especially in the field of acute diseases like heart attack and stroke, the factor time is very critical for the quality of care outcome. Often few minutes decide here about the degree of heart muscle or brain damage and with it about the quality of life or even patient's life itself. Therefore, the quality of care can be evaluated using the time factor. In this context, we propose an approach for event-oriented process time measurement based on clinical IT-systems aiming the performance monitoring during the clinical workflow.

Therefore we present a novel way of standardized investigation of Key Performance Indicators (KPIs) by means of a newly developed reference process model, which represents a promising opportunity to support clinical performance measurement. We propose this model for generic description of the complete clinical patient-centered process and quality measurement in hospitals provided by generic clinical KPIs which we derive using clinical indicator libraries and clinical guidelines. Our proposed model provides the basis for a standardized analysis and evaluation of hospital-specific time-critical workflows. Furthermore, the proposed process-oriented view enables a continuous, automated and standardized quality controlling, benchmarking and comparative reporting.

2 Clinical Pathways and Guidelines

Clinical guidelines are systematically developed disease specific statements, which assist practitioners and support physicians as well as health care organizations in the treatment of patients in disease specific circumstances. They aim to improve patient care and limit unjustified treatment variation. As recommendations for clinical diagnostics and treatment of certain diseases they are important for quality of care in hospitals [1]. In most cases they are based on results deduced from recent clinical studies. However, clinical guidelines are not binding and additionally they cannot address the variations in different health care institutions where the guidelines are applied [3]. For example, the German guidelines for treating heart attack recommend the coronary intervention until 60 minutes at the latest [4] while US guidelines recommend a timeline until 90 minutes [5]. Since each hospital has different equipment (e.g. different departments, personnel, imaging devices etc.), it is not possible to follow the guidelines in all cases. Clinical guidelines often leave such organizational conditions under which a certain diagnostic or treatment procedure can be performed.

Because of their universal character regarding the organizational details of health care organizations (e.g. hospitals), clinical guidelines are interchangeable between different hospitals [1]. So-called clinical pathways are in contrast locally limited and individually defined for each hospital. Therefore they have to be adapted to local organizational conditions. This includes their individual adaptation to environmental and personnel conditions in the hospital [6]. Clinical pathways are evidence-based, multidisciplinary treatment processes, which were developed for a certain hospital. They focus on the entire clinical process and aim continuous quality improvement in patient care [6]. In addition, clinical pathways can be used for data collection and

Table 1. Differences between clinical pathways and guidelines

Clinical Pathways	Guidelines
locally limited individually for each hospital multidisciplinary focuses on entire clinical process clear outcome objectives	locally not limited universal specifically focuses on part of clinical process outcome objectives often not clear

process monitoring [7]. They are important in the field of quality management and clinical process standardization in hospitals. Table 1 shows the main differences between clinical pathways and guidelines.

Since the construction of clinical pathways has to be performed for each disease in each hospital anew, it is very time-consuming. From the hospital's point of view, a support of construction and evaluation of clinical pathways is therefore very important. Therefore, the development of clinical pathways should be supported and based on approved clinical guidelines. After realization, the process quality should be evaluated using clinical meaningful measures and indicators. Both efforts are supported by the newly developed clinical reference process model and the generic KPIs, which are presented in the following.

3 Event-Based Process Modeling Language

With regard to the objective of event-based process performance measurement, the Event-driven Process Chain (EPC) is used as process modeling language [8]. The EPC is a popular methodology for a semi-formal description and notation of business processes [9], [10], [11]. Furthermore, it is also applicable for clinical process modeling [12]. However, the EPC notation is not oriented towards clinical tasks and has therefore some limitations regarding the clinical requirements. Instead, it is a general purpose modeling language and it is the responsibility of the modeler not to model irrelevant details (e.g. events that are irrelevant for the medical professionals).

On the other hand, the EPC notation is very valuable for the objective of event-based process performance measurement. Furthermore, the EPC provides the flexibility of extensive event enlargement within the model without changing the workflows. This requirement is one of the main advantages regarding the usage of EPCs for event-based process monitoring. Additionally, EPCs are mostly used at the lower levels of process hierarchy. Especially this level of detail is also necessary for performance measurement.

EPCs consist among others of functions, events, control flows and connectors. Functions are active elements and describe actions, i.e. the (sub-)processes or tasks [1], [9]. On the other side, events describe as passive elements the state of the process and are in contrast to functions not time-consuming [1]. They are starting points of activities (functions) and are themselves results of such functions. Both elements – functions and events – have to be connected in an alternating manner. Furthermore, sequences, branches, loops, joins and splits can be modelled using logical connectors like AND, OR and XOR [9], [13].

4 A Novel Generic Clinical Reference Process Model

A methodological approach is essential to provide the comprehensive and comparable definition, documentation, analysis and monitoring of clinical processes. In order to come to a common understanding about clinical workflows, events and activities as well as performance measurement, we developed a clinical reference process model with the objective of clinical pathway definition and process cycle time measurement in the field of time-critical diseases like heart attack and stroke. The model itself, as well as the clinical modeling requirements resp. challenges are presented in the following.

Clinical process modelling is a challenging task. Workflows performed in hospitals are people-intensive and require closely interaction between physicians and clinical staff to appropriately perform patient's care and treatment [14]. Furthermore, clinical processes are often affected by various departments and fields of responsibility as well as hospital information system-catchment areas [15]. The flexibility requirements for clinical process models complicate the task additionally [16], [17]. Therefore, modeling in clinical domain requires interdisciplinary consultation and cooperation between all involved stakeholders [18]. These are all persons performing or supporting the clinical process, like responsible physicians or clinical staff and IT-specialists. Furthermore the common clinical standards have to be considered for modeling in clinical domain. Clinical guidelines, as written before, are important universal quality standards defined by independent organizations for specific diseases. For measurement purposes they can be used as reference for comparison to the hospital's workflow and also for deriving clinical important performance indicators which should be observed.

Because of the specified requirements, physicians were interviewed in preparation of the modeling task. Since the model should enable time measures in the field of time-critical diseases like heart attack and stroke, cardiologists and neurologists were involved. Furthermore standardized best practice pathways were took into consideration. These are hospital independent and disease-specific best practice process descriptions developed based on clinical guidelines for heart attack and stroke. Thirdly, a task-oriented Customer Reference Process Model was consulted, which was developed at Siemens Healthcare based on the Siemens Process Reference House for interoperability analyses purposes [19]. Furthermore, to define clinical meaningful KPIs, we studied common national and international indicator libraries as well as clinical guidelines for heart attack and stroke. As clinical processes are performed in hospitals, where IT-systems (like Hospital Information System, Radiology Information System etc.) are used, the involved IT-system degree was analyzed finally in order to define IT-based events, which could automatically provide timestamps for the clinical meaningful KPIs.

Based on the results, a set of generic clinical tasks which are performed to treat acute diseases were identified. For example, imaging procedures are always necessary to attend and treat patients with heart attack or stroke. First of all, the identified set of tasks was separated into several categories and assigned to several process modules aiming a higher flexibility of the model. In the next step the modules were arranged hierarchically building a total of six process levels. Fig. 1 shows schematically the modules at level 1 to level 4. On the top level, the patient process itself is the main activity which is modeled. Using the term “patient process” we mean all activities and tasks, which are performed during a patient’s hospital stay starting with the admission of the patient and finishing with his discharge. Because the onset phase (the duration from the first pain resp. signs of heart attack or stroke until the admission at hospital) is time-critical in acute diseases, the transport activity was inserted into the model additionally to enable more exactly time measures [19].

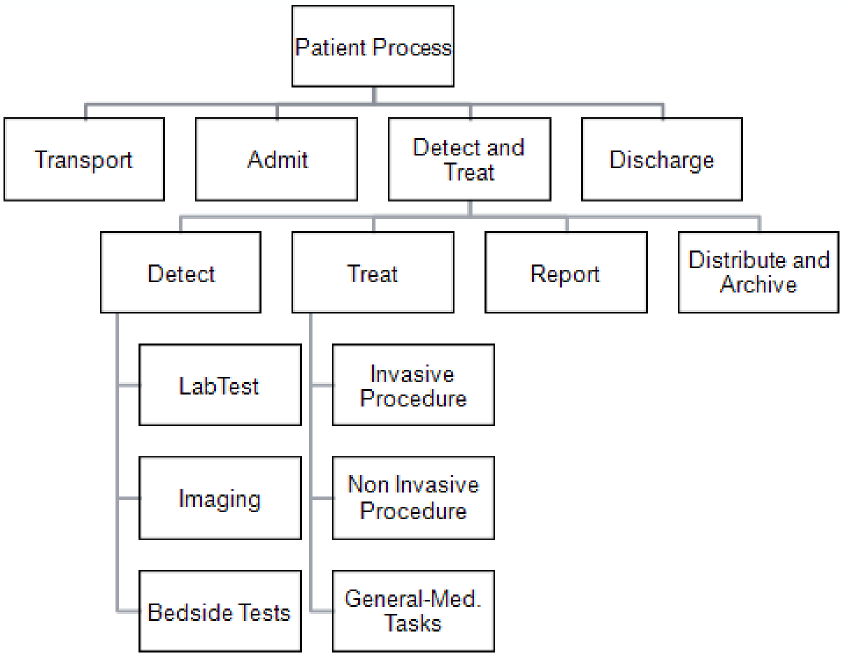


Fig. 1. The newly developed clinical reference process model: The schematically structure of the first four process levels (cf. [19])

An EPC-based counterpart was assigned to each of the modules presented in Fig. 1. Fig. 2 shows the EPC-module which implements the 2th level module “Detect and Treat”. The modules on level 3 are called using the EPC process path elements. Process path elements are commonly used following the EPC-syntax for connecting several EPCs on different hierarchy levels. In our approach we use them for providing a possibility to connect the EPC-modules. For example the 3th level module “Detect” in Fig. 1 is called by the process path element “Perform diagnostic procedure” in Fig. 2, the module “Treat” by the interface “Perform treatment procedure”, “Report” by “Perform documentation” and the last 3th level module “Distribute and Archive” by “Distribute and archive results”.

Each of the boxes in Fig. 1 represents at least one module (i.e. one EPC chain). Thereby, the presented boxes represent different aspects resp. sub-processes of the clinical patient process. It should be noted that each of the boxes is defined only once. For example, the module “Report” can be also used (i.e. called) by the modules “Detect” and “Treat” and further other modules. The module “Distribute and

Archive" can be analogical called by several other modules too. In summary, all modules can be called separately and run repeatedly without the necessity of embedding redundancies in the model [1]. Beyond that, new activities can be appended depending on the disease in question and hospital's needs.

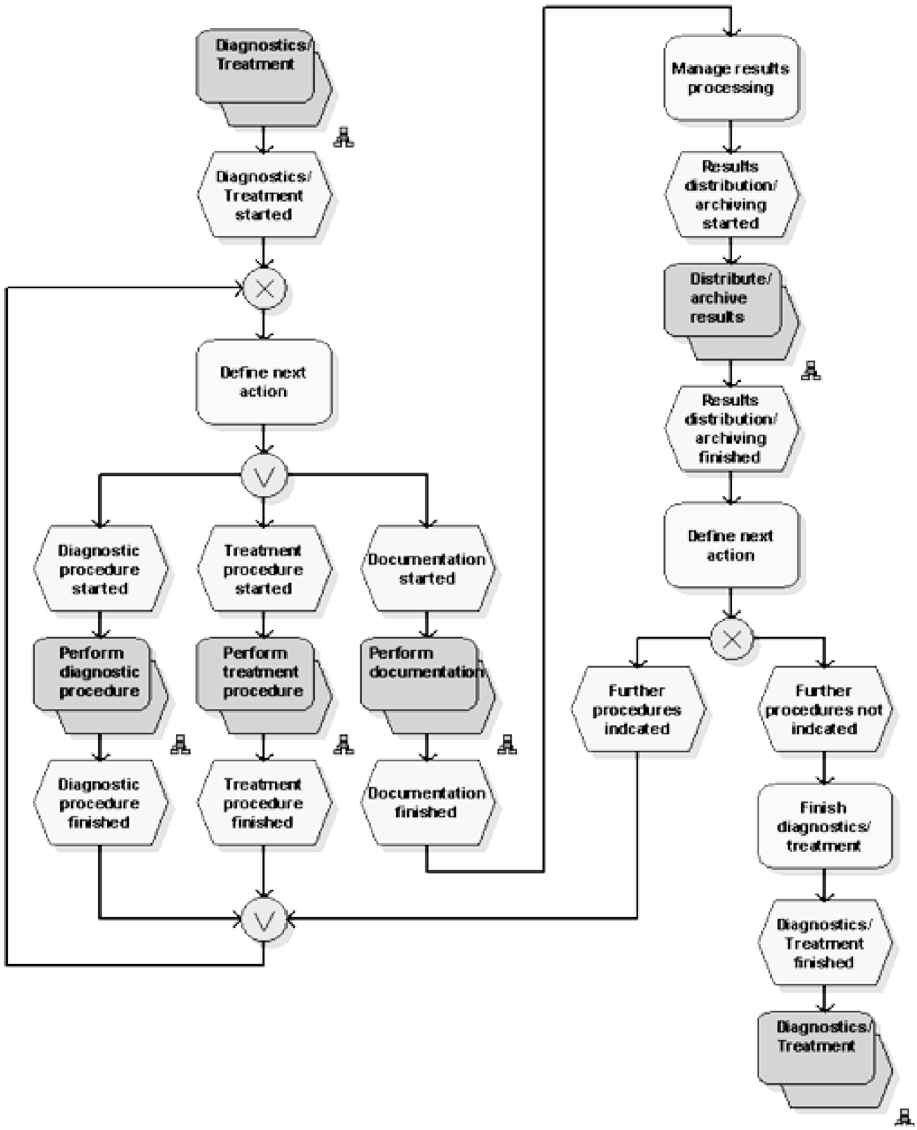


Fig. 2. The newly developed clinical reference process model: The 2[th] level module "Detect and Treat"

Considering the patient process, the transport, admission and discharge can be performed only once during one patient's hospital stay (cf. Fig. 1, 2th level). However, after patient's admission several diagnostic and treatment procedures can be performed in the reiterate manner. Because of the variance, which is necessary to be considered in clinical models, attention was paid to the flexibility of the EPC modules. As shown in Fig. 2, each of the modules on level 3 can be called separately and in the reiterate manner. For example, after finishing one diagnostic procedure e.g. electrocardiography (ECG), a further diagnostic procedure e.g. computer tomography (CT) can be performed. Afterwards the treatment procedure can be started, which could be a coronary intervention in a heart catheter laboratory. Our model is however flexible enough, to allow further diagnostics after the treatment (e.g. lysis etc.).

As already mentioned before, we consulted common national and international indicator libraries and clinical guidelines for heart attack and stroke to derive clinical meaningful, feasible, relevant, understandable, non-ambiguous and disease-oriented KPIs. First, we identified the most important clinical events, which allow to draw conclusions about the quality of care. For example, the time until when an imaging procedure like CT was started is very important in patients with heart attack or stroke. The reason is that a subsequent appropriate treatment procedure can be started as recently as the imaging procedure has started and was finished. For the measure of the time when the diagnosis was made, the starting point of the imaging procedure is an important indication.

Table 2 shows in the right column a set of disease specific KPIs using the example of heart attack. On the left side, the generalized KPIs are assigned, which are used in the clinical reference process model. The generic KPI "Door-to-Treatment" for instance complies to the KPIs "Door-to-Balloon" and "Door-to-Lysis" in patients with heart attack. The term "door" means hereby the time, when the patient arrived "at the door" at hospital. With regard to automatically retrieving of the timestamp, this is often the time, when the patient was registered in the concerned IT-system (usually it is the Hospital Information System (HIS)). The term "balloon" means the interventional treatment (i.e. balloon dilatation) of the occlusion caused the heart attack. Especially the "Door-to-Balloon"-time is very important regarding process performance in the field of heart attack treatment and decides about the quality of care outcome.

To establish process performance measure points in order to measure the generic KPIs we added new events into the model [19]. As in Fig. 2 shown, each process path element is introduced by a specific starting event. It is used to timestamp the start of the sub-process represented by the following process path element. For example, in the EPC shown in Fig. 2 the start of each treatment procedure can be captured. The finishing events were inducted with the same objective. They indicate the finishing of the sub-process represented by the previous process path element. Using the timestamps of the starting and finishing events, the duration of the whole sub-process can be calculated. For instance, the time until when the diagnostics is done can be captured.

Table 2. A selected set of generic and disease-specific KPIs using the example of heart attack

Generic KPIs	Heart Attack KPIs
Door to 1st physician Contact	Door to 1st Physician Contact
Door to Lab Order	Door to Lab Order
Door to Imaging	Door to CT
Door to Data	Door to ECG
Door to Decision	Door to Decision
Lab Test Turnaround Time	Lab Test Turnaround Time
Order to Image Interpretation Time	Order to CT Interpretation Time
Order to Image Report Time	Order to CT Report Time
Preparation Time for Intervention	Catheter Laboratory Arrival to Balloon
Transport Time for Intervention	Decision to Catheter Laboratory Arrival
Door to Treatment	Door to Balloon / Door to Lysis

KPI: Key Performance Indicators
Lab: Laboratory
ECG: Electrocardiography
CT: Computertomograhy

5 Event-Based Process Times Measurement Using Generic KPIs

As our approach focuses time-critical diseases, we have chosen time-based KPIs for process performance measurement purposes and for further investigation. Fig. 3 shows selected generic KPIs, as for the most part also presented in Table 2, using the example of a patient process in the field of diagnostics including the “Diagnostics” sub-processes such as “Clinical Examination”, “Laboratory” and “Imaging”, the “Imaging” sub-processes such as “Request/Order”, “Scheduling”, “Planning”, “Preparation”, “Performing” and “Finishing” as well as the “Performing” sub-process “Specific Performing”. These clinical diagnostic sub-processes are drawn in the middle of the figure. The “Admission” sub-process is added into the consideration because several KPIs start already during admission (e.g. “Door to Admission”, “Door to Imaging” or “Door to Lab Order”). We assign each of the in presented KPIs to the adequate process events in our developed clinical reference process model.

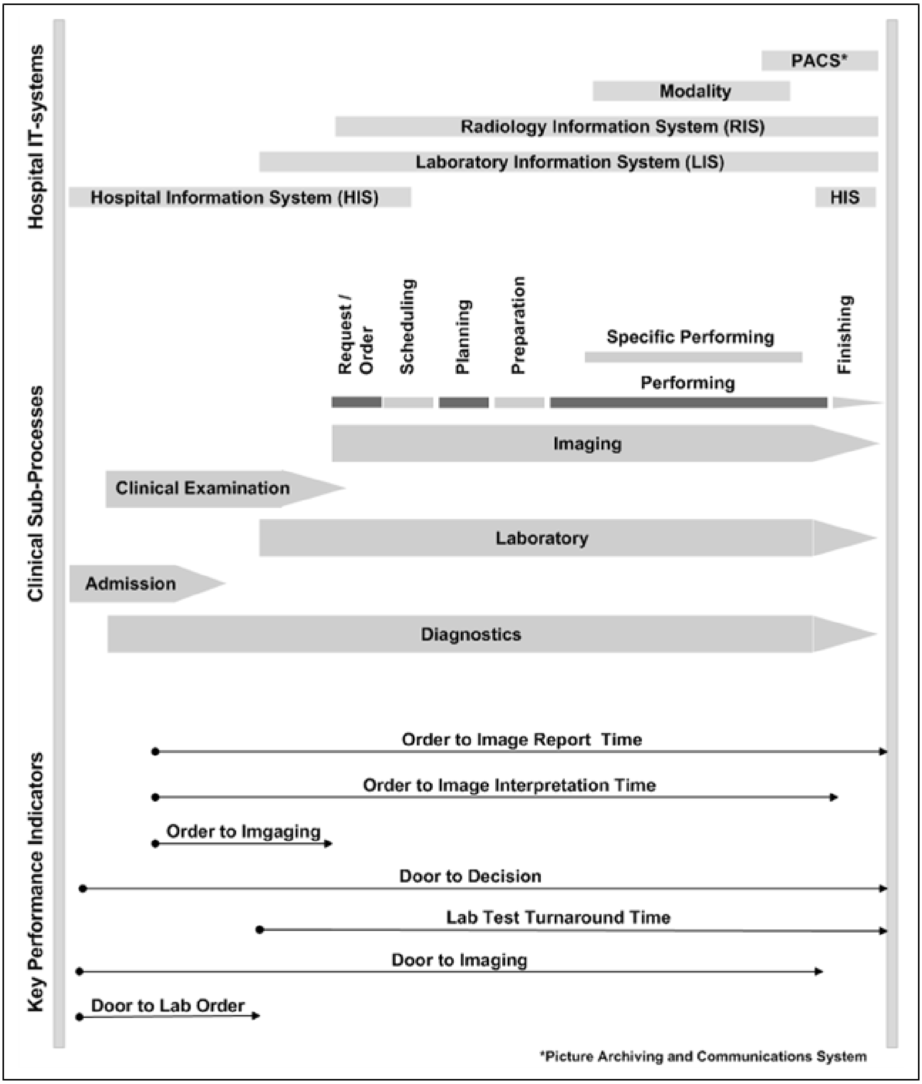


Fig. 3. The newly developed clinical reference process model: Mapping of the generic KPIs and involved clinical IT-systems to the clinical processes using the example of Diagnostics

There is a correlation between Fig. 1 and Fig. 3: The sub-process "Diagnostics" in Fig. 3 correlates with the 3th level module "Detect" in Fig. 1. The sub-processes "Laboratory", "Imaging" and "Clinical Examination" comply with the 4th level modules "LabTest", "Imaging" and "Bedside Tests" in Fig. 1. The further sub-processes ordered along the "Imaging" sub-process in Fig. 3 comply with appropriate 5th level modules, which are not visible in Fig. 3. These are sub-processes assigned to the 4th level module "Imaging". The sub-process "Performing" in Fig. 3 contains a further 6^{th} level sub-process “Specific Performing”, which is used for time measurement during an imaging procedure.

Each of the presented sub-processes is performed in a catchment area of at least one specific clinical IT-system as shown in Fig. 3. Thereby, the Hospital Information System (HIS) is responsible for all administrational tasks during a patient’s stay in hospital. In the radiology department: the Radiology Information System (RIS), imaging devices (Modalities) and the Picture Archiving and Communications System (PACS) are used to support and perform necessary diagnostic procedures. Outside the radiology the Laboratory Information System (LIS) is responsible for supporting all laboratory tests during diagnostics [20].

6 Related Work

In the field of clinical performance measurement our approach is best of our knowledge new. There are indeed several general overviews which address the clinical quality indicator topics. However, they are neither disease-oriented nor patient process-based. On the other hand, clinical studies-based guidelines are often addressed. Thereby, the clinical procedures order and the time duration until the treatment is finished are focused on as important quality of care aspects in acute diseases. However, none of the approaches investigate the situation in such a comprehensive manner as it is necessary to provide a standardized performance measurement in the field of time-critical diseases like heart attack and stroke.

Regarding the workflow performance analysis there are approaches, which provide event-based process monitoring and mining with the objective to optimize the workflow [21], [22], [23], [24], [25]. However, they were not specifically designed for health care. The 3LGM-Meta-Modell focuses indeed sub-processes in hospitals and provides the modeling of the clinical activities in the catchment area of the Hospital Information System (HIS) [26], [27]. In summary, all approaches have in common that they don’t support event-based time measurement. Furthermore, clinical and disease-specific performance measurements are not supported.

7 Conclusion and Outlook

Deriving time-measure-oriented KPIs from the presented process model based on clinical IT-systems is a promising way to support efficiency and quality improvement efforts in health care. Therefore our approach provides a generic set of feasible, relevant, meaningful, understandable, non-ambiguous and disease-oriented KPIs. Additionally, we propose the standardized usage of time-oriented KPIs in acute diseases.

We verified our generic set of KPIs by consulting several health care organizations and clinical experts regarding the topics mentioned above. In the following, implementation tasks will be started to show additionally the possibilities of standardized KPI-extraction and process mapping. The limitations of an automated workflow performance measurement occur because of the lack of interoperability in a typical hospital IT-environment. Interoperability problems are caused among others by the heterogeneity of a lot of manufacturer systems and the clinical standards as well as systems diversity.

In summary, our presented novel approach for standardized performance measurement as well as our newly developed clinical reference process model provides a comprehensive description of the complete clinical patient-centered process in hospitals. It fulfils performance measurement requirements particularly in the field of time-critical diseases like heart attack and stroke. We defined our process quality measures commonly and standardized them. Thereby, our presented comprehensive performance measurement approach enables quality controlling and sustainable care quality improvement as well as a comparative quality reporting.

References

1. Gattnar, E., Ekinci, O., Detschew, V.: Event-based Workflow Analysis in Healthcare. In: Proceedings of the 9th International Workshop on Modeling, Simulation, Verification and Validation of Enterprise Information Systems (MSVVEIS 2011), In Conjunction with the 13th International Conference on Enterprise Information Systems (ICEIS 2011), pp. 61–70 (2011)
2. Birdsell, J.M., Hayden, K.A., Lewis, S.: Performance Measurement in Healthcare: Part II. Healthcare Policy 2(1), 57–78 (2006)
3. Ghattas, J., Peleg, M., Soffer, P., Denekamp, Y.: Learning the Context of a Clinical Process. In: Proceedings of the Business Process Management Workshops (Prohealth 2009), pp. 545–556 (2009)
4. Deutsche Gesellschaft für Kardiologie, `http://www.dgk.org`
5. American Heart Association, `http://www.americanheart.org`
6. Kahla-Witzsch, H.A.: Interprofessionelles Qualitätsmanagement als unverzichtbare Grundlage für eine erfolgreiche Pfadarbeit. In: Hellmann, W. (ed.) Praxis Klinischer Pfade: Viele Wege führen zum Ziel, Ecomed, Landsberg/Lech (2003)
7. Kahla-Witzsch, H.A., Geisinger, T.: Clinical Pathways in der Krankenhauspraxis, Kohlhammer, Stuttgart (2004)
8. Keller, G., Nüttgens, G., Scheer, A.-W.: Semantische Prozeßmodellierung auf der Grundlage Ereignisgesteuerter Prozeßketten (EPK), Veröffentlichungen des Instituts für Wirtschaftsinformatik (IWi), Universität des Saarlandes, Heft 89, Saarbrücken (1992)
9. Fettke, P., Loos, P.: Perspectives on Reference Modeling. In: Fettke, P., Loos, P. (eds.) Reference Modeling for Business Systems Analysis. Idea Group Publishing Inc., London (2007)
10. Status Quo Prozessmanagement 2010/2011: BPM expo, `http://www.bpm-expo.com`
11. vom Brocke, J.: Referenzmodellierung: Gestaltung und Verteilung von Konstruktionsprozessen. Logos Verlag, Berlin (2003)
12. Sarshar, K., Dominitzki, P., Loos, P.: Einsatz von Ereignisgesteuerten Prozessketten zur Modellierung von Prozessen in der Krankenhausdomäne. In: EPK 2005: Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten, Proc., 4. Workshop der Gesellschaft für Informatik e.V (GI), Hamburg, pp. 97–116 (2005)
13. Davis, R.: Business Process Modelling with ARIS. Springer, London (2001)
14. Gattnar, E., Ekinci, O.: Qualitätsmanagement im Krankenhaus – Konzepte, Rahmenbedingungen und Standards. In: Podium der Wirtschaft, Recht & Soziales, Band, vol. 21, pp. 67–130 (2011)

15. Lenz, R., Reichert, M.: IT support for healthcare processes – premises, challenges, perspectives. Data & Knowledge Engineering 61, 39–58 (2007)
16. Anyanwu, K., Sheth, A., Cardoso, J., Miller, J., Kochut, K.: Healthcare Enterprise Process Development and Integration. Journal of Research and Practice in Information Technology 35(2), 364–376 (2003)
17. Dadam, P., Reichert, M., Rinderle-Ma, S.: Prozessmanagementsysteme – Nur ein wenig Flexibilität wird nicht reichen. Informatik Spektrum 34(4) (2011)
18. Becker, J., Fischer, R., Janiesch, C.: Optimizing U.S. Health Care Processes - A Case Study in Business Process Management. In: Proceedings of the 13th Americas Conference on Information Systems, AMCIS 2007 (2007)
19. Gattnar, E., Ekinci, O., Detschew, V.: Clinical Process Modeling and Performance Measurement in Hospitals. In: Proceedings of the AQuSerM 2011: 5th International Workshop on Advances in Quality of Service Management - In conjunction with the 15th IEEE International Enterprise Computing Conference (IEEE EDOC 2011), pp. 132–140. IEEE Computer Society Press (2011)
20. Gattnar, E., Ekinci, O., Detschew, V.: Interoperable Process Monitoring using Clinical IT-Standards and Healthcare Information Systems. In: Proceedings of the IADIS International Conferences Informatics 2011, Wireless Applications and Computing 2011, Telecommunications, Networks and Systems 2011, pp. 59–66 (2011)
21. Rubin, H., Pronovost, P., Diette, G.: The advantages and disadvantages of process-based measures of health care quality. International Journal for Quality in Health Care 13(6), 469–474 (2001)
22. zur Mühlen, M., Rosemann, M.: Workflow-based process monitoring and controlling-— technical and organizational issues. In: Sprague, R. (ed.) Proceedings of the 33rd Hawaii International Conference on System Science (HICSS-33), pp. 1–10. IEEE Computer Society Press (2000)
23. Aalst, W.M.P.: Business Alignment: Using Process Mining as a Tool for Delta Analysis. Requirements Engineering 10(3), 198–211 (2005)
24. Aalst, W.M.P., et al.: Workflow mining: A survey of issues and approaches. Data & Knowledge Engineering 47(2), 237–267 (2003)
25. van der Aalst, W.M.P., ter Hofstede, A.H.M., Weske, M.: Business Process Management: a Survey. In: van der Aalst, W.M.P., ter Hofstede, A.H.M., Weske, M. (eds.) BPM 2003. LNCS, vol. 2678, pp. 1–12. Springer, Heidelberg (2003)
26. Winter, A., Haux, R.: A Three Level graph-based Model for the Management of Computer-Supported Hospital Information Systems. Methods of Information in Medicine 34(4), 378–396 (1995)
27. Buchauer, A., et al.: 3LGM: Method and Tool to support the management of heterogeneous hospital information systems. In: Computers in Medicine, Polish society of Medical Informatics, pp. 72–82 (1997)

Quality Improvement Models for Business Process Change – A German Case Study

Matthias Kaiser[1], Frederik Kerssenfischer[1], Frederick Steinke[2], and Tobias Fritsch[3]

[1] Technische Universität München, Garching b. München, Germany
[2] Humboldt-Universität zu Berlin, Berlin, Germany
[3] Freie Universität Berlin, Berlin, Germany
Matthias.Kaiser@mytum.de, ozodis@gmail.com,
SteinkeF@student.hu-berlin.de, T.Fritsch@gmx.net

Abstract. This paper describes the role of business phase models with regards to modern IT-service design. Hence, the usability of two models incorporated for IT-service design in modern business environments will be evaluated. The two examined models are namely the classic four-step model also referred to as PDCA-Circle of Deming on the one hand and the lean Six Sigma model on the other hand.

The approach includes qualitative interviews with six different specialists of German companies in disjoint practices. This practice-diversity is needed to profile the current process of IT-service design over cross-sectional areas. Based on the standardized interviews, seven drivers of the research framework have been identified as quantification guidelines for the performance of phase models in current IT-service design. Thereafter, a case study of a large German insurance IT-department was analyzed. The results are showing differences in the explanatory power. Thereby, the lean Six Sigma model offers a wide set of tools for customer orientation, whereas the classic four-step model scores lower on average. Based on these insights, this paper analyzes if the findings can be generalized for other practices and how future IT-services can be supported by a modification of the theoretical frameworks.

1 Introduction

The IT-evolution of the last decade has shown rapid trends in terms of complexity and interface design towards business applications. Currently, nearly every company (despite of their business practice) is using IT-systems and their functionality is still growing.

This leads to new challenges for Managers, since the market pressure is reducing the available budget and the expectation for high-quality applications as well as innovation is growing ever since. The IT itself faces a change from a pure tool, which is used to automate processes, towards an enabler for new business models [WBK07]. In most business practices the flexibility and value add relies significantly on a performing, depending, flexible and adoptable IT.

The leverage of IT-standardization and partial automation is concluded in the current trend of "IT industrialization". This includes the cross-industry best-practice

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 77–90, 2011.

exchange from production of real assets and services [We09]. The arising IT-service management aims to tailor customized IT products to the necessities of the customers. This change reflects a constant migration from technical aspects and feasibility towards a support oriented IT-product perspective, which consequently includes the needs of the customers [OB91].

The key question for the analysis in this paper is: How well can modern IT-service design be predicted by using theoretical concepts. On par with this topic the question about how future IT-service concepts can be further supported by modifying the existing approaches.

2 Background

This section contains background information about the history of process optimization and both evaluated models.

2.1 History of Process Optimization

[ZBP05] state that IT-services usually only require a performance approach in sole phases. For these areas typical production related approaches are often used, such as total quality management or Six Sigma. With regard to the performance, the total quality management ensures the necessary level of quality through a systematic and ongoing improvement management. Six Sigma uses statistical methods for controlling, hence trying to minimize the variation for strategic aspects of the product [Br06].

Business processes in service industrial practices show a different level of standardization. This difference is based on the level of predictability. The current controlling literature is concluded in [Ri08]. The authors state that the immateriality and the missing customer focus are the main drivers for the limited quantification of the business influence factors. Thus, the management is often not capable of finding first-best-solutions due to the complexity and interdependence of business processes [Ku08].

A typical process classification between type A and type B processes helps to identify, how the underlying problems need to be engaged. Type A processes are mostly standardized application flows. The input and output as well as the interfaces are mostly clarified. In contrast, type B processes are neither standardized with regard to the output nor towards the process itself. They pose a high degree of freedom, which leads to uncertainty and in some situations (with inadequate specifications/exception handling) missing stability.

In this context both, [Ti07] and [ZBP05] discuss the current trend of IT-services shifting from mainly type A processes towards a growing percentage of type B processes. With this fundamental change in the role of IT-processes, a more flexible and structural background is needed. While common models may not support the current evolution of IT-services enough, this paper analyzes the usability of the four-step model and the Six Sigma approach for an IT-case, which includes type A as well as type B problem fields.

2.2 Four-Step Model

The four-step model is based on [De82], who listed 14 different points to improve the productivity and quality in business processes. These improvement potentials indicate guidelines for the management, mainly focusing on services. His focus within the optimization is the improvement of processes and operation towards a best-in-class quality.

In order to reach the benchmark quality level, Deming's model is based on four steps. It is commonly known as the Deming Circle or PDCA-Circle (Plan, Do, Check, Act). The rudimentary idea is an iterative circle that supports constant quality improvement in four given steps:

1. **Plan:** Establish the necessary objectives and processes to deliver results in accordance with the expected output. By focusing on the expected output, it differs from other techniques by integrating the completeness and accuracy of the specification into the improvement process.
2. **Do:** Implement the new processes. This is often done on a small scale and, if possible, via test-scenarios of prototyping.
3. **Check:** Implement the new processes and compare the results with the expected results to identify any differences.
4. **Act:** Prove the differences to clarify their causes (root-cause-analysis). Each will be part of either one or more of the P-D-C-A steps. Determine where to apply changes that will lead to an improvement. These four steps complete the 'circle' – the journey from facing the problem (Plan phase) to solving the problem (Act phase).

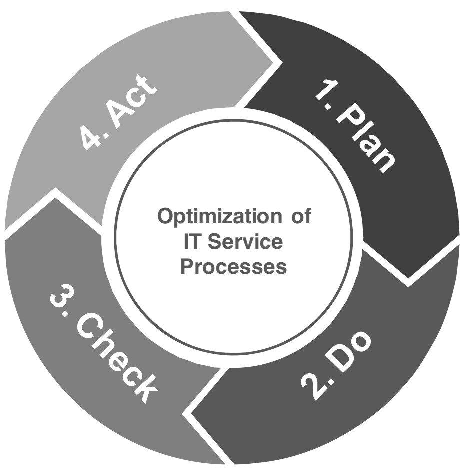


Fig. 1. PDCA-circle (Four-step model) from [DM92]

The PDCA circle or four-step model is commonly used from a management perspective. For IT-processes [Ol08] gives an overview about an ITIL v3 continual service improvement with this method. The four steps and their results are explained. For this paper the PDCA circle will act as one of the two chosen models for IT-service improvements.

2.3 Six Sigma

The Six Sigma model is a business management tool to evaluate strategies to improve quality and reduce variances of a process. The name Six Sigma is the result of the initial goal to achieve an error rate of less than 0.00034%, which is six times the standard deviation in the normal distribution. Complementary to the PDCA circle of the four-step model, the six sigma model uses a similar approach of different steps to follow in order to improve a process. The main difference lies in the addition of a fifth step. Even though it follows the Deming Circle, it offers a wider variety of input while improving an existing business process [Re06].

The acronym of these five steps is DMAIC:

1. Define the problem, including all stakeholders of the process chain. Set a project goal.
2. Measure the data of the current process and try to collect important key figures.
3. Analyze the collected data and try to find interactions between the parameters, which can be improved. In order to reduce or raise a certain parameter, make sure to consider the influencing factors.
4. Improve the process by adjusting important key factors that were identified before. Use test- and pilot-runs to ensure an improvement of the new process.
5. Control the new process and compare the results with the expected results. Finally, deviations from the new calculated figures can be corrected by reacting and fine tuning the new work process.

Six sigma aims to improve the quality of manufacturing and business processes by identifying and eliminating the source of errors. Variability is analyzed with statistical methods and has to be avoided. The method was developed originally by Motorola in 1981 and is used today in many sectors of industry with adoptions for example for service industries [KA06]; [PNC07]; [St07]; [Wi07].

Since it is focused and because it is offering a framework, the hypothesis for the PDCA circle and Six Sigma performance in modern IT-service can be described as:

The PDCA circle and Six Sigma are usable for type A processes but fail to deliver the necessary business support in terms of tools to address more dynamic problems (type B processes).

3 Research Framework

After describing the history of process optimization as well as giving an introduction of the two theoretical business management tools (four-step model and Six sigma model), the following part focuses on the paper's framework parameters and approach.

Due to the increasing complexity and growing importance of IT-processes, a structural procedure at enterprise level is essential. Subsequent to the literature meta-analysis, seven success factors are derived. These factors should be used for structuring the improvement of IT-processes. Moreover, the factors serve as a starting point for the quantification of the expert interviews. With the help of these interviews, a transfer of the theory to the practice is possible. A detailed explanation will follow in the next chapter.

3.1 Literature Meta-analysis

For a first general overview about the complexity of the topic a meta-analysis of the literature was conducted. The authors used three different channels for their research in order to reach the widest possible coverage. The search terms used in the literature research were used in German and the according terms in English to assure good coverage of the topic.

1. **Textbook-Research:** The electronic media search of the library of the Technical University of Munich has relevant sources (in terms of papers and journals). The following key words were used: "IT-Services, Operational Management, Business Process Management and Quality Management".
2. **Research in scientific databases:** To determine current scientific articles "Google Scholar" and "Ebsco" were used and likewise searched for the above keywords.
3. **Systematic evaluation of scientific journals** (amongst others "International Journal of Service Industry Management" and "Zeitschrift für Wirtschaftsinformatik"): The vintages 2006, 2007, 2008 and 2009 of different magazines were analyzed using abstract titles and texts.

Following the meta-analysis more than 150 selected documents with the above-named keywords were available.

3.2 Development of Seven Success Factors

These documents were evaluated and recurring facts collected. Clusters for different topics were built to identify the most important success factors for the optimization of IT-processes. As the result of the literature analysis the following table presents seven success factors and the most important associated sources (Table 1).

Table 1. Success factors in the optimization of IT processes (authors design)

Success factor	Description and sources
(1) Customer Integration	The typical interaction in the service sector between supplier and customer in service delivery has to be optimized [Fl09]; [Kr10]; [ZBP05].
(2) IT- Services Characteristics	The optimization should consider the characteristics of IT service providers, namely hardware (using standards), software (customization) and service (customer involvement) of the hybrid products, which are available and specifically tailored to customer needs [Bo04]; [Za07]; [ZBP05].
(3) Process View and Industrialization	The optimization should examine standardized processes, which follow the expiry of a predefined schema and return a certain result [Ri08]; [Ti07]; [Wi07].
(4) Operational Control	The optimization shall support the implementation of processes in daily execution and improvements through appropriate measures [Bl04]; [Fl09]; [MW08].

Table 1. (Continued)

(5) Effectiveness and Efficiency	The optimization should consider both internal efficiency targets for economic performance and efficiency targets for the satisfaction of customer needs [Br96]; [GHK08].
(6) Quality Management	The optimization should include the observation, measurement and control of the customer perceived service quality [BH01]; [Br06]; [DM92]; [Py03] [Sh39].
(7) Roles in Process	The optimization should distinguish between different roles, including IT-managers, employees, suppliers, customers and users in the process and respond to their interests and information needs [Kr10]; [ZBP05].

3.3 Expert Interviews

The next part of the approach includes standardized interviews with six different specialists from German companies in disjoint practices to highlight the performance of the business phase models regarding the seven success factors in the corporate context. In general, the qualitative research aims to receive a deeper understanding by open questions and smaller, more intense information gathering. Moreover the interview implies a quantitative component in which the experts have to rate the seven success factors relating to the two theoretical management process tools. The 5-point Likert scale serves as quantification guideline for the suitability of phase models in the current IT process optimization [Li32]. Based on the approach to qualitative research [Ei89] the procedure for the interviews in this paper is arranged. Some phases are reduced or combined and a quantitative part is supplemented.

In the following the selection of interviewees was made by using defined selection criteria. Respondents with a deep understanding project management and control and optimization of IT-processes were selected. This ensures that the interviews cover the desired topic. Six employees of the internal IT-staff of independent IT-services, as well as employees of other chemical and insurance companies were selected. Due to the sensitivity of the information the names of the companies will not be published. The interviews were conducted in a semi-structured way. This implies that the interview process was only roughly outlined in advance in order to allow a responding to the interview situation. The developed standardized questionnaire guaranteed a consistent structure for the six interviews.

The interviews in the present work were conducted face-to-face or by telephone. The interviewees had no personal interest in the results and could describe their objective point of view. Moreover, reliability and validity of the statements were ensured. All interviews were recorded to facilitate subsequent evaluation. For the analysis of the individual interviews, transcripts of the conversations were prepared. For evaluating the suitability of the developed success factors with reference to the four-step model and the Six sigma model the participants were asked for their estimation on a 5-point Likert scale.

After describing the analytical proceeding of the separate interviews, the search for comprehensive patterns will be brought into focus in the next part. The mean of the responses of the six interviewees marked on the presented 5-point Likert scale is shown in the chart below (Figure 2). The left hand side (--) represents a limited suitability of the particular model in terms of the success factor whereas on the right hand side (++) there is a major suitability. For example, for the factor (5) Effectiveness and Efficiency the four phase model shows a very low and the Six sigma model a high suitability.

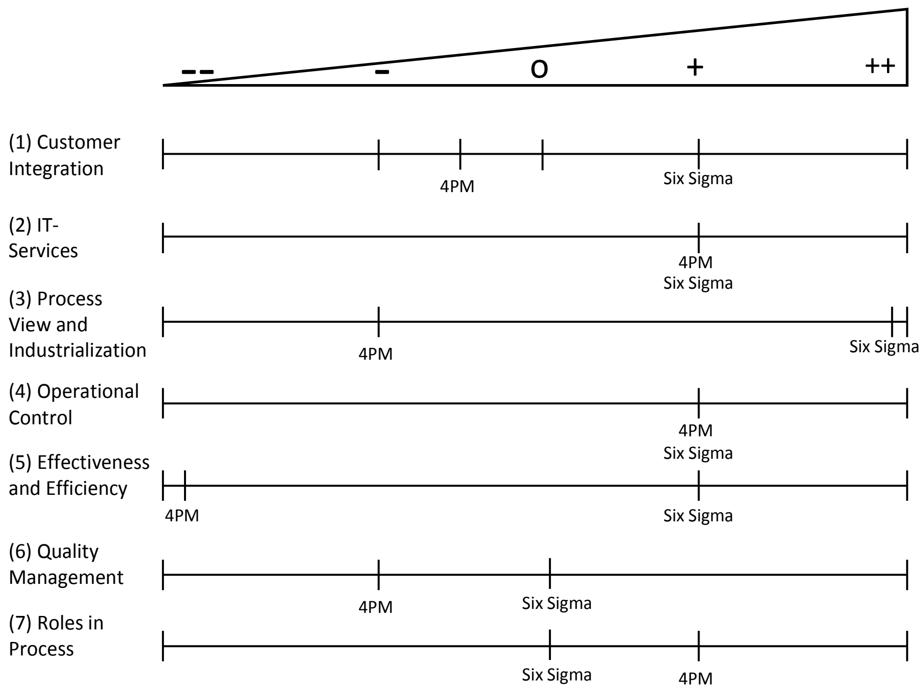


Fig. 2. Expert opinion on the suitability of the two theoretical models (authors design)

The responses given by the six experts have shown noticeable differences between the suitability of the two analyzed models in reference to the developed factors. The graphic illustrates the cumulative assessment of the six interviewees. With regard to the seven success factors, the advantages and disadvantages of the four-step model and the Six sigma model in practice can be clarified. It can be seen that four success factors display advantages for the Six sigma model whereas only one shows advantages for the four phase model (Figure 2).

The additional scientific value of the literature meta-analysis, the subsequent creation of the seven success factors for the IT-process optimization, and the application on the two theoretical models has to be reviewed with help of the case study. In order to underline the individual interviews' findings, it is useful to analyze a current practical IT case study and monitor the results related to the previously made results below.

4 Case Introduction

In the digital age the manual processing of paper files is not productive. Therefore, the transformation of traditional paper files into metafiles and integration in a uniform database is an absolute requirement. During the transitional period it is necessary to scan old paper files and provide them electronically.

In this context, the optimization of the digitalization process of different types of paper files plays an essential role. Using the case of paper file digitalization in a large German insurance company, this section describes the use of business phase models for process optimization. The original digitalization process consisted of relatively complex steps with a significant proportion of manual work. For an optimization of this process, the daily manual work had to be minimized. Moreover, an improvement of this component has influences on the processing time, error rate, and cost factors and brings quality intensification.

This case was chosen as a representative one for the application of the Six sigma model. It is both IT-service specific (i.e. the technical scanning process) and relevant for the daily business of the insurance company which is the customer of the process. Additionally, the chosen framework is a standard tool for process optimization of service processes in the focused group of companies. The activities of one process optimization project are shown as follows to give an insight on the application of one of the models.

4.1 Structure and Results

First, the setting of the case is described. Both new files and all kinds of incoming mail for existing files are stored digitally. However, there still are archives of paper files for which storage and access cause notable costs. A paper file digitalization process is used for scanning files systematically in order to empty archives.

Both the use and the results of a Six sigma approach with five phases for analyzing and optimizing that process are described now. For confidentiality reasons, relative improvements are presented rather than absolute figures.

1. Define

In workshops with the all stakeholders of the process (clerks, IT-specialists, management) the present process chain was modeled and goals for processing time, error rate, and average costs per file were set. A set of 17 optimization points was discussed and the best ones were selected.

The first process step is the selection of the files that should be scanned. These files are shipped to a service provider, who prepares the files for scanning and digitalizes the files using high performance scanning devices. Metadata is collected to connect the images with the corresponding digital files.

Goal of the optimization is to ensure that the process design fits into the customer requirements regarding the days until the digital file is available. A certain error rate is accepted because achieving a lower error rate would be more expensive than dealing with a rather small amount of failures.

2. Measure

To have a qualitative reference to which results can be compared to, important parameters of the present process are collected over a period of two months. Therefore, automatically generated reports are used as well as collection by employees that are responsible for process execution. Result of the measurement is a gap between the actual as-is-process and the goal values of the define-step. The long term sigma value of the present process is 2.4.

3. Analyze

To realize a process improvement, the usability of the actions defined in the plan-step has to be evaluated. Therefore, all gathered data is analyzed and a business cases is calculated concerning the costs and use of implementation regarding the defined goals. A result of this analysis was that the most promising action is a change from using an external scan service to scanning the files with free in-house resources using already available hardware and processes. To illustrate this, the further steps of the proceeding will be outlined in what follows.

4. Improve

To realize gains calculated in business cases the best actions were conducted. Therefore, the as-is-process was adapted as planned by informing the right people and changing technical implementations. To use the in-house resources instead of the external provider, necessary process adjustments were taken and a pilot run was used to ensure correct results.

5. Control

Finally, the new process was controlled to measure the results of the optimization, to decide whether another run through the five steps is necessary, and to make sure that these achievements are used in the future as well. The outcomes of the new process are a reduction of average processing time by 32% while reducing cost per file by 40%. The sigma value of the process could be improved to 4.2. Compared to other service processes the result is satisfying and so another run through the steps was not necessary.

4.2 Assessment and Influence Factors

After the introduction of the case and the presentation of the results, the performance of the used Six Sigma model will be discussed in the following, using the 7 success factors presented in chapter 3.2. Therefore, project members were asked to answer the questionnaire presented in 3.3. The average results are presented in figure 3.

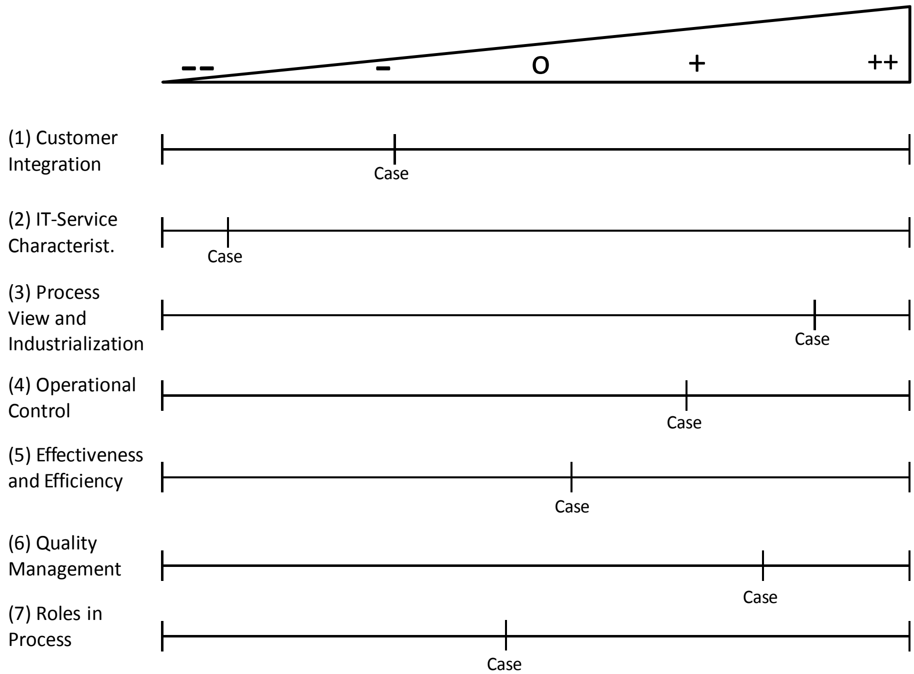


Fig. 3. Team members' opinion about the performance of the Six Sigma model (authors design)

To interpret these quantitative assessment qualitative interviews with four project members were conducted and the results for the seven success factors merged:

(1) Customer Integration tools of the Six Sigma Model were applicable to some extent for the digitalization steps of the process.

(2) IT-Service Characteristics are only partially supported by the Six Sigma model but were incorporated anyway in the project, i.e. the optimal fit of hardware, software, and staff for service execution. Skills of the project team had compensated that gap.

(3) Tools of the Six Sigma model for Process View and Industrialization could successfully be applied.

(4) Operational Control: Six Sigma instruments for measurement and improving daily work could gainfully be applied for the paper file scanning process.

(5) As proposed by the Six Sigma model, economically and functionally-oriented stakeholders were distinguished between to consider effectiveness and efficiency goals.

(6) Quality Management from a customer perspective was conducted with Six Sigma model tools for measurement and optimization of error rates, i.e. minimization of missed pages in the scanning process.

(7) Six Sigma tools could be applied to match process steps with different roles in the process, even though there were no predefined hints for each role. It was also part of the project to frame the different interactions during process execution.

The project members summarized that the Six Sigma model was very useful for the evolution of the paper file digitalization process. The framework for the project composition could successfully be applied. Solutions for very IT specific parts of the project could be generated using the skills of the team members.

The following chapter broadens the perspective of a single project and process optimization framework and compares the outcomes with the results of the interviews of chapter 3.3.

5 Analysis

In the following section, the results of the expert interviews are compared to the case results regarding the usability of the Six Sigma model for the optimization of IT-processes. As an indication we use the seven success factors that were rated on the 5-point Likert scale. The results show that there is a slight discrepancy between the expert opinions and the opinion of the team involved in the optimization process.

Referring to the first success factor Customer Integration, the gap between the experts and the team valuation of this factor is clear-cut. While the experts valued this parameter clearly suitable for the Six Sigma model the team members did not find this represented in their case.

Ongoing with the next factor, a severe gap for the suitability regarding the IT-Services characteristics was found. The opinions differed, ranging from the rather good usability evaluation of the Six Sigma tools by the experts to the lowest possible rating from the team members involved in the optimization of processing paper files. As mentioned before, we believe that this gap shows how the optimization process for IT-Services is not dealt with in an optimal way by using the Six Sigma model for it. This factor could be sufficiently considered during the process evaluation just by using the project staff's IT knowledge.

The remaining factors of the Six Sigma model, namely Process View and Industrialization, Operational Control, Effectiveness and Efficiency, Quality Management and Roles in Process where rated equally useful by the experts and the team members.

Therefore, the focus of the analysis should lie in the difference discovered for the first two factors concerning the Six Sigma performance. As mentioned in section 2.1 of this paper, two types (A and B) of processes could be roughly described for the field of IT-services. Whereas Type A is mostly standardized and input and output can sufficiently be described, Type B processes offer a higher degree of uncertainty about important parameters and mostly display more complex situations. Due to the lack of far-reaching IT-knowledge on the customer side, the requirements actually needed are most often only vaguely defined. This leads to a low rating for Customer Integration when using the Six Sigma model tools because the model fails to help creating the necessary information needed and defining adequate target figures for the IT-optimization. While on a theoretical level, Six Sigma appears to be a valid method also for IT-processes, in practice it shows it roots for optimizing manufacturing and business processes. The complexity of IT-solutions is hardly captured by the attempt to optimize one or two target figures that can be named by the typical IT-customer.

In accordance with the result of Customer Integration we find the most significant gap for the IT-Service characteristics where the team members stated the worst performance of the Six Sigma model, while the experts listed a good theoretical performance of the tools. As for integration, the characteristics in IT-processes do not fit in the usual manufacturing optimization scheme. Influence factors like hardware, software, and support service with their many different characteristics are at best only poorly addressed when using the Six Sigma model with a sparse number of indicators. Especially this discrepancy leads to our understanding of the actual lag in IT-process optimization.

6 Conclusion

The evolution of IT-processes rapidly progressed in the past years. The complexity rises with the linkage of customer-tailored software, hardware, and service solutions. As shown in this paper, different methods exist for process optimization in the IT-sector which are used by IT managers to reduce this complexity and optimally use the potential to create a performing, depending and flexible IT-solution.

Starting with an overall analysis of important factors for the optimization of IT-processes, this paper shows in an exemplary case the lag of optimization tools concerning IT-processes. We find that the seven factors which were named most important for the success of an optimization by experts where only partly matched with the current methods used in large companies for process optimization. By identifying two meanderings from the theoretical framework and the actual application of the methods in a real process optimization we discover the disadvantage of methods like Six Sigma for IT-processes. As stated before, both analyzed optimization models (Four-step model and Six sigma) show their weaknesses in customer integration and the characteristics of IT-services. Only with a larger involvement of the customer and a broadened approach regarding the target figures which are optimized, the complexity of modern IT-processes can be handled. Due to the interferences caused by the design of the mostly hybridly constructed modern IT-products, the next step has to be the development of a theoretical model that allows a wider variety of input.

The identification of future standard methods for complex IT-processes is to be developed and evaluated in further research. This new model framework is important because the limitations of the classical optimization tools become evident for IT-processes. As long as all seven success factors identified by this paper can be successfully integrated in the conceptual framework, we strongly believe it will improve IT-process optimization.

References

[BH01] Benkenstein, M., Holtz, M.: Qualitätsmanagement von Dienstleistungen. In: Bruhn, M., Meffert, H. (eds.) Handbuch Dienstleistungsmanagement, Wiesbaden, pp. 193–209 (2001)

[Bl04] Bleicher, K.: Das Konzept integriertes Management: Visionen – Missionen – Programme, Frankfurt (2004)

[Bo04] Böhmann, T.: Modularisierung von IT-Dienstleistungen, Wiesbaden (2004)
[Br96] Broglie, M.: Steigerung der Performance von Informatikprozessen, Braunschweig (1996)
[Br06] Bruhn, M.: Qualitätsmanagement für Dienstleistungen, Heidelberg (2006)
[Bu07] Burianek, F., Ihl, C., Bonnemeier, S., Reichwald, R.: Typologisierung hybrider Produkte – Ein Ansatz basierend auf der Komplexität der Leistungserbringung. In: Arbeitsbericht Nr. 01/2007 des Lehrstuhls für Betriebswirtschaftslehre – Information, Organisation und Management der TU München (2007)
[DM92] DeLone, W., McLean, E.: Information Systems Success: The Quest for the Dependent Variable. Information Systems Research 3(1), 60–95 (1992)
[De82] Deming, W.E.: Out of the Crisis, Cambridge (1982)
[Ei89] Eisenhardt, K.M.: Building Theories from Case Study Research. Academy of Management Review 14(4), 532–550 (1989)
[Fl09] Fließ, S.: Dienstleistungsmanagement – Kundenintegration gestalten und steuern, Wiesbaden (2009)
[GHK08] Goeken, M., Heck, R., Klein, H.: Mit dem Service Value Lifecycle zur wertorientierten Serviceorganisation. HMD - Praxis der Wirtschaftsinformatik 264, 19–28 (2008)
[Kr10] Krcmar, H.: Informationsmanagement, Berlin (2010)
[Ku08] Kumar, V., Smart, P.A., Maddern, H., Maull, R.S.: Alternative Perspectives on Service Quality and Customer Satisfaction: the Role of BPM. International Journal of Service Industry Management 19(2), 176–187 (2008)
[KA06] Kwak, Y.H., Anbari, F.T.: Benefits, obstacles, and future of six sigma approach. Technovation 26(5-6), 708–715 (2006)
[Li32] Likert, R.: A technique for the measurement of attitudes. Archives of Psychology 149, 5–55 (1932)
[MW08] Macharzina, K., Wolf, J.: Unternehmensführung: Das internationale Managementwissen – Konzepte – Methoden – Praxis, Wiesbaden (2008)
[Ol08] Olbrich, A.: ITIL kompakt und verständlich, Wiesbaden (2008)
[OB91] Orlikowski, W.J., Baroudi, J.J.: Studying Information Technology in Organizations: Research Approaches and Assumptions. Information Systems Research 2(1), 1–28 (1991)
[PNC07] Pande, P.S., Neuman, R.P., Cavanagh, R.R.: The Six Sigma Way. In: Boersch, C., Elschen, R. (eds.) Das Summa Summarum des Management: Die 25 wichtigsten Werke für Strategie, Führung und Veränderung, Wiesbaden, pp. 299–308 (2007)
[Py03] Pyzdek, T.: The Six Sigma Handbook: A Complete Guide for Green Belts, Black Belts and Managers at all Levels, New York (2003)
[Re06] Reckenfelderbäumer, M.: Konzeptionelle Grundlagen des Dienstleistungscontrolling – Kritische Bestandsaufnahme und Perspektiven der Weiterentwicklung zu einem Controlling der Kundenintegration. In: Bruhn, M., Stauss, B. (eds.) Dienstleistungscontrolling, Wiesbaden, pp. 31–51 (2006)
[Ri08] Riedl, C., Böhmann, T., Rosemann, M., Krcmar, H.: Quality Management in Service Ecosystems. Information Systems and E-Business Management 7(2), 199–221 (2008)
[Sh39] Shewhart, W.A.: Statistical Method from the Viewpoint of Quality Control, Mineola (1939)
[St03] Stamatis, D.H.: Six Sigma and Beyond: Statistics and probability, Boca Raton, Florida (2002)

[St07] Stiehler, A.: Interdig – Systematisierung und Handelbarkeit von IT Dienstleistungen (2007), http://www.berlecon.de/studien/downloads/Interdig_Report1.pdf (last checked March 22, 2010)

[Ti07] Tiemeyer, E.: Organisation und Führung im IT-Bereich. In: Handbuch IT-Management, München, pp. 347–384 (2007)

[WBK07] Walter, M., Böhmann, T., Krcmar, H.: Industrialisierung der IT – Grundlagen, Merkmale und Ausprägungen eines Trends. HMD - Praxis der Wirtschaftsinformatik 256, 6–16 (2007)

[We09] Wei, C.J.: Theories and Principles of Designing Lean Service Process. In: Ji, G. (ed.) 2009 6th International Conference on Service Systems and Service Management, pp. 821–826 (2009)

[Wi07] Wilkinson, J.: IT Service Management - An Introduction, Based on ISO 20000 and ITIL V3. 1 (2007)

[Za07] Zarnekow, R.: Produktionsmanagement von IT-Dienstleistungen, Heidelberg (2007)

[ZBP05] Zarnekow, R., Brenner, W., Pilgram, U.: Integriertes Informationsmanagement: Strategien und Lösungen für das Management von IT-Dienstleistungen, Berlin (2005)

Entropy-Based Indicator for Predicting Stock Price Trend Reversal

Virgilijus Sakalauskas and Dalia Kriksciuniene

Department of Informatics, Vilnius University,
Muitines 8, 44280 Kaunas, Lithuania
{virgilijus.sakalauskas,dalia.kriksciuniene}@khf.vu.lt

Abstract. Predicting changes of stock price long term trend is an important problem for validating strategies of investment to the financial instruments. In this article we applied the approach of analysis of information efficiency and long term correlation memory in order to distinguish short term changes in trend, which can be evaluated as informational 'nervousness', from the reversal point of long term trend of the financial time series. By integrating two econometrical measures of information efficiency – Shannon's entropy (SH) and local Hurst exponent (HE) – we designed aggregated entropy-based (EB) indicator and explored its ability to forecast the turning point of trend of the financial time series and to calibrate the stock market trading strategy.

Keywords: Shannon entropy, informational efficiency, financial market, local Hurst exponent, stock price.

1 Introduction

In this article we assume the possibility that the characteristics of long term correlation memory can be employed for analysis of the financial time series. Changes in this characteristic can enable us to predict the turning point of stock price long term trend. Long term correlation memory means the interdependences existing among observation periods of the time series even if these periods are separated. This assumption contradicts to efficient market hypothesis or at least states that there exist periods of temporary inefficiency of the financial markets. In case of discovery of long term dependencies of the stock prices, the investors could have better opportunities for decisions related to portfolio structure and investment timing.

The level of stock market inefficiency and long term correlation memory can be measured by applying several statistical and econometrical indicators. The most widespread measures among the researchers are Shannon's entropy and Hurst exponent. These indicators are designed in such way that we can measure the level of efficiency of the market information even without knowledge of the underlying causal factors. The research works which employ these indicators mostly analyze the stock market index, as it entails the information accumulated by all listed stocks of the market.

Most popular measure of information efficiency is Hurst exponent (*HE*). It allows not only to evaluate long term correlation memory, but also to classify the time series.

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 91–98, 2011.

The range of *HE* falls to the interval of [0; 1]. It equals to 0.5, if the analyzed time series is completely random. If *HE>0.5* and its value approaches to 1, it indicates, that the analyzed time series is persistent and has particular trend which most probably will not change its direction in nearest future. If *HE<0.5* the trend component of time series is vague and unstable, its upward direction is changed by down value or vice versa, which is denoted as antipersistent series [1].

Some authors hesitate about the reliability of the Hurst exponent for identification of long term correlation [2,3], even for the *HE* value, different from *0.5*. Nevertheless its ability to make comparisons of the efficiencies of different markets was noticed by many authors. In Cajueiro and Tabak [4,5] we can find comparative evaluation of the efficiency of Asian countries stock markets with those of Latin America, J.Barunik and L.Kristoufek [6] have estimated *HE* for S&P500 index, S.Danilenko [7], has calculated *HE* for the Baltic sector indices. The Hurst exponent also can be used for predicting financial time series [6] or even for crash prediction in finance [8].

The other indicator applied for evaluation of market efficiency is Shannon's Entropy *(SE)* [9,10]. *SE* is the measure, which can be applied for evaluating level of uncertainty of symbolized time series. Differently from *HE*, the Shannon's entropy cannot classify time series to persistent and antipersistent. As the *SE* measure is less affected by noise, it is widely applied to solving problems related to Information theory and Symbolic time series analysis. The noisiness of data is very effectively lowered by symbolizing time series, therefore application of SE opens new analysis possibilities for high-resolution financial data.

Shannon's entropy measure takes values in the interval of [0,1]. *SE=1* means, that the analyzed time series is completely random, and its changes are unpredictable. The bigger the difference of *SE* values from 1, the better forecasting possibilities are created.

If the time series is formed of the financial data, such as market index values, the *SE* values approaching to 1 can be interpreted as high market efficiency. The analysis of changes of Shannon's entropy gives more opportunities to find regularities if the values are more different from 1.

In this work we explore, how to extract and to combine the most effective qualities of Shannon's entropy and Hurst exponent measures for forecasting financial time series. The suggested idea is based on creating aggregated entropy-based (*EB*) indicator which could recognize forthcoming turning point of long term trend from temporary market fluctuations, and to investigate the behavior and ability of the indicator to forecast changes of the long term trend of the financial time series, and to adequately calibrate trading strategy for the particular stock market.

In the following chapter we analyze the calculation methods of Shannon's entropy and Hurst exponent. Also we introduce the entropy-based indicator, the suggested methodology for its calculation and application for time series analysis.

In the third chapter the experimental research is presented, the *EB* indicator is evaluated and analyzed for predicting stock prices long term trend turning points. The research results and advantages of the suggested methods are summarized in the conclusions section.

2 The Methodology for Evaluation of Entropy-Based Indicator

In this chapter we present the methodology of computing Hurst exponent and Shannon's entropy measures for the series reflecting the fluctuations of returns of the stock market prices or market index, expressed by formula 1:

$$F_t = \ln(P_t) - \ln(P_{t-1}) \tag{1}$$

where $\{P_t\}$ is the stock price or market index value, F_t –Return, t- particular time interval t.

The application of the Shannon's entropy method starts with symbolizing the time series, aimed to divide the series into the parts, giving best prediction opportunities. This procedure reduces the noisiness of time series. Symbolizing means that the elements of the time series are coded by applying the selected symbolic notation (in most cases 2 or 3 symbol systems) for the series within the time frame of sufficient length (the size of the time frame in the scientific research literature varies between 150 and 400 days). Two symbol coding system means denoting positive returns as 1, and the negative returns as 0. We apply the 2-symbol method, as it was used in [11].

Let the resulting symbolic time series is $\{s_1, s_2, ..., s_N\}$, where N is the size of calculation window and s_i takes the values 0 or 1. Within the time frame of length *N,* we create the sliding window of length *L*, and segment the array $\{s_i\}_1^N$ to the words consisting of *L* symbols. As we use two symbols coding system, the number of different words within the calculation window is equal to 2^L.

Size of word *L* is optional and is selected in such way, that relative frequencies of all different words of length *L* could be calculated:

$$r_i = \frac{\textit{frequency of selected word } i}{2^L}$$

Shannon's entropy, otherwise the measure of uncertainty of time series information, is defined as in formula 2:

$$S = -\frac{1}{\log_2 2^L} \sum_{i=1}^{2^L} r_i \log_2 r_i \tag{2}$$

In case all the relative frequencies r_i were equal, it can be stated that the symbolic time series $\{s_i\}_1^N$ is completely random, therefore its Shannon's entropy measure $S=1$ (formula 2). If the symbolized array is designed of the stock prices or market index series, the financial market can be considered as efficient. If the frequencies are different, the *S* value is less than 1 and falls to the interval [0,1]. If the differences among the relative frequencies of the words are more evident, the market is less efficient, and the chance to forecast time series is increased.

In order to explore evolution of Shannon's entropy measure in time, the calculation window of length *N* is slide further and the computation procedure is repeated for each successive element of the time frame. The obtained series of *SE* values represent changes of time series efficiency.

Hurst exponent is the statistical measure of time series long-range dependence. Several methods can be applied for its evaluation: Rescaled Range Analysis (R/S), which was proposed by H.E.Hurst [12], Detrended Fluctuation Analysis (DFA) developed by Peng et all [13], Detrended Moving Average [14] and other methods.

In this article we apply Rescaled Range Analysis (R/S) technique for evaluating the Hurst exponent.

$\{F_t\}_{t=1}^{N}$ time series of returns are calculated by formula (1) from the stock price or market index. Hurst exponent is calculated by applying the 4-step procedure:

1. Firstly the cumulative mean adjusted series Y are computed: $Y_t = \sum_{i=1}^{t}(F_i - \overline{F_N})$,

where $\overline{F_N}$ is the mean value of series $\{F_t\}_{t=1}^{N}$, e.g $\overline{F_N} = \frac{1}{N}\sum_{i=1}^{N} F_i$.

2. Then the differences R are calculated:

$$R_t = \max(Y_1, Y_2, ... Y_t) - \min(Y_1, Y_2, ... Y_t),\ t=1,2,...N$$

3. Standard deviation of the values for all t are estimated:

$$S_t = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(F_i - \overline{F_t})^2}$$

4. The ratios (R/S) for each $t=1,2,...N$ are calculated: $(R/S)_t = \frac{R_t}{S_t}$

It was shown by Hurst (1951) that (R/S) ratios follow the relationship:

$$(R/S)_t = c \cdot t^H, \tag{3}$$

where c is the constant and H denotes the Hurst exponent value.

The Hurst exponent value is calculated by taking the logarithm of both sides of the expression (3) and defining the slope H of equation by applying the least squares method:

$$\log(R/S)_t = \log(c) + H\log(t).$$

The H value belongs to interval [0,1]. If $H=0.5$, it can be stated that the respective time series is completely random and behaves as the white noise process. In this case any long term forecasting is irrelevant.

If value of H increases comparing to 0.5, the time series obtain better expressed trend, the probability of trend reversal lowers, therefore the time series is persistent. The opposite situation occurs when the Hurst exponent value is less than 0.5 and approaches to 0. In this case the time series is antipersistent, the probability of change in trend is more likely. As in the case of calculating Shannon's entropy, the H value is measured for the entire time frame, therefore we can compute H values as time function for each moment t by sliding the calculation window of size N.

It can be noted that Shannon's entropy measure is not able to separate status of persistent series from the antipersistent. Therefore this indicator can be used only as long range dependence measure. The aggregated indicator could strengthen *HE* ability to define likeliness of changes in trend by adding *SE* valuation of efficiency.

The entropy-based (*EB*) indicator is designed by aggregating *SE* and *HE*, and is defined by the following expression (4):

$$EB = \frac{1}{2} - (\frac{1}{2} - H) \cdot (1 - S), \tag{4}$$

where H - Hurst exponent value, S - Shannon's entropy value.

In case of randomness of the time series, where H equals to 0.5 or $S=1$, then the indicator *EB* is also equal to 0.5 and characterizes the time series as white noise process. In all the other cases *EB* indicator can take any value from [0,1] and similarly to Hurst exponent characterizes persistent series by *EB* values greater than 0.5, and antipersistent series by values lower than 0.5.

The main difference of aggregated indicator *EB* from *HE* is created by adding influence of *SE* values. If S value is very near to 1, but H value is significantly different form 0.5, the compound value of *EB* will be much nearer to 0.5 than H value. And vice versa, if H value is near 0.5, but S is significantly different from 1, the *EB* value is shifted from 0.5.

The indicator *EB* as defined by expression (4), aggregating H and S, can be considered as time series long range dependence measure and applied for defining the revolving moment of time series trend direction. We expect that this measure can be especially useful for predicting financial time series, characterized by high noisiness.

By exploring the suggested measure it was noticed, that fluctuations of EB indicator and frequent crossing points of the 0.5 level do not provide opportunity for result interpretation, therefore the results cannot be used for long term forecasting purposes. In order to reduce have high noisiness of indicator values and to research likeliness of changes of long term trend analysis we suggest to apply smoothing by n - points moving average technique. The smoothing value n can be selected according to the analysed financial time series. It can be assumed that the smoothing value n has to be sufficiently big - from two weeks to two month and it has to be selected and adjusted to each analysed market due to their differences in efficiency.

3 Experimental Investigation of *EB* Indicator

The experimental research was designed in order to explore the performance of aggregated indicator *EB* for forecasting reversal points of long term trend direction of financial data series, and to explore its performance by applying smoothing. The research database was formed of daily index values of NASDAQ OMX [15] Vilnius stock exchange (OMXV) from 2004.03.18 to 2009.07.30.

The length of calculation window $N = 215$ line was selected by taking into account the main requirements: N should be sufficiently large that it could reflect long range changes of financial time series, and at the same time is has to encapsulate not too long period (e.g. not longer than 300 days), that it could react to local changes of the time series [11,16]. The measures of efficiency *SE* and *HE* were computed for each window of size 215, which was further sliced along the database of 1355 cases. Consequently we obtained 1140 values for tracing the evolution of market efficiency by measuring each efficiency measure *SE* and *HE*. For calculating Shannon's entropy we symbolized the log return values of OMXV index by using two-symbol system: 0 for negative returns and 1 for positive and zero returns. The length of sliding word $L=4$ was selected and the calculations were performed for all 1140 *SE* values.

The aggregated entropy-based (*EB*) indicator is further calculated by using formula (4). As the indicator is composed from both *HE* and *SE* values, it should be less affected by the noisiness of initial time series, which can increase precision of predicting the event of long range trend reversal point. In Fig. 1 the charts of *EB* indicator smoothed by applying span = 20 and the OMXV values are presented.

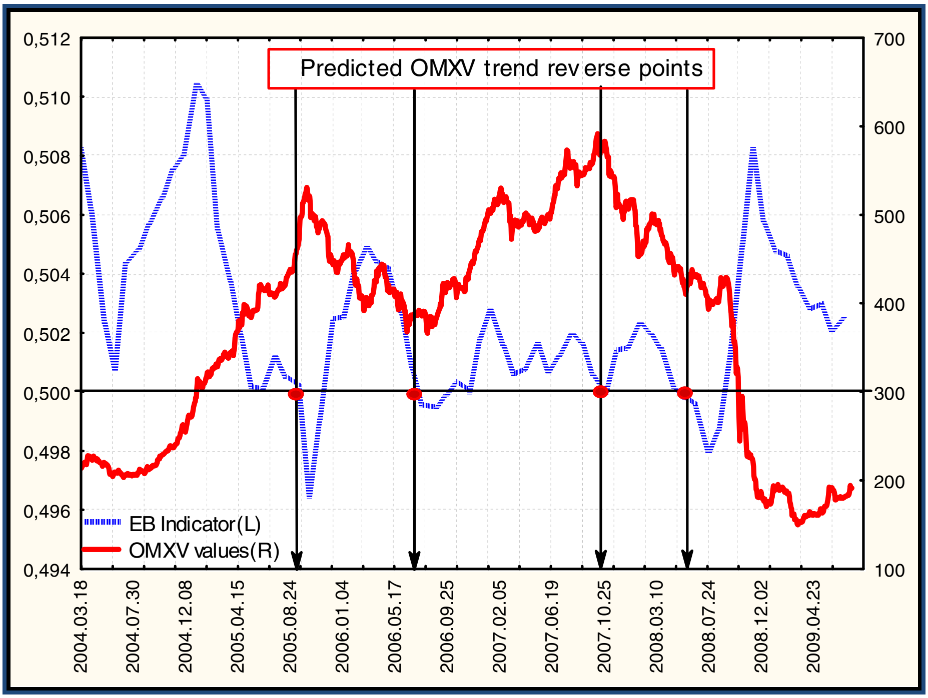


Fig. 1. Aggregated *EB* indicator and NASDAQ OMX Vilnius Index values

This figure allows to not only compare the behaviour of the indicator and index values, but also to analyse some of their important correlations. Each time when the declining *EB* value crosses the level of 0.5, the trend of time series reaches its reversal point quite soon. Four such points are marked in Fig.1, when the important reversal points of OMXV trend took place: periods of raise of September of 2005 and August of 2006, and the crises of November 2007 and fall of 2008.

The time interval between the decrease point of *EB* to the level of 0.5 and the anticipated reversal of OMXV trend is different in each case: from two months time interval before the crisis of the fall 2008 to one week interval before the decline of OMXV index in November 2007.

4 Conclusions

The presented research explores the possibility to predict stock price or market index long term trend changes. In this research two measures of information efficiency and long term correlation memory were applied: Hurst exponent (*HE*) and Shannon entropy (*SE*).

It was investigated that forecasting by applying the measures *HE* and *SE* is complicated due to their high noisiness. The values of Shannon's entropy are near 1 and they can be mostly applied for evaluation of long term correlation of time series. Hurst exponent values can be applied for classification of the time series according to the existence of trend: either persistent (supporting current trend) or antipersistent (not having any constant trend).

In this work we explore the finding, that the moment when the time series changes its character from persistent to antipersistent indicates approaching moment of reversal point of the time series trend.

The aggregated entropy-based (*EB*) indicator was designed and the experimental research was conducted in order to test the finding. The NASDAQ OMX Vilnius (OMXV) daily index values from 2004.03.18 to 2009.07.30 were applied for performing calculations and exploring forecasting possibilities for long range trend changes.

The indicator *EB* is defined as function of two arguments *HE* and *SE*. This function enhances long-term correlation effect, if at least one of indicators *HE* and *SE* show a high correlation. Also the *EB* indicator can easily distinguish persistent series from antipersistent.

To predict the time series using the *EB* indicator, we still need to smooth them in order to remove random noise. The width of smoothing depends on the nature of financial series. Smoothing width can be identified from several weeks to several months depending on the issue of market efficiency.

In our case, we use the aggregation span equal to 20 and can conclude that the *EB* indicator provide the very precise prediction of OMXV long term trend turning point.

The research results can be useful for the long term investors for forecasting necessary changes in strategies. The application of the aggregated measure *EB* has to be further investigated in order to discover the influence of the selected parameters, including selection of the smoothing span, which could possibly vary from several weeks to the few months.

References

1. Qian, B., Rasheed, K.: Hurst exponent and financial market predictability. In: IASTED Conference on Financial Engineering and Applications, pp. 203–209 (2004)
2. Bassler, K., Gunaratne, G., McCauley, J.: Markov processes, Hurst exponents, and nonlinear diffusion equations: With application to finance. Physica A 369(2), 343–353 (2006)
3. McCauley, J., Gunaratne, G., Bassler, K.: Hurst Exponents, Markov Processes, and Fractional Brownian Motion. Physica A 379(1), 1–9 (2007)
4. Cajueiro, D., Tabak, B.: Ranking Efficiency for Emerging Markets. Chaos, Solutions and Fractals 22, 349–352 (2004)

5. Cajueiro, D., Tabak, B.: Ranking Efficiency for Emerging Markets II. Chaos, Solutions and Fractals 23, 671–675 (2005)
6. Barunik, J., Kristoufek, L.: On Hurst exponent estimation under heavy-tailed distributions. Physica A 389(18), 3844–3855 (2010)
7. Danilenko, S.: Hurst Analysis of Baltic Sector Indices. In: Applied Stochastic Models and Data Analysis (ASMDA 2009), Vilnius, pp. 329–333 (2009)
8. Grech, D., Mazur, Z.: Can one make any crash prediction in finance using the local Hurst exponent idea? Physica A: Statistical Mechanics and its Applications 336, 133–145 (2004)
9. Shannon, C.: A Mathematical Theory of Communication. Bell System Technical Journal 27, 79–423, 623–656 (1948)
10. Cover, T., Thomas, J.: Elements of Information Theory. Wiley ed. (1991)
11. Sakalauskas, V., Kriksciuniene, D.: Evolution in information efficiency in emerging markets. In: Advances in Intelligent and Soft Computing: Soft Computing Models in Industrial and Environmental Applications: 6th IC SOCO 2011, pp. 367–377. Springer, Heidelberg (2011)
12. Hurst, H.E.: Long-term storage of reservoirs: an experimental study. Transactions of the American society of civil engineers 116, 770–799 (1951)
13. Peng, C., Buldyrev, S., Havlin, S., Simons, M., Stanley, H., Goldberger, A.: Mosaic organization of dna nucleotides. Physical Review E 49(2) (1994)
14. Guide to Baltic Market (2010), `http://www.nasdaqomxbaltic.com/files/baltic/NASDAQ_OMX_Baltic_10.pdf`
15. Risso, W.A.: The Informational Efficiency and the Financial Crashes. Research in International Business and Finance 22, 396–408 (2008)

Temporary Belief Sets Management in Adaptive Training Systems

Robert Andrei Buchmann, Anamaria Szekely, and Delia Pulcher

Babes Bolyai University,
Faculty of Economic Sciences and Business Administration,
Business Information Systems Dpt., str. Teodor Mihali 58-60,
400591 Cluj Napoca, Romania
robert.buchmann@econ.ubbcluj.ro,
{b_ana_sz,delialoman}@yahoo.com

Abstract. The paper proposes a semantic view on the notion of „learning object" and an application model based on RDF-based learning objects and learning processes. Direct feedback is individualized for test subjects and learning tasks, according to requirements defined for corporate training. The knowledge model allows contextualization and subjectivity, which, in turn, are used to dynamically generate temporary belief sets, compare them to the (theoretically) objective belief set underlying the learning content and adapt learning recommendations to each particular user. The semantic models also determine learning prerequisites and the screen flow adapted to each individual learner, thus influencing usability.

Keywords: Learning object, qualified knowledge, context, RDF, SPARQL.

1 Introduction

The goal of the paper is to propose a knowledge subjectivity management model for adaptive computer-based training systems, based on a notion of "temporary belief sets", which designate the intermediate states of a trainee's knowledge, mapped on learning content modeled with the Resource Description Framework [1]. The study is guided by individualization requirements for corporate learning adaptivity in business settings, as defined by [2].

Learning objects are abstract notions with imprecise definitions, boundaries and granularity. In the context of our approach, we consider that one of their most important features is reusability. Some authors also point to self-containment [3], but this attribute is slightly forced, as the knowledge expressed through a learning object collection, although it might be decomposable and modular, must reflect some dependencies with other concepts and learning objects. Furthermore, in natural language most term definitions are based on other terms, which in turn are based on other terms and so on, until one can reach circular references, or primitive concepts that are defined through themselves.

RDF is a knowledge representation framework underlying the paradigm of Semantic Web [4]. In RDF, concepts are built in the above mentioned manner: the

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 99–110, 2011.

notions of **class** or **property** are defined primitively and axiomatic: a class is a member of the class of all classes (*rdfs:Class rdf:type rdfs:Class*), a type is a type of property (*rdf:type rdf:type rdf:Property*). Most belief systems are built on such axiomatic, self-referential conventions, more or less formally and rigorously developed over the fundamental set of axioms. Learning content is built on the same principles, even in what are considered to be „exact” disciplines such as mathematics (which is itself based on relatively primitive notions such as the object, the relation, the number, the set etc.).

These are some reasons why, throughout our studies, we prefer to ignore the attribute of self-containment and insist on the attribute of reusability for learning objects. Also, employing RDF in knowledge representation forces us to emphasize connectivity in the detriment of self-containment.

The proposed model also has some subtle implications on the usability of computer-based training, due to its adaptivity and intelligent response to learning needs, based on the user input while he’s taking evaluation tests. This impacts the learning efficiency and user’s feeling that the system is “supportive” to his/her needs.

The next section states the problem and some of its background (including related works). Section 3 lists the low cost instruments needed for implementation. Section 4 describes implementation dilemmas and solutions regarding the subjective knowledge, ending with a general overview of the application in whose context this issue was raised, followed by a SWOT evaluation and final conclusions.

2 Problem Statement and Background

Most implementations of training systems and platforms (Moodle [5], Blackboard [6] etc.) are basically formatted computer-mediated communication systems aiming at eliminating costs and inconveniences regarding the management of a learning environment, usually aspects regarding time and space. They are less involved in automating the training processes themselves, as the intervention of the human tutor is still considered essential.

Although our proposed model is still based on human intervention, it provides a superior level of semantic automation by storing a formal representation of the knowledge states through which the trainee goes during his training sessions. This implies that the main issue to be modeled here is knowledge subjectivity and temporality which are the main attributes of a “temporary belief set”.

The paper discusses several alternatives in expressing subjective knowledge, falseness and evaluated knowledge in a computing environment based on the Resource Description Framework and its inference vocabularies (OWL [7], custom production rules [8]). The instrumentation is provided by the coupled knowledge management system provided by OpenRDF Sesame and OWLIM, managed through a REST interface and extending a Python web application acting as a traditional e-learning application. The proposed model requires a knowledge engineer to define the mapping between the learning content and the learning objects, which are defined and delimited as RDF resources.

As for related works, first of all we mention a paper discussing the multiple interpretations assigned to the notion of learning object [9], an ambiguity that allows us to take the position mentioned above. Also, significant efforts have been made for the modeling of learning objects: The Learning Technology Standards Committee defined the Learning Object Metadata [10], while more ontology-oriented approaches can be found in [11]. A solution instrumentally closer to our approach (in the sense that uses Sesame as a storage platform) is [12] – it works with SeRQL as a query language and does not employ OWLIM for inferences. Learning object formalizations using OWL and semantic technologies have been also proposed by [13] (for the domain knowledge of first-aid), which in turn is inspired by the recomposition methodology proposed by [14].

Our work emphasizes subjectivity of knowledge rather than learning object structures and taxonomies, insisting on the idea that semantic repositories backing up adaptive learning systems must be flexible enough to support unstable knowledge.

3 Instrumentation

The tools required for implementation can be obtained freely:

- OpenRDF Sesame –RDF management system developed by Aduna Software, with weak inferential capabilities (RDF Schema and the Direct Type vocabulary) [15];
- OWLIM – rule-management system developed by Ontotext AD [16] as a an extension to Sesame, thus bringing inference capabilities and production rules to Sesame;
- Python is the language of choice for prototyping, which connects to the Sesame+OWLIM platform through its REST interface and Python's HTTP capabilities provided by the urllib package. SPARQL [17] queries can be run over the REST interface, with results returned as a JSON mapping of the the standard SPARQL Result Format [18].

4 The Proposed Model

An essential requirement for learning processes is the subjectivity of expressed knowledge and viewpoints – we have to express at least two sides – the trainer's and the trainee's, and on each of these sides there might be a large diversity of viewpoints. By stating this, we don't promote the philosophical debate that knowledge is volatile and non-absolute, but rather the fact that the trainee's knowledge set shifts through **multiple subjective states**, from the initial, realistic and „unlearnt" state, to a final, ideal and theoretical state, after the trainee had acquired all the knowledge provided by the training programme. Shifts between these states can be triggered during a test-driven learning recommendation system:

An evaluation of the initial state might reveal that the „unlearnt" responder chooses answers mostly randomly. This random set represents the first temporary belief set. When confronted with the right and wrong answers, some of the trainee's knowledge shifts towards the desired one, creating another temporary belief set. This confrontation shouldn't take the form of simply displaying the right answers, but rather the one of pushing forward (through reading recommendations) the learning objects containing the right answers. A new evaluation will reveal new problems, with new requirements for adjusting the belief set, and so on.

An essential problem with this is the representation of subjective knowledge – this includes all the temporary belief sets of all trainees, and a reference set that must be acquired through learning. This reference may also vary based on the tutor's interpretation, authorities, sources, without being necessarily wrong, but rather perfectible, unstable and temporary. Thus, the representation of training content must be aligned to the principle of AAA (anyone could say anything about anything), which is well served by the Semantic Web paradigm [19].

We have to emphasize that under the umbrella term of **subjectivity**, we fit both subjective viewpoints and a lack of viewpoints (for a beginner trainee). That's why the initial state of the trainee's belief set, although generated by a general lack of knowledge, is still represented as a subjective view rather than a lack of view.

Representing knowledge subjectivity opens capabilities for multiple analysis scenarios. The most important ones for our studies are filtering of right and wrong knowledge at any given time (according to a trainer's evaluation), delta analysis (differences between the temporary knowledge sets) and a general training strategy driven by evaluation (and concept dependencies) rather than content.

In the next section we discuss several alternatives in representing subjectivity:

4.1 Representation of Contextual Knowledge

It is well known that in RDF the basic knowledge unit is the triple, which establishes a binary, directional relationship between a subject and an object. The subject-predicate-object structure, defined by the RDF standard, is fit to express most reality descriptions but it is based on the assumption of objective reality – every triple occuring in knowledge base is considered true and valid (until inconsistency detection).

Reification is a knowledge representation pattern that can be defined as "qualified knowledge" or, in other words, "contextual knowledge". This means that reification allows for the representation of subjective knowledge, by assigning a context to a knowledge unit.

Although weakly supported by RDF frameworks, there is a subset of the N3 syntax that allows for expressing reification (which should translate in a 4-arity relationship, also defined by the standard RDF vocabulary but avoided by many implementations) [20].

Several scenarios where contextual assertions would prove useful are:

- when the context is spatio-temporal and limits the validity of the statement:

```
{:John :hasAgeOf 20} :inYear 2000 .
{:Humans :SkinColor :Yellow} :InRegion :China.
```

- when the context is the subjective source of the statement:

```
:Mary :ThinksThat  {:John :hasAgeOf 20}.
:John :SaidThat
               {:Mary :ThinksThat
                       {:John :hasAgeOf 20}}.
```

- when the context is an evaluation (adverbial or quantified):

```
{:John :Plays :Footbal}      :Evaluation       :Well.
{:England :Weather :Rainy} :withProbabilityOf "70%".
```

In RDF, relationships with arity greater than 2 can be expressed by using an intermediary anonymous graph node [21]. Reification is a 4-arity relationship between the 3 components of a statement and the reifying context:

A sensitive problem in this respect is the truth value of statements. We consider it to be an evaluation, so it can be expressed as a reification:

```
{:Sun    :Orbits  :Earth}  :Evaluates  :False.
```

In our proposed model, falseness is essential in the temporary belief sets. Knowledge that evaluates to false will trigger more tutoring (recommendations) on the learning objects involved in the false knowledge. Here, the learning objects are the reified triple components (:Sun, :Earth, :Orbits, with this prioritization). Each learning object has several learning sources (web pages) assigned to it with annotation properties:

```
:Sun   rdfs:isDefinedBy <http://mycompany/lessons/Sun>.
:Sun   rdfs:seeAlso <http://someencyclopedia.com/Sun>.
```

Due to the **open world principle** which is fundamental to the Semantic Web paradigm, every RDF assertion is considered true, while the triples that are absent are NOT considered false, but rather missing (undefined, yet). Our proposed model requires a way of explicitly assigning falseness. Other contexts are also needed to effectively model subjectivity, such as the holder of the temporary belief set. With respect to a single triple, this would require at least a 5-arity relationship.

Explicit knowledge falseness can be accomplished through several artifices, of which the last two are relevant to our proposal:

1. The basic reification pattern, by defining a 4-arity relationship between the assertion elements and the truth value (of false, in this case):

```
_:MyStatement :EvaluatesTo :False.
_:MyStatement rdf:subject :Linda;
              rdf:predicate :PlayedIn;
              rdf:object :Avatar.
```

2. Relying on inference rules provided by restrictive RDF vocabularies of high expressivity, such as OWL:

```
:AvatarActors rdf:type owl:Restriction;
              owl:onProperty :PlayedIn;
              owl:hasValue :Avatar.
:NonAvatarActors owl:complementOf :AvatarActors.
:LindaHamilton rdf:type :NonAvatarActors.
```

(if the restriction occurs in the same knowledge base as :LindaHamilton :PlayedIn :Avatar, it would trigger an inconsistency, which is the basic mechanism of expressing negation in OWL).

3. The solution that we opted for defines negated variants for the predicates:

```
:LindaHamilton :PlayedInNot :Avatar .
```

This solution is problematic since it is:

- *convention-based* (we convene that all properties ending in "Not" are negations of the same properties without the "Not" particle). Of course, from a natural language perspective, it would be more intuitive to place the negation in front (NotPlayedIn, or DidntPlayIn) but keeping in mind that prefixed names are mapped to full URIs (<http://myorganzation.com/myconcepts#NotPlayedIn>) placing the negation particle at the end makes it easier to identify and extract using regular expressions; another variant would be to place a delimited token expressing the truth value:

```
:LindaHamilton :PlayedIn-True  :Terminator.
:LindaHamilton :PlayedIn-False :Avatar.
```

- based on a set of rules expressing *the mutual exclusiveness* of the true and false versions of the predicate. This is only supported by OWL 2.0 through its proposal of disjoint properties. An OWL 1 pattern for describing an aproximation of this exclusiveness is:

```
:AvatarActors rdf:type owl:Restriction;
              owl:onProperty :PlayedIn-True;
              owl:hasValue :Avatar.
:NonAvatarActors rdf:type owl:Restriction;
                 owl:onProperty :PlayedIn-False;
                 owl:hasValue :Avatar.
:AvatarActors owl:disjointWith :NonAvatarActors.
:LindaHamilton rdf:type :NonAvatarActors.
```

(this triple set states that LindaHamilton belongs to a class that does not have common elements with the class of entities who have the PlayedIn-True relationship with Avatar).

An alternative to this pattern of mutually exclusiveness may be provided by production rule systems such as OWLIM, which allows for the customization of rule sets. Custom rules are expressed in a textual format such as:

```
<:PlayedIn-True> <:MutuallyExclusive> <:PlayedIn-False>
```

(this would be a custom axiom)

```
Relationship1 <:MutuallyExclusive>      Relationship2

X   Relationship1     Y

X   Relationship2     Y

Relationship1 <:TriggeredInconsistency> Relationship2
```

(this rule defines a conveniently named inconsistency when a predicate and its negation hold between the same entities).

4. A second viable solution, much more efficient, but limited with regard to inferences, is implemented in the very foundation of the Sesame platform. It proposes the extension of the basic triple structure with a fourth concept – the context.

```
:False :LindaHamilton :PlayedIn :Avatar.
:True :LindaHamilton :PlayedIn :Terminator.
```

The limitations of these are:

- It breaks the RDF model, by employing the unit of quadruple instead of the triple, so it won't be interoperable with strict RDF systems. On the other hand, it is compliant with the RDF query language – SPARQL – which supports the so-called „named graphs". In this respect, :False and :True become graph identifiers: all true assertions can be merged in the context named :True, all false assertions can be merged in the context named :False;
- The context is not supported by the inference engines (of both Sesame and its OWLIM extension). More precisely, one could assign context to every asserted triple at upload time, but all inferred triple lose their context. This means that it's not possible to filter inferred knowledge in the same intuitive and efficient way as to filter asserted knowledge.

The context usefulness is not limited to truth values. It can be used as a replacement for any kind of reification:

```
:Mary :LindaHamilton :PlayedIn :Avatar .
```

This states that the statement of :LindaHamilton :PlayedIn :Avatar belongs somehow to the context Mary (this can be interpreted freely, in this case, as a subjective opinion of Mary).

Furthermore, as the context itself is an RDF resource, it can be subjected to its own description, under a supercontext:

```
:False rdfs:comment "this is the context of all false
beliefs".
:Mary rdfs:comment "this is the set of Mary's beliefs".
```

As reifications can be nested, supercontexts can be applied to contexts:

```
:Ann :Mary rdfs:comment "this is the set of Mary's
beliefs".
```

This states that the comment on :Mary is itself a subjective opinion of Ann (Ann thinks that Mary thinks that....). Consequently, we can combine owners of subjective knowledge with truth values or other types of contexts (spatio-temporal, adverbial, probabilistic etc.):

```
:TempBeliefSet1 :LindaHamilton :PlayedIn :Avatar.
:TempBeliefSet1 :Sun :Orbits :Earth.
:TempBeliefSet2 :LindaHamilton :PlayedIn :Terminator.
:TempBeliefSet2 :Earth :Orbits :Sun.
```

These represent two qualified temporary sets of beliefs. They can be further described through metabeliefs which might be the trainer's evaluations:

```
:Trainer1 :TempBeliefSet1 :Source :Trainee1;
                          :Evaluates :False.
:Trainer2 :TempBeliefSet2 :Source :Trainee2;
                          :Evaluates :True.
```

(the two belief sets could come from the same source, thus separating the correct knowledge from the wrong knowledge of a given trainee).

Obviously, there's no limit to the number of levels on which we can develop the knowledge qualification, in a much more efficient manner than using previously described techniques. And, we can use the named graph feature when querying the knowledge base, using the contexts as graph names:

```
SELECT ?X ?Y ?Z
WHERE
{
GRAPH ?G {?X ?Y ?Z}
GRAPH :Trainer1 {?G :Evaluates :False}
}
```

This query will extract all the knowledge that was evaluated as false according to Trainer1. The evaluations could be further filtered based on the trainee, subject or other criteria, limited only by the SPARQL capabilities. By using contexts as graph names, the query engine provides a native way of filtering knowledge based on subjectivity and correctness.

The temporary belief sets can also be qualified by temporality (timestamps), for any given trainee. This expresses how the belief set shifts from one state to another, thus recording the progress towards the ideal state. The state shifting is recorded during an on-line evaluation, while the trainee takes an on-line multiple choice test. Each question is backed by learning objects consisting in RDF assertions that express the knowledge behind the question (usually the subject and predicate maps on the question, while the object maps on the correct answer but this is not mandatory, the tutor or a knowledge engineer working with the tutor may assign more complex RDF triple sets, depending on the granularity desired for the learning objects with respect to the question meaning). The RDF assertions are qualified by its source (the tutor who defined the question) and the truth value assigned by the source. Thus, every question is represented in the system's knowledge repository.

When the trainee selects an answer, the corresponding assertions are also qualified by his identity. After such an evaluation, the system produces the trainee's temporary belief set, split in false and correct subgraphs according to the tutor's qualifications. From the false subset, the learning objects are linked to their defining pages/lessons, which are recommended for more in-depth reading.

As previously mentioned, we need to employ both the techniques 3 and 4 as long as OWLIM doesn't adopt the underlying Sesame's contextualization model at production rule level. Inferences (or even rules) limited to a given context would be of great help in delimiting subjective knowledge in an intuitive way (and, more important, in detecting inconsistencies of subjective knowledge).

4.2 Knowledge Dependencies

Usually the concepts of a learning content are linked through various degrees of dependencies: in order to understand concept x, you have to understand concept y, and so on. Learning object dependencies are easy to express in RDF, using a taxonomy of prerequisites for each learning object. This allows the system to evaluate the trainee gradually – by displaying questions as they are enabled by the fact that the trainee acquires the prerequisite concepts. From the trainee's point of view, this is perceived as a usability feature, as it affects the "feel" of a learning process – reading/learning is guided by wrong answers in the temporary belief set and by concept dependencies.

An initial dependency taxonomy (further refined by the knowledge engineer) is generated using a CONSTRUCT query which creates a (Subject-DependsOn-Object) triple for every learning object (if the Object is a resource):

```
:Earth rdf:type :Planet .
:Planet rdfs:subClassOf :CelestialBody .

CONSTRUCT {?x :DependsOn ?y}
WHERE {GRAPH :Trainer1
            {?x ?predicate ?y. FILTER isURI(?y)}
```

OWL restriction classes are used for grouping categories of enabled/enabling questions and answers and their concepts. A sample of the triplestore is expressed as follows (prefixes avoided for brevity; partially inspired by an example from [19, chapter 10]):

```
:PossibleAnswer a owl:ObjectProperty;
    rdfs:domain :Question;
    rdfs:range :Answer.
:AnswerText a owl:DatatypeProperty;
   rdfs:domain :Answer;
   rdfs:range xsd:string.
:QuestionText a owl:DatatypeProperty;
    rdfs:domain :Question;
    rdfs:range xsd:string.
:Enables a owl:ObjectProperty;
    rdfs:domain :Answer;
    rdfs:range :Question.
:SelectedAnswer a owl:ObjectProperty;
    rdfs:subClassOf _:x.
_:x a owl:Restriction;
    owl:onProperty :Enables;
    owl:allValuesFrom :EnabledQuestion.
:AnsweredQuestion a owl:Restriction;
    owl:onProperty :SelectedAnswer;
    owl:someValuesFrom :Answer.
```

```
:DependsOn a owl:ObjectProperty;
   rdfs:domain :LearningConcept;
   rdfs:range :LearningConcept.
:EvaluatedBy a owl:ObjectProperty;
   rdfs:domain :LearningConcept;
   rdfs:range :Question.
:AcquiredConcept a owl:Restriction;
   owl:onProperty :DependsOn;
   owl:allValuesFrom :CorrectAnsweredConcept.
```

The store is managed and queried in order to extract contextual semantics based on the current learning objects approached by the student. The semantics control the question flow (by making sure the content is not delivered unless the student acquires all prerequisite concepts).

4.3 General Architecture

All queries are executed over Sesame's REST interface, via HTTP requests triggered by Python's urllib module. Data is returned as JSON, easily parsable in Python dictionaries for further processing. Due to the limitations of the open world notion of inequality, aggregate computations (such as counting the wrong/correct triples) are not supported by the (current) implementation of SPARQL [22], so they are executed with Python, over the query result dictionaries.

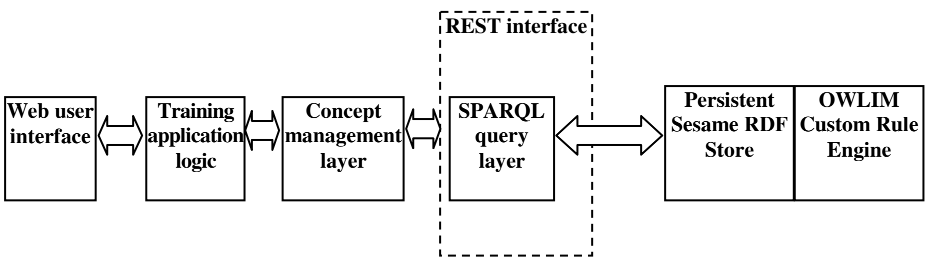


Fig. 1. The general architecture

In this respect, the proposed architecture involves a concept management layer, working as a wrapper for SPARQL queries, applying:

- pre-processing to the data extracted from the interface (building the belief sets to be uploaded on Sesame based on the trainee evaluation or the knowledge engineer interface)
- post-processing to the SPARQL results (generating recommendation links based on incorrect or unstable beliefs, displaying inferred inconsistencies); further development is needed to integrate on this level a delta analysis module for detecting differences between the temporary belief sets of the same trainee.

5 SWOT Evaluation and Conclusions

Strengths

The paper presents a low-cost methodology for representing temporary, subjective knowledge which can be tracked during the learning process, in the context of a computer-based training system.

The proposed solution adds value to learning systems by monitoring closely the evolution of a knowledge portfolio of a trainee and by optimizing the training process with suggestions of content directly related to the trainee's detected insufficiencies.

Weaknesses

In order to respect the low cost requirement, the model employs the free version of OWLIM, SwiftOWLIM, which does not support inconsistency detection. This is compensated by the possibility of customizing rulesets using an intuitive syntax for Horst rules. The lack of OWLIM's support for contextualization of inferences complicates the management of inferences over subjective knowledge.

Opportunities

The emergence of Semantic Web, although slow, will greatly affect the field of e-learning, which must prepare for this and adopt as soon as possible the semantization of content.

Threats

E-learning has its typical threats regarding student motivation, technology adoption, arbitration and security. A more specific threat is related to the slow adoption of Semantic Web technologies in Web development in general, as legacy systems are considered sufficient for the current requirements of eliminating the spatial and temporal costs of learning.

The paper presented some issues regarding the representation of subjective knowledge and its implications on learning systems and learning objects within the context of a computer-based training system architecture based on semantic storage.

Future efforts will be invested in more advanced processing of the temporary belief sets aimed mainly at pattern detection and the relationship between the evolution of belief sets and the concept dependencies.

Acknowledgment. This paper represents results of the research project 2443/2009 funded by the UEFISCSU-CNCSIS Program IDEI, managed by Lect. Robert Buchmann Ph.D.

References

1. The Resource Description Framework, `http://www.w3.org/RDF/`
2. Höver, K.M., Steiner, C.M.: Adaptive Learning Environments: A Requirements Analysis in Business Settings. International Journal of Advanced Corporate Learning 2(3), 27–33 (2009)

3. Beck, R.: What are Learning Objects?, http://www4.uwm.edu/cie/learning_objects.cfm?gid=56
4. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American 284(5), 34–43 (2001)
5. Moodle – the Official Website, http://moodle.org/
6. Blackboard – the Official Website, http://www.blackboard.com/
7. The Web Ontology Language, http://www.w3.org/2004/OWL/
8. Bishop, B.: Primer Introduction to OWLIM, http://www.sirmaitc.com/confluence/display/OWLIM/Primer+Introduction+to+OWLIM
9. Gonzalez-Videgaray, M., Hernandez-Zamora, G., del Rio-Martinez, J.H.: Learning Objects in Theory and Practice: A Vision from Mexican University Teachers. Journal Computers & Education 53(4), 1330–1338 (2009)
10. The IEEE Learning Object Metadata Standard, http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf
11. Bennacer, N., Bourda, Y., Doan, B.: Formalizing for Querying Learning Objects Using OWL. In: Looi, C.T., Sutinen, E., Sampson, D.G., Aedo, I., Uden, L., Kähkönen, E. (eds.) Proceedings of the IEEE International Conference on Advanced Learning Technologies, ICALT 2004, pp. 321–325. IEEE Computer Society, Washington (2004)
12. Bouzeghoub, A., Defude, B., Ammour, S., Duitama, J.F., Lecocq, C.: A RDF Description Model for Manipulating Learning Objects. In: Looi, C.T., Sutinen, E., Sampson, D.G., Aedo, I., Uden, L., Kähkönen, E. (eds.) Proceedings of the IEEE International Conference on Advanced Learning Technologies, ICALT 2004, pp. 81–85. IEEE Computer Society, Washington (2004)
13. Ahmed, S., Parsons, D., Ryu, H.: Supporting Adaptive Learning Interactions with Ontologies. In: Ryu, H., Champion, E. (eds.) Proceedings of the 11th International Conference of the NZ Chapter of the ACM Special Interest Group on Human-Computer Interaction, pp. 17–24. ACM Press, New York (2010)
14. Jovanovic, J., Gasevic, D., Devedzic, V.: TANGRAM for Personalized Learning Using the Semantic Web Technologies. Journal of Emerging Technologies in Web Intelligence 1(1), 6–21 (2009)
15. Open RDF Official Website, http://www.openrdf.org/
16. OWLIM Official Website, http://www.ontotext.com/owlim/
17. SPARQL Official Website, http://www.w3.org/TR/rdf-sparql-query/
18. The Serialization of SPARQL Query Results in JSON, http://www.w3.org/TR/rdf-sparql-json-res/
19. Allemang, D., Hendler, J.: Semantic Web for the Working Ontologist. Morgan Kaufmann, San Francisco (2008)
20. N3 Resources, http://www.w3.org/DesignIssues/N3Resources
21. Defining N-ary Relations on the Semantic Web, http://www.w3.org/TR/swbp-n-aryRelations/
22. SPARQL Issues, http://www.w3.org/2001/sw/DataAccess/issues#countAggregate

Momentum Effect: Developed vs. Emerging Stock Markets

Andrius Svolka[1], Donatas Pilinkus[2], and Edverdas Bartkus[2]

[1] Copenhagen Business School, Solbjerg Plads 3, 2000 Frederiksberg, Denmark
[2] Vilnius University Kaunas Faculty of Humanities,
Muitines 8, 44280 Kaunas, Lithuania
sabas22@gmail.com, pilinkus@pilinkus.com,
edvardas.bartkus@ktu.lt

Abstract. In the recent debates over the informational efficiency of the stock markets, particular attention has been paid to the momentum violating the traditional finance theory which is being taught at business schools. Although many empirical findings support the existence of price momentum, there are still many angles to be researched for the sake of complete understanding of the momentum phenomenon. This paper will emphasize the importance of country selection for the profitability of momentum strategies since academic literature lacks the comparison of the momentum profitability across Emerging and Developed stock markets. Therefore, the main objective of the paper is to investigate the profitability difference of momentum effect between Emerging and Developed stock markets. The study will embrace back-testing stock markets across pre-crisis (year 2005 to 2007) and crisis (year 2008 to mid-2010) periods.

Keywords: Momentum effect, stock markets, efficient market hypothesis, behavioural finance.

1 Introduction

The intuition behind the "momentum strategy" concept is very simple as it effectively suggests that stocks that have done well in the past will also do well in the future. Thus the proposition implies that an investor is capable of making above average profit by simply placing his bets solely on historical price information. However, if past stock prices provide any indication of how stocks will perform in the future, this violates the very core assumption of the traditional finance theory which is being taught at business schools, namely market efficiency.

Although many empirical findings support the existence of price momentum, there are still many angles to be researched for the sake of complete understanding of the momentum phenomenon. Some of the major issues under consideration are: a) statistical significance of transaction costs and risk adjusted profitability of momentum, b) momentum across different countries and asset classes, c) roots and causes of momentum effect, etc.

To extend the empirical research of momentum effect, this paper will emphasize the importance of country selection for the profitability of momentum strategies.

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 111–122, 2011.

Although the existence of price momentum, its duration and magnitude has been documented in a series of empirical studies conducted mainly on major stock markets (Jegadeesh, 2001; Bird and Casavecchia, 2004), academic literature lacks the comparison of the momentum profitability across Emerging and Developed stock markets. Moreover, there is a significant limitation of evidence recorded on the momentum profitability in Central and Eastern Europe (CEE) region and the Baltic countries (Lithuania, Latvia, and Estonia) in particular (Pilinkus, 2010). Therefore, the main objective of the paper is to investigate the profitability difference of momentum effect between Emerging and Developed stock markets.

In order to see whether the level of economic country development has any explanatory power of momentum effect, it will be divided country sample into 2 parts to find out if stock markets of less developed countries are prone to higher or lower over-reaction and subsequently stronger momentum effect. The analysis will be supplemented by back-testing stock markets across 2 time periods: pre-crisis (year 2005 to 2007) and crisis (year 2008 to mid-2010) periods. The purpose of doing it is to see if over-reaction and momentum effect is strengthened by increased volatility in financial markets worldwide. Lastly, it will be investigated the momentum effect from the behavioural finance angle and it will be looked for potential explanation of the momentum effect within the context of Emerging and Developed markets.

2 Theoretical Framework

Basically, there are two major schools of thoughts pertaining concept of market efficiency and subsequently momentum effect. Those are the efficient market hypothesis (EMH) and behavioural finance theory. As central finance theory, EMH lost its ground in recent decades. Dozen of studies conducted by academics provided a number of anomalies and cases where EMH has been disproved, and was unable to reconcile with the phenomenon of the financial markets. Alternatively, behavioural finance advocates provided sound arguments and explanations for certain anomalies and market inefficiencies. Main propositions of both theories and their links to momentum effect are presented next.

2.1 Efficient Market Hypothesis (EMH)

EMH is the proposition that prices fully reflect information available to market participants, i.e. hedgers and speculators, and there are opportunities neither for the hedgers nor for the speculators to make super-normal profits, meaning both speculative efficiency and arbitrage efficiency exists in the markets (Nguyen, 2000).

The general conclusion from numerous studies in Developed countries, beginning with Fama (1970) is that the weak-form of market efficiency holds and that no exploitable risk and transaction costs adjusted patterns in past trading records exist.

Evidence from stock markets in the emerging markets, however, is mixed. For example, Dickson and Muragu (1994) found evidence consistent with EMH in their study of the Nairobi Stock Exchange, whereas Barnes (1986) study of the Kuala Lumpur Stock Exchange provided only limited support of the weak form of the EMH. Zychowicz et al. (1995) concluded that on the Istanbul stock exchange, daily and weekly returns diverge from random walk, while monthly returns are consistent with weak-form EMH.

Mobarek and Keasey (2000) in their work did not arrive to find a supporting argument for weak-form efficiency in Dhaka Stock market (Bangladesh). Balaban and Ercan (1996) presented the evidence from Turkish capital markets, which once again was not consistent with weak-form EMH. Gupta and Basu (2007) tested weak-form efficiency hypothesis on Indian Stock Exchange and found that series do not follow random walk and there are an evidence of autocorrelation, thus rejected weak-form EMH.

Market efficiency test performed on Central European Stock Markets (such as Budapest, Prague and Warsaw Stock Exchanges) by Bechev (2004) found some evidence of weak-form efficiency in Budapest Stock Market. However, Bechev didn't find any support for market efficiency in Prague and Warsaw stock markets.

Studies performed by other authors on the stock market efficiency hypothesis in emerging markets were mostly unfavorable to EMH, even though some authors did find slight convergence of the emerging markets to something more like weak-form EMH (Hall et al. 1998; Hanousek and Filer, 1999; Bechev, 2004).

Aforementioned empirical evidence suggests that the panacea of efficient markets is quite remote from the reality in the financial markets. Although developed markets are considered to be converging to higher efficiency, there are quite a few "holes" to be exploited by market players. One of the "holes", which could potentially provide abnormal profits, is momentum investing.

2.2 Momentum Effect

In its very basics, momentum investing is a principle of buying stocks that have had high returns over the recent periods, and selling those stocks that have been performing poorly. More specifically, momentum characterizes certain stock price behaviour as prolonged price movement continuation. A formal definition of price momentum is the persistence of observed past price change.

The existence, duration and magnitude of price momentum have been documented in a series of empirical studies conducted on stock markets in different countries (Jegadeesh, 2001; Bird and Casavecchia, 2004). Most of the authors came up with the conclusion that momentum investing is statistically profitable strategy (Jeegadesh & Titman, 1993; Rouwenhorst, 1998; Dijk & Huibers, 2003; Griffin, Ji & Martin, 2003). The main disagreement between the academia is, however, the source of the momentum effect.

2.3 Behavioural Finance (BF)

Behavioural Finance (BF) theory is a new approach to finance. It emerged during the late 1980's. Up until this point the EMH dominated the academic research within the field of finance and there existed a vast amount of evidence supporting the hypothesis.

BF in its essence is the application of psychology to financial behaviour. Basically, the idea is that irrational behaviour among investors causes prices to deviate from fundamentals. The same irrationality prevents the correction of prices back to fundamentals through the process of arbitrage, because irrationality introduces a new dimension of risk. This new dimension of risk is known as noise trader risk. It is systematic and cannot be diversified away. One of the greatest achievements of BF is that is showed, that due to limits to arbitrage, deviations from fundamentals can persist.

These propositions by BF rest on two pillars – psychology and limits to arbitrage. BF has shown, by employing research done by cognitive psychologists, that the behaviour of investors is often irrational. This irrationality stems from several factors broad on by limited mental capacity to process information and emotions. These factors cause severe biases (Tversky and Kahneman, 1979; Shefrin, 2000) when investors form their expectations and this leads to mispricing in financial markets. The discovery that there is limits to arbitrage, help facilitate that these mispricing will not be corrected by arbitrageurs and hence may persist also in the long run.

3 Methodology

In order to test the profitably of momentum effect across international stock markets, the back-testing technique will be applied on the daily time series of country stock indexes. Real investing environment will be replicated by back-testing technique. This method is superior to standard (normal distribution based) models due to its up-front set-up of investing rules. This diminishes the threat of data mining and represents the results based on attainable data (not statistical approximation).

Firstly, the back-testing method will be applied on the sample of 35 international stock market indexes worldwide. The central question of the paper is to investigate whether there is a difference of momentum investing profitability in Emerging versus Developed stock markets. Therefore, the sample is divided into two groups: a) stock indexes of emerging countries (total of 16 indexes), and b) stock indexes of developed countries (total of 19 indexes). Although country stock index does not fully reflect the behaviour of the whole country, it is the best approximation attainable. To distinguish between country development level, GPD per capita measure is chosen, i.e. the higher GDP per capita indicates higher level of country development, thus lower irrationality by investors and subsequently lower profit potential by momentum investing strategies.

Secondly, in order to evaluate the significance of market volatility as a component of momentum profitability, sample countries are back-tested on two different time periods: a) pre-crisis period (normal state of volatility), and b) crisis/post-crisis period (increased volatility). Selected pre-crisis period ranges between mid-2005 to 2007, while crisis/post-crisis period ranges from 2008 to mid-2010. Structurally, the whole sample to be tested is divided into four data sets.

3.1 Investment Strategy

The back-test of momentum effect will be kept within simple investing strategy. Selecting simple strategy makes the predictability more compelling and minimizes inevitable data snooping issues. The focus will be placed on only widely used price and simple moving average (SMA) cross-over strategy. A moving average of past prices is formed by taking

$$SMA_t^m = \frac{1}{m}\sum_{i=0}^{m-1} p_{t-i} \tag{1}$$

where the prices (p_t) are sampled on the daily basis. Standard trend following momentum trading signal will be generated when price (p_t) crosses SMA_t from the above or below. If price (p_t) crosses SMA_t from below, the strategy would generate a buy signal suggesting strengthening momentum relatively to historic prices. The position is maintained until the price crosses SMA_t from above, at which point the position will be reversed to the short-sell. Buy and sell signals are executed on the opening price of the subsequent day after the signal has been generated.

All simple moving average strategies are described by a single parameter, m, which indicates the number of periods taken into average, and in this paper will vary between minimum of 3 days (m=3) to maximum of 50 days (m=50). Simply speaking, there are total of 48 different strategies for each stock market index to be back-tested.

After back-testing, 50% of best performing strategies per each country will be filtered for further investigation and profit evaluation, i.e. 24 best performing strategies per stock index (total 35 indexes) in each of two time periods. Overall, there are 24 x 35 x 2 = 1680 strategies to be investigated.

3.2 Data Set and Transaction Costs

In order to examine the differences between Emerging versus Developed stock markets, it will be looked into the local stock indexes and it will be performed back-testing technique based on the daily data of countries' stock indexes. The data is provided by Bloomberg terminal. The software used to perform Back-testing optimization is also provided by Bloomberg terminal integrated functionality.

There are total of 35 stock indexes to be examined through the back-testing technique. 19 of these stock indexes represent Developed stock markets, while remaining 16 stock indexes are drawn from Emerging stock markets.

To replicate real investing environment, every buy and sell trade will be adjusted for transaction costs, which is 0.25% of the trade value. In reality the transaction cost will vary country by country and should be smaller for Developed stock markets compared to emerging stock markets. Although short-selling restriction might be in effect in some countries, the assumptions of both buy and short-sell possibility is made. The assumption will not impair the results, as the main purpose of the paper is to investigate the momentum effect presence, but not to precisely estimate the profitability of each strategy.

4 Back-Testing Results

The output of this part is twofold. Firstly, the results of momentum profitability of Emerging versus Developed stock markets are presented across two time periods.

Secondly, the statistical significance of hypothesized relationship of higher momentum profitability across Emerging stock markets will be tested with the help of regression.

4.1 Pre-crisis Period Results

The main output from the back-testing process is presented in tables below. It is seen from table 1 that on average the sample of stock market indexes across Emerging markets resulted in 10,2% annual return.

Table 1. Momentum investing profitability across Emerging stock markets during pre-crisis period

Country	Annual return	P/L	%max DD	Sharpe	Average duration	Dummy
Brazil	-1,8%	0,94	26,2%	-0,54	18,02	0
Bulgaria	38,3%	6,42	10,9%	1,64	37,83	0
China	-3,4%	0,86	25,1%	-1,26	21,62	0
Croatia	40,8%	4,57	16,2%	2,01	21,57	0
Czech	-9,0%	0,67	29,1%	-1,54	18,48	0
Estonia	31,2%	3,72	14,2%	1,29	25,96	0
Hungary	-11,2%	0,59	42,2%	-1,28	14,53	0
India	19,9%	1,79	23,0%	1,48	24,59	0
Latvia	18,2%	1,74	20,8%	0,66	17,09	0
Lithuania	30,3%	2,60	12,6%	1,76	23,72	0
Mexico	9,9%	1,47	17,6%	1,29	24,27	0
Poland	-13,7%	0,67	48,6%	-0,24	13,87	0
Romania	8,2%	1,28	32,5%	0,66	18,73	0
Russia	10,4%	1,35	29,2%	1,13	23,02	0
Slovakia	-5,5%	0,75	28,6%	0,19	19,53	0
Turkey	0,8%	1,02	33,3%	0,89	15,93	0
Average	**10,2%**	**1,90**	**25,6%**	**0,51**	**21,2**	

The profitability result, however, varies from country to country, e.g. on one extreme there is Polish stock index, which provided 13,7% annual loss if traded on momentum based strategy. On the other extreme there is Croatia and Bulgaria with an impressive annual return of 40,8% and 38,3% respectively. Highest risk-adjusted result is provided by Croatia, Lithuania and Bulgaria.

Overall, although the momentum investing resulted in an average 10% annual profit, the “buy and hold” strategy in Emerging stock markets would have been more profitable. Equally weighted portfolio of Emerging stock markets would have resulted in an annual return of 27,4%, with no single loosing country stock market. “Buy and hold” strategy outperforms the momentum based investing mainly due to lower transaction costs.

The results from the sample of Developed stock markets are presented in table 2. Overall the profitability of momentum strategies across Developed markets is negative (-3,6%), with 14 out of 19 stock indexes resulting in absolute loss. In comparison, “buy and hold” strategy across Developed stock markets would have given a positive result, 13,6% per annum. The difference in results is mainly explained again by higher transaction costs of momentum strategies.

Table 2. Momentum investing profitability across Developed stock markets during pre-crisis period

Country	Annual return	P/L	%max DD	Sharpe	Average duration	Dummy
Australia	-10,3%	0,58	32,0%	-1,12	17,74	1
Austria	-0,5%	1,01	29,7%	0,46	19,13	1
Belgium	-1,5%	0,95	22,9%	0,50	22,00	1
Denmark	-8,8%	0,69	31,8%	-1,15	13,79	1
Finland	5,0%	1,23	21,6%	0,95	20,88	1
France	-12,0%	0,51	31,8%	-1,67	17,13	1
Germany	1,5%	1,12	17,1%	-0,33	23,94	1
Greece	-4,1%	0,86	29,6%	-0,03	19,46	1
Ireland	5,2%	1,28	17,8%	-0,08	20,30	1
Italy	-5,7%	0,69	20,0%	-1,70	21,23	1
Japan	-8,6%	0,76	37,3%	-0,23	12,71	1
Netherlands	-1,5%	0,92	15,7%	-1,19	23,09	1
Norway	1,1%	1,05	28,1%	0,77	19,90	1
Portugal	17,3%	2,27	14,4%	1,23	28,41	1
Spain	-9,5%	0,64	33,9%	-1,04	15,30	1
Sweden	0,0%	1,01	26,6%	0,49	21,17	1
Switzerland	-6,0%	0,75	29,2%	0,28	19,02	1
UK	-13,9%	0,41	38,9%	-0,80	15,96	1
US	-16,4%	0,32	36,5%	-2,80	15,58	1
Average	**-3,6%**	**0,9**	**27,1%**	**-0,39**	**19,3**	

It is worth noticing that two most developed stock market in the world (US and UK) in terms of its volume traded and popularity among investors have the highest negative return of momentum strategies. This is an indication of both being somewhat more efficient than its peers.

During the period of normal market volatility (pre-crisis period), the momentum strategies on average recorded greater momentum profitability potential in Emerging stock markets compared to Developed markets. In fact, most momentum strategies across Developed markets were loss making strategies. Nevertheless, the momentum investing did not outperform the "buy and hold" strategy neither for Emerging nor Developed stock market indexes. Risk-adjusted efficiency measure of the strategy is presented by Sharpe ratio. Emerging markets posted significantly better Sharpe ratio as compared to Developed markets, 0,51 versus minus 0,39.

The average duration measure indicates an average length of how long each trade was kept before reverting it into the opposite position. Across Emerging markets it is slightly longer (21.2 days) as compared to Developed markets (19.3 days). The difference is not significantly different, thus it is difficult to draw any viable conclusions. There is an inverse relation between the level of country development and the profitability of momentum strategies, i.e. as once country becomes richer and subsequently more efficient, the profitability potential derived from momentum strategies decreases. However, momentum profits are highly feasible across less developed countries.

For the statistically confirmation, the regression analysis is performed. Each country is assigned a dummy variable, either 0 or 1, according to the definition of its economic development level. Countries which are considered as Emerging are assigned dummy variable 0, while Developed countries are assigned 1. The following regression has been run:

$$Annual\ return = \alpha + \beta \times Dummy + \varepsilon \tag{2}$$

The regression result from the pre-crisis period is presented in table 3.

Table 3. Pre-crisis period results of the regression analysis

Annual return =	***0,1 –***	***0,14 x Dummy***
t-test	*3,01*	*-2,89*
p-value	*0,005*	*0,007*

Both intercept and slope coefficients are statistically significant, meaning that the inverse relationship between the level of country development and profitability of momentum investing actually exits. In other words, the level of country development could be used as determinant when choosing country index for momentum investing exploitation. The explanatory power of such regression is 45% according to R^2.

4.2 Crisis/Post-crisis Period Results

The period of crisis/post-crisis is defined as the one with historically unseen volatility. Due to the crash of financial system, fear has been spread all around the financial markets globally with extreme levels of volatility. This paper intends to analyze the importance of volatility for the improved profitability results. From table 3 below, it is seen that during crisis/post-crisis period there was a huge profitability potential for momentum strategies. The average return on stock indexes of Emerging markets was about 30% on annual basis. Bulgaria, Lithuania and Estonia are the leaders of the table with impressive return ranging from 60% to 80% per annum. There was only one country (Slovakia) which had negative result (-4,7%).

Buy and hold strategy of equally weighted portfolio across Emerging markets would have resulted in an annual loss of 15,4%, thus indicating a strong under-performance in comparison to momentum strategies. The Sharpe ratio did not increase significantly as compared to pre-crisis period, due to increased risk in the market. Average position holding duration has not changed significantly either as compared to pre-crisis period.

Table 4. Momentum investing profitability across Emerging stock markets during crisis/post-crisis period

Country	Annual return	P/L	%max DD	Sharpe	Average duration	Dummy
Brazil	2,2%	1,10	34,9%	-0,14	20,4	0
Bulgaria	82,8%	3,83	15,3%	0,78	22,4	0
China	13,2%	1,43	25,5%	1,14	21,2	0
Croatia	58,9%	4,04	20,9%	2,06	27,1	0
Czech	11,7%	1,26	37,1%	0,48	17,1	0
Estonia	68,2%	3,01	18,8%	1,36	18,7	0
Hungary	17,0%	1,38	38,6%	0,85	17,8	0
India	17,1%	1,51	28,1%	0,58	22,2	0
Latvia	17,2%	1,52	31,5%	0,74	21,1	0
Lithuania	78,7%	5,55	15,2%	1,97	25,6	0
Mexico	13,8%	1,75	17,5%	0,21	26,8	0
Poland	1,5%	1,09	33,1%	-0,07	18,2	0
Romania	57,3%	2,57	23,5%	0,75	25,2	0
Russia	15,7%	1,31	39,8%	0,52	18,8	0
Slovakia	-4,7%	0,79	24,8%	-1,21	25,5	0
Turkey	27,1%	1,79	34,5%	0,94	19,3	0
Average	**29,8%**	**2,12**	**27,4%**	**0,68**	**21,7**	

Increased volatility also improved the profitability of momentum strategies across Developed markets. Although loss making during low volatility period, Developed markets on average posted 7% return during the turmoil period. During the period, only 9 out of 19 countries were loss-making, moreover, the size of loss was not as big as in pre-crisis period. Buy and hold strategy for Developed countries was negative at minus 16,4%, once again confirming the proposition that increased volatility has a significantly positive affect on the momentum investing. Sharpe ratio in crisis/post-crisis period was also positive.

Table 5. Momentum investing profitability across Developed stock markets during crisis/post-crisis period

Country	Annual return	P/L	%max DD	Sharpe	Average duration	Dummy
Australia	1,9%	1,09	21,5%	0,10	17,7	1
Austria	23,2%	1,60	27,7%	0,74	21,4	1
Belgium	-2,2%	0,95	35,4%	0,33	14,0	1
Denmark	-10,0%	0,80	39,8%	-0,55	10,9	1
Finland	-5,7%	0,88	32,4%	-0,40	15,8	1
France	1,1%	1,03	25,1%	0,29	19,8	1
Germany	-0,9%	0,99	30,3%	0,34	15,7	1
Greece	55,0%	2,74	22,0%	0,80	25,1	1
Ireland	22,7%	1,77	24,4%	0,62	21,6	1
Italy	28,0%	2,31	17,9%	1,10	24,0	1
Japan	11,6%	1,63	20,3%	0,56	27,7	1
Netherlands	6,1%	1,14	31,7%	0,38	17,5	1
Norway	27,1%	1,79	29,8%	0,38	22,8	1
Portugal	19,2%	2,02	21,0%	1,36	22,9	1
Spain	-1,7%	0,96	31,4%	0,00	18,1	1
Sweden	-18,9%	0,59	52,0%	-1,30	16,1	1
Switzerland	-13,8%	0,61	39,1%	-1,88	16,0	1
UK	-2,5%	0,94	34,1%	0,20	18,6	1
US	-6,8%	0,80	34,2%	-1,02	17,9	1
Average	**7,0%**	**1,3**	**30,0%**	**0,11**	**19,1**	

Analogical check of statistical soundness has been run on the sample during high volatility period. The results are shown below:

Table 6. Crisis/post-crisis period results of the regression analysis

Annual return =	***0,28 –***	***0,19 x Dummy***
t-test	*4,53*	*-2,33*
p-value	*0,00*	*0,03*

Both coefficients are significant within 95% level of confidence, indicting a strong relation between the state level of country development and momentum potential. Less developed the country is, the better momentum investing opportunity it provides. Moreover, profitability opportunity of momentum investing is amplified by increased volatility.

5 Conclusions

The efficiency of international stock markets has been historically questioned for many years. EMH was the foundation of finance theory suggesting at least weak-form of stock market efficiency. However, a substantial amount of empirical evidence has been presented recently, which to the large extent confirmed market deviation from the weak-form efficiency. Specifically, there were quite a few market anomalies documented, which EMH was incapable to reconcile with. One of them is “momentum” effect.

The empirical study performed in this paper questioned the feasibility of profits derived from solely momentum investing strategies. The sample of total of 35 countries was sub-divided in to 2 groups according to the level of its economic development. The back-testing technique was performed over the period of 2005 to 2010. Total of 1680 momentum strategies were investigated.

Major findings of the paper are:

a) In low volatility environment, momentum effect is significantly higher among Emerging countries as compared to Developed countries. However, momentum investing would have not out-performed simple “buy and hold” strategy across both Emerging and Developed market indexes.
b) In high volatility environment, momentum effect is strengthened to the level that momentum investing across both Emerging and Developed countries would have significantly out-performed “buy and hold” strategy. There is also evidence in favour of substantially stronger momentum effect across Emerging as compared to Developed countries.
c) The profitability of momentum strategies is only marginally caused by higher risk undertaken, as suggested by risk adjusted profitability measure. Therefore, the explanation must hide under the psychological state of local stock market players.

References

1. Balaban, E., Kunter, K.: Financial Market Efficiency in a Developing economy: The Turkish Case. The Central Bank of the Republic of Turkey. Discussion paper no: 9611 (1996)
2. Bechev, I.: Efficient Market Hypothesis: Budapest, Prague and Warsaw Stock Exchanges. Central European University (2004)
3. Bird, R., Casavecchia, L.: The Profitability of Price and Earning Momentum across the European Markets. University of Technology, Sydney (2004)
4. Brush, S.J.: A flexible Theory of Price Momentum. The Journal of Investing 16(1), 36–42 (2007)
5. Daniel, K., Hirshleifer, D., Subrahmanyam, A.: Investor Psychology and Security Market Under- and Over-reactions. The Journal of Finance LIII(6), 1839–1885 (1998)
6. DeBondt, W.F.M., Thaler, R.: Does the Stock Market Overreact? The Journal of Finance XL(3), 793–805 (1985)

7. Dickson, J.P., Muragu, K.: Market Efficiency in Developing Countries: A Case Study of Nairobi Stock Exchange. Journal of Business Finance and Accounting 21(1), 133–150 (1994)
8. Dijk, R., Huibers, F.: European Price Momentum and Analyst Behavior (2002)
9. Du, D., Zhao, X.: Momentum and Autocorrelation in Stock Returns. International Research Journal of Finance and Economics (50), 157–173 (2010)
10. Fehr, E., Tyran, J.-R.: Individual Irrationality and Aggregate Outcomes. Journal of Economic Perspectives 19(4), 43–66 (2005)
11. Griffin, J.M., Ji, X., Martin, S.: Momentum investing and Business Cycle Risk: Evidence from Pole to Pole. The Journal of Finance 58(6), 2515–2547 (2003)
12. Gupta, R., Basu, P.K.: Weak Form Efficiency in Indian Stock Market. International Business & Economic Research Journal 6(3), 57–64 (2007)
13. Hall, S., Urga, G., Zalewska-Mitura, A.: Testing for Evolving Market Efficiency with an application to Russian stock prices. London Business School Discussion Paper No. DP 12-98 (1998)
14. Hanousek, J., Filer, R.: The Extent of Efficiency in Central European Equity Markets. Charles University in Prague, Center of Economic Research and Graduate education (1999)
15. Helmenstein, C.: Capital markets in Central and Eastern Europe. Journal of Banking and Finance 24(12), 1997–1999 (2000)
16. Hong, H., Stein, C.J.: A Unified Theory of Under-reaction, Momentum Trading and Over-reaction in Asset Markets. The Journal of Finance 54(6), 2143–2184 (1999)
17. Jegadeesh, N.: Momentum. Univeristy of Illinois (2001)
18. Jagadeesh, N., Titman, S.: Returns to buying Winners and Selling Losers: Implications for Stock Market Efficiency. The Journal of Finance 48(1), 65–91 (1993)
19. Mobarek, A., Keasey, K.: Weak-form market efficiency of an emerging Market: Evidence from Dhaka Stock Market of Bangladesh. University of Leeds, UK (2000)
20. Nathan, D.: Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. Bar Ilan University (2006)
21. Pilinkus, D.: Macroeconomic Indicators and Their Impact on Stock Market Performance in the Short and Long Run: The Case of the Baltic States. Technological and Economic Development of Economy 16(2), 291–304 (2010)
22. Rouwenhorst, K.G.: International Momentum Strategies. The Journal of Finance 53(1), 267–284 (1998)
23. Swinkels, L.: Momentum Investing: a Survey. Journal of Asset Management 5(2), 120–143 (2003)

Assistive Tools for the Motor-Handicapped People Using Speech Technologies: Lithuanian Case

Vytautas Rudžionis, Rytis Maskeliūnas, and Algimantas Rudžionis

Department of Informatics,
Vilnius university Kaunas faculty, Lithuania
Kaunas university of technology, Lithuania

Abstract. The paper presents analysis of the possibilities to use voice technologies for the partial integration of people with disabilities. The particular interest has been expressed to the motor-handicapped people. The special wheelchair with the voice command recognition capabilities has been designed. Evaluation of command's recognition accuracy shows high dependency on the proper detection of the utterance boundaries. The acoustic boundaries detection algorithm has been proposed. This algorithm allowed achieve high accuracy of the detection of acoustic events boundaries such as words or phrases even in the presense of high noise. The proper detection leads to the increased accuracy of voice commands recognition and the overall satisfaction of users.

Keywords: voice technology, voice command recognition, motor-handicapped people, acoustic events, detection of people speaking.

1 Introduction

People with disabilities meet barriers of all types. However, technology can help to lower many of them. Using computing technology tasks such as reading and writing documents, communicating with others, searching information on the internet, and controlling or even adapting the surrounding environment people with disabilities are capable of handling a wider range of activities independently. However typical input modes used in modern control and communication devices – keyboard input and visual output – are often not well suited for the disabled people. Voice technology is often the most preferable way for such people – to use speech recognition for recognizing commands as a substitute for keyboard or mouse based control and to use speech synthesis to read the content of computer screen as a substitute of typical screen reading by eyes. The ability to control the home is an essential aspect of independence and e-inclusion.

The implicit richness of human speech communication gives the user many degrees of freedom for control and input of various devices [1]. The speed of speech recognition also gives it a potential advantage over other input or control methods. Applications of speech technology can be grouped in the areas of access, control, communication and rehabilitation\therapy. For people with different impairments different types of speech technologies are more important: for people with visual

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 123–131, 2011.

impairments speech synthesis is essential as a way to access information, for people with hearing impairments perceptual speech processing and amplification are crucial, for other disabilities other areas of speech technology can be more important. But it is really difficult to find people with some sort of impairment that can not benefit from one or another aspect of voice technology.

The barriers arising for the people with disabilities may be well shown on the example of the GUI development. One of the major advances of the human-machine interfaces in the recent decades was the advent of the graphical-user interfaces (GUI). It is hard to predict how much GUI was the reason of the massive computerization of the society but it is undoubtedly that GUI had significant impact. GUI popularized icons and the use of mouse for computer navigation which is generally extremely comfortable way to communicate with the computer for the majority of people. GUI is wonderful for the WYSIWIG (what you see is what you get) world, but is inconvenient for the people with various impairments such as visual, hearing loss or motor-handicapped persons. It is noted that prior to the GUI era [2] it was easier for impaired people to work with the computer using a Braille reader to assist them. In those days mainframes with character-based user interfaces were mainly used. System navigation was done using menus, tabbing, and function keys – all of which could be learned on a keyboard by the blind people. If the situation for the hearing impaired or the motor-handicapped people might look different in fact it is not a far better. And everyone can see the obvious reason for that: impaired or disabled people can't use one or more modalities that conforms the foundation of modern human-machine interfaces – monitor, keyboard or mouse. Other modalities are necessary in such situations. Well known fact is that speech is the preferable modality for the majority of the disabled people. If the person is mobility impaired, and cannot use their hands to move a mouse or type, issuing a voice commands is the natural solution. If people can't see, then applying text-to-speech technique to read the content is the most convenient solution. Voice technologies are the key element in the devices that are developed to satisfy the needs of many impaired people.

The success of the development of specialized tools for impaired people mainly relies on two factors: development of the voice technology being used as well as the knowledge of the special requirements of the disabled people. It should be emphasized that disabled people are especially valuable users from the point of view of voice technologists since due to the physical limitations they are ready to use even technologically restricted applications and not very well developed technologies that normal people often simply refuse to do.

The main group of interest which needs is addressed in this study is the motor-handicapped people. The characteristic property of such category of people is that they often simply can't use traditional keypad based control systems independently or the use of such systems is significantly restricted. Environmental Control Systems (ECS) or Smart home control interfaces are available which address many elements of home management for disabled people, such as control of audio-visual equipment, telephones, household appliances, doors and curtains as well as the ability to summon assistance. Most ECSs utilize switch-scanning or keypad interfaces for control. More recently, ECSs with speech recognition have been introduced and a number of such systems are available on the market. Their success depends on a number o factors most important of them being maturity of voice processing technology used. Even better results could be achieved implementing multimodal approach – combining several different modalities to work in parallel or supplementing each other. In example, a multi-modal interaction framework using speech recognition and computer vision to model a new generation of interfaces in the residential environment was developed in [3]. The design is based on the use of simple visual clues and speech interaction. The latter system incorporates video information processing block which moves this system to the class of multimodal systems. Experience shows that motor-handicapped people are keen to use voice technology. This is especially true for people with hand movement restraints were the use of voice recognition is the only mode to transfer a computer control commands.

Very important characteristic of voice based interfaces is the dependability of the phonetic, syntactic and lexical properties of the language spoken by the user. This means that it is impossible to move technologies developed for the recognition of one language for the recognition of another automatically. Some sort of adaptation would be necessary. Since major developers of speech technologies aren't particularly interested in less spoken languages such as Lithuanian the need for adaptation in such cases is even more important. One of the possible solutions for some class of applications is the adaptation of foreign language based speech engines via the selection of proper phonetic transcriptions. In our previous studies the advantages of such method and its possible uses were established [4, 5]. But the success of voice based interfaces significantly depends on the proper detection of speech in long recordings: such devices typically work in the continuous recording mode while only very small part of recording contains useful information. At the same time often users need to operate in the noisy environments what makes the problem of detection of the speech boundaries not trivial.

Further the paper is organized as follows: the second part presents voice controlled wheelchair with the possibilities to implement multimodal control opportunities. The third part presents acoustic events detection algorithm and it’s evaluation in different acoustic environments. And finally achieved results and further prospects are summarized.

2 Wheelchair with Multimodal Control Possibilities

Wheelchair is one of the main assistive tools used by the motor-handicapped people in their daily lives. Wheelchair provides the opportunity to overcome the main limitation of this type of people – the inability to move independently. There were many different types of wheelchairs proposed and used in practice. The simplest type is manually controlled and relies on the assistance provided by the third person: the supporting person moves the wheelchair to the place were disabled person wish to be. Such wheelchairs has obvious disadvantage: they need the assistant person to be in a close proximity to the disabled person to be called when necessary. Often the need to ask for an external help arouses psychological problems to the disabled person since the necessity to ask for a help each time emphasizes dependency from other people. The reliance on own hands and the possibility to use own force to move such type of wheelchairs isn't possible for all people with motoric system disabilities. Another type of wheelchairs uses the electric motor and battery to move the wheelchair. The user typically controls the equipment with the small keyboard or even some kind of joystick type device. Being significantly more convenient than pushcart wheelchairs such kind of tools has also several drawbacks. One of them is the dependency from battery: the more it is used the more likely it will require loading. The loading means that wheelchair will be unavailable for the disabled person for a while. Another drawback is that it is difficult and inconvenient to use it in the small spaces such as living rooms, corridors, etc. This means that it would be highly desirable to free the person from the necessity to move when doing such tasks as switching on/off lights, turning on/off radio, etc. And for some types of diseases even the using of keyboard or joystick to control the wheelchair is problematic.

There were successful attempts to do wheelchairs using voice commands as the mode for the control. Such wheelchairs typically has embedded voice command recognition and control unit designed to recognize and process pre-specified set of commands. From a Lithuanian speaker perspective it is very important that such wheelchairs recognizes only English (or some other language) commands and it is difficult to embed the recognition of Lithuanian commands.

These considerations suggested us to propose client-server based architecture for wheelchair control: the user is provided with the PDA type device which serves as the recorder and does some initial processing of speech signal and transmits it to the server. Server runs speech recognition engine, receives voice commands, recognizes them and makes appropriate turns. The PDA client and server are linked using Bluetooth or wireless connection. Such approach enables to expand human-computer interaction with the additional modes easier and in a more flexible way in the future. The Fig. 1 shows the principal schema of client-server structure of wheelchair voice command based control system:

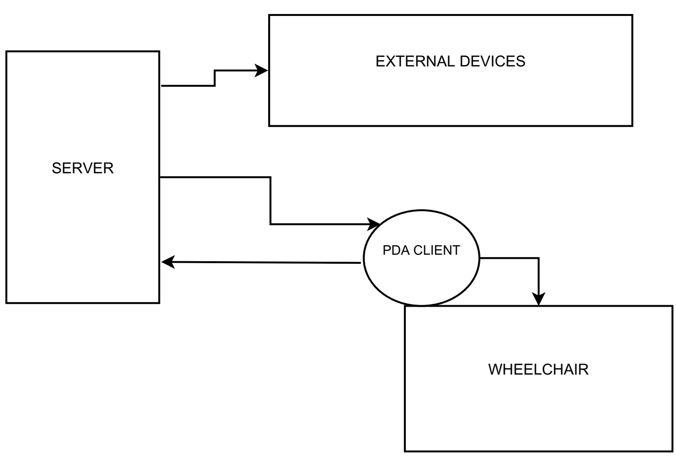


Fig. 1. Client-server architecture of voice command based control system for wheelchair and additional devices

The Fig. 2 shows the wheelchair equipped with the voice commands recording and processing equipment.

Fig. 2. Wheelchair with equipment for the recognition of voice commands

The main advantage of such approach is it flexibility: since nowadays houses are equipped with a set of household appliances ranging from simple lights to feature-rich hi-fi systems, DVD players and TV sets then it is possible to use voice command recognition system to control this big variety of home appliances not only the wheelchair. In this case speech server has controller with executive unit connected with selected appliances. Another advantage is the possibility to implement other than voice commands control possibilities or to use them as the additional channel for the commands. In this case it is possible to organize multimodal wheelchair control system.

From the developer's point of view the main advantage is the possibility to implement wide range of voice commands recognition algorithms. It also includes the possibility to use third party or foreign language speech recognition engines and to implement the experience gathered adapting foreign language speech recognition engines for the recognition of Lithuanian voice commands. It has been shown earlier that such adaptation enables to achieve voice command recognition accuracy necessary for the commercial applications (95% and above) while applying only limited resources. To achieve this goal in open and noisy room conditions some additional requirements should be met. One of them is the proper detection of acoustic boundaries in the noisy recordings: since it is more convenient for user to record audio signal continuously and to recognize command just after the utterance was finished rather than to press a key and begin to talk we need to detect reliably the boundaries of acoustic events (words, phrases, utterances). The reliability should be maintained even if the environment is noisy. The next chapter describes proposed method for the detection of the acoustic events in long or continuous recordings.

3 Detection of Acoustic Events in Long Utterances

The detection of the boundaries of acoustic events such as utterances in the long recordings, utterances in the noisy environment or the phoneme boundaries within a word is one of the most fundamental problems in the area of speech processing. It is not surprisingly that a lot of activities were devoted to solve this problem. Various algorithms proposed for the detection of speech and segmentation of spoken utterances are presented in [6-9] and others sources. Most of the algorithms exploit such spoken signal properties such as the articulatory movement's features or the differences between the actual signal spectrum and the spectrum prediction using its first or second order regression. The selection of those features are based on the analysis of the physical properties of speech signal, e.g. articulatory movements features describe the particular structure of the speech signal spectrum which is typical only for the transitions between various.

Many methods also are based on the signal energy changes as one of the factors, which reflect best the acoustic changes in a speech signal.

In this study detection of acoustic boundaries is important as a template approach for the detection of phoneme boundaries using visual features: the quality of the detection using visual features should be compared with the results achieved using the detection based on articulatory features. Of course it could be possible to use manual segmentation of spoken signal but in this case the complexity of the study will grow enormously. The combined use of the acoustic and visual segmentation (e.g. for the improved recognition of speech or the detection who from a group of people is speaking) is still the future goal and isn‘t covered by this study. But we need to use robust algorithm to segment the spoken signal to use automatically generated boundaries of acoustic events as a templates for visual features. In this study we used proper algorithm for acoustic events detection which showed to be accurate and robust enough for the segmentation of spoken speech in previous study [10].

In this study we used slightly modified algorithm. The essence and the background could be explained as follows. It is well known that the speech is non-stationary process over longer time spans. At the same time speech could be considered as a quasi-stationary process over shorter time periods (a time frame is no longer than 30 ms though the exact duration of the stationarity depends on the phonetic content of a signal). Most algorithms count on the periods, where the statistical properties of stochastic process change moderately. Since speech signal could be described as a process with time varying frequency, the properties of a speech signal in the frequency (or spectral) domain is rather informative.

At the initial stage the logarithmic spectrum was derived using 8-10 msec step. Experimental evidence proved that the spectrum based on the recursive IIR filter bank is more robust to fluctuations of the spectrum properties on the adjacent speech frames. These vectors were used to construct the likelihood function of the changes in spoken speech. The changes in the likelihood function values enables to capture such highly indicative acoustic event features as the articulatory movements. Comparing with the method described in [10] we used shorter period for the integration of the likelihood function. The integration is necessary since it allows obtaining smoother likelihood function and helps to avoid the random type of fluctuations which are characteristic for the changes in spectral properties of many phonetic units. The function f_{SR} was used to perform smoothing by converting the sequences of the likelihood function parameters r_t to smoothed sequence of the parameters ρ_t:

$$\rho_t^{SR} = f^{SR}(r_t, \tau, \nu) \qquad (1),$$

were the parameters τ and v defines the smoothing duration and the smoothing type. The sequence ρ_t is used for the detection of the boundaries of acoustic events and it is called an acoustic events response (AER). The examples of AER are presented in the Figure 1, which consists of the oscilogram of the original and the differentiated syllable, the spectrogram of the same syllable and the AER curve. It is expected that the changes in the acoustical content of a signal will occur on the places where the AER curve reaches the local maximum. The higher is a peak of the AER the higher is the likelihood value of the boundary between the different acoustic units.

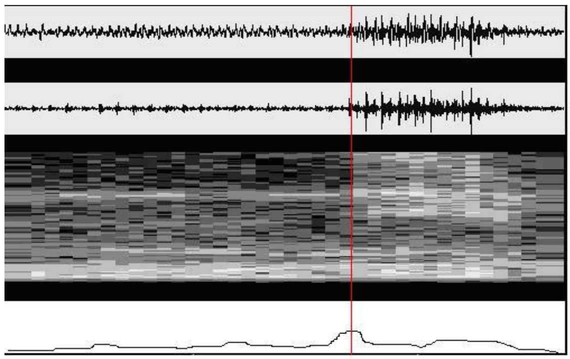


Fig. 3. The acoustic events likelihood function within an isolated word

The efficiency of basic algorithm for the detection of endpoints of spoken utterance in noisy recordings has been investigated in [10]. Then was showed that the algorithm allows detect the boundaries of words in noisy recordings with high accuracy even when the SNR ratio falls down to -30 dB. In this study we analyzed the efficiency of the modified algorithm for fixing the boundaries of phonemes in diphones composed of the nasals (m,n) and six vowels (a,o,e,ee,u,i). The noise has been added by carrying out the analysis with different SNR levels. An expert labeler analyzed diphone recordings and checked the starting and ending positions of the diphone as well as the boundary between a nasal and a vowel. Those labels were considered as the references for the automatic algorithm. In case the boundary detected by the algorithm felt into the range $x \pm 20$ msec of the manually detected boundary, the result was considered as correct. Otherwise an error was fixed. Two types of errors were analyzed: false alarm P_{fa} and false rejection P_{fr}. Table 1 presents the averaged errors per speaker for different SNR values.

Table 1. Average number of false alarm and false rejection errors detecting the boundaries of phonemes in diphones nasal-vowel

SNR, dB	Average number of errors per speaker	
	P_{fa}	P_{fr}
12	1.5	0
6	2.2	0
0	8.3	4.9

It could be concluded that the algorithm based only on audio information reliably detects the boundaries of phonemes in diphones when the noise level is low or moderate. When the noise level substantially increases the algorithm produces relatively great number of errors (8.5 false alarms and 5.4 false rejections per speaker; having in mind that each speaker pronounced 24 diphones). In the presence of a noise the information of lip movements should form a complementary source for improving the detection of acoustic event boundaries.

4 Conclusions

The wheelchair with voice command based control was developed. The wheelchair control interface has embedded capabilities to be expanded to the multimodal interface implementing several modes of human-computer interaction.

The recognition of voice commands is performed using client-service architecture: the recognition engine is implemented at the server side while client serves only as speech recorder and transmitter to the server. Such approach enables to achieve higher recognition accuracy exploiting higher resources of the server side computer and consequently more sophisticated algorithms. This approach enables to expand the human-computer interaction mode with more modalities in the future.

The user's satisfaction is affected by the commands recognition accuracy. The recognition accuracy is affected by many factors. One of affecting factors is the proper detection of the spoken utterance boundaries. The algorithm for the detection of acoustic events boundaries in long and noisy recordings has been proposed. The algorithm enables to detect the boundaries of acoustic events even when SNR level goes down to 3-6 dB. The proper detection of boundaries should lead to the overall increase in the voice commands recognition accuracy.

Acknowledgements. Parts of this work are part of the research project "Lietuviškų balso komandų atpažinimui orientuoto, multimodalinio išmaniųjų įrenginių asociatyvinio valdymo algoritmo sukūrimas ir modeliavimas", No.: 20101216-90 funded by European Union Structural Funds project "Postdoctoral Fellowship Implementation in Lithuania" within the framework of the Measure for Enhancing Mobility of Scholars and Other Researchers and the Promotion of Student Research (VP1-3.1-ŠMM-01) of the Program of Human Resources Development Action Plan.

References

1. Hawley, M., Green, P., Enderby, P., Cunnigham, S., Moore, R.: Speech technology for e-inclusion of people with physical disabilities and disordered speech. In: Proc. of Interspeech 2005, Lisbon, Portugal (2005)
2. Rogoff, B., Goodman Turkanis, C., Bartlett, L.: Learning together: Chikdren and adults in school community. Oxford University Press, New York (2001)
3. Macek, T., Kleindienst, J., Krchal, J., Seredi, L.: Multi-modal telephony services in hometal Intelligent Environments. In: 3rd IET International Conference, pp. 404–410 (2007)

4. Maskeliunas, R., Rudzionis, A., Rudzionis, V.: Advances on the Use of the Foreign Language Recognizer. In: Esposito, A., Campbell, N., Vogel, C., Hussain, A., Nijholt, A. (eds.) Second COST 2102. LNCS, vol. 5967, pp. 217–224. Springer, Heidelberg (2010)
5. Maskeliunas, R., Rudzionis, A., Ratkevicius, K., Rudzionis, V.: Investigation of Foreign Languages Models for Lithuanian Speech Recognition. Electronics and Electrical Engineering – Kaunas: Technologija 3(91), 37–42 (2009)
6. Rabiner, L.R., Sambur, M.R.: An Algorithm For Determining the Endpoints in Isolated Utterances. Bell System Tech. J. 54, 297–315 (1975)
7. Ying, G.S., Mitchell, C.D., Jamieson, L.: Endpoint Detection of Isolated Utterances Based on a Modified Teager Energy Measurement. In: Proc. of ICASSP 1993, pp. 732–735 (1993)
8. Hoyt, J., Wechsler, H.: Detection of Human Speech in Structured Noise. In: Proc. of ICASSP 1994, pp. 237–240 (1994)
9. Scheirer, E., Slaney, M.: Construction of Robust Multifeature Speech / Music Discriminator. In: Proc. of ICASSP 1997, pp. 1331–1334 (1997)
10. Rudzionis, A., Rudzionis, V.: Noisy speech detection and endpointing. In: Proc. of ISCA Workshop “Voice Operated Telecom Services”, Ghent, Belgium, pp. 79–84 (May 2000)

Purchase Intent, Online Offers and Product Innovation: Misunderstandings in the Ménage à Trois

Davor Meersman[1] and Christophe Debruyne[2]

[1] Digital Ecosystems and Business Intelligence Institute,
Curtin University, De Laeter Way, Technology Park 4,
6102 Bentley, Australia
d.meersman@curtin.edu.au
[2] Semantics Technology and Applications Research Lab,
Vrije Universiteit, Brussel Pleinlaan 2, 1050 Brussel, Belgium
chrdebru@vub.ac.be

Abstract. We discuss a semantic platform that matches a customer's purchase intent against vendor offers. The customers' perception on particular products, including evolving needs and preferences, were captured in a request and product ontology, in turn used to annotate vendor offers. During the project, however, we observed an important gap between the intent descriptions of users and the available data in product descriptions. We argue that through the inclusion of peripheral data, vendors are able to innovate according to customer preference, and users receive increasingly relevant results. We present a method that is essentially a customer-driven innovation system using product innovation ontologies.

Keywords: product ontology, product innovation, online commerce, purchase intent, peripheral data.

1 Introduction

When consumers want to buy a certain item on the Web today, they have to browse through literally hundreds of offers and results and this number is expected to increase in the future. In this model, the vendors drive the process by publishing products and providing means to buy these online. For example travel agencies in the Netherlands need to query many different tour operators to find holiday packages meeting their customers' requirements. They often have an API that facilitates this process, but the granularity of the specific search is often limited due to the heterogeneous nature of all vendor databases.

A solution to this problem would be to allow the consumers to specify their requirements and to match these to offers of different vendors. The COMDRIVE RFP project [5] resulted in a platform enabling consumers to drive the requirements process by expressing their intent to buy a certain product in a tool and language they are comfortable with. This platform sends out the request to a distributed vendor infrastructure, which responds to the request with offers.

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 132–143, 2011.

For this solution to be effective, a common vocabulary between the consumers and the vendors has to be established. Such a vocabulary can be captured in an ontology. An ontology is commonly defined as: *a [formal,] explicit specification of a [shared] conceptualization* [8]. Ontologies are necessary to enable semantic interoperability between information systems and services on the Web [9]. In general, interoperability is defined as the ability of two or more information systems or their (computerized) components to exchange data, knowledge or resources and to interpret the information in them [4], in this case the COMDRIVE RFP platform and the different vendor applications.

The content of this paper is organized as follows: Section 2 presents the platform and its different components. Section 3 presents the ontology engineering methodology adopted for this project and provides details on the ontology construction processes as well how these ontologies can be used to annotate data in vendor applications. Section 4 evaluates the ontology with respect to goal of the project and points some of the problems we have encountered during the pilot. Section 5 discusses these problems in more depth, arguing the need for a methodology to create a feedback loop from the customer to the producer, presented in Section 6. Throughout this paper, the examples used stem from the domain of winter holiday packages (including winter sports, accommodation, facilities).

2 The Comdrive RFP Platform

In the COMDRIVE RFP project a semantic platform [5] was developed that consists of the following components:

- **Request and Product Ontology.** A semantic and extensible conceptualization of requests and products. An ontology that describes the domain-specific requests, allowing communities of customers and domain experts to contribute product- and domain specific concepts through collaborative ontology engineering.
- **Automated Group Buying Module** allowing community leaders to organize their group buying activities online with their community members with support for member group buying process initiation.
- **Semantic Matching Engine.** Matching of customer intent and vendor offering based on shared, personal customer purchasing profile and community profile for accurate offerings based on predefined and implicit criteria. Matching of semantic product data and flat vendor data annotated with the ontology.
- **Rapid Semantic Node Cloud Navigation.** Dynamic node cloud underpinned by semantic product data for rapid navigation through correlated product concepts. Combines the ease of use of tag clouds with the ability to process structured data. Fig. 1 contains a screenshot of the user interface.

Fig. 2 shows how the different components interact within the platform. Users are able to express their purchase intent through the portal, which has a dynamic interface using the ontology. When requests are entered, the matching engine interprets the request and annotated vendor databases to find matches, which are sent back to the costumer. The whole platform is driven by the request and product ontology and is semantically underpinned in that sense.

Fig. 1. The interface, driven by the ontology, aids the user in expressing their intent. In this example, the user is asked to give one (or more) possibilities for Ski Area ("Skigebied" in Dutch), Auto-completion relies on accessing the data through the application commitment.

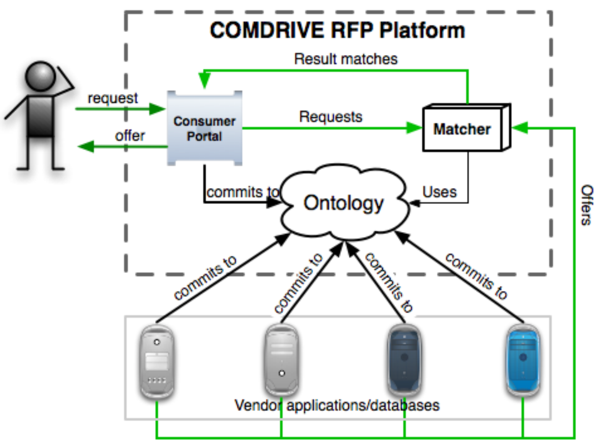


Fig. 2. The interaction between the COMDRIVE RFP Platform's different components

The ontologies do not emerge by themselves, buyers and vendors need to share and reach an agreement on a common vocabulary of the domain. More specifically, software agents need to interpret the information in the customer's purchase intent to automatically match this information with vendor offers based on their semantics. A conceptualization provides a shared agreement on the semantics of core concepts and the relationships between them, imposing a structure on the domain that is readable by both humans and machines.

3 Ontology Engineering Methodology and Related Work

Out of the many collaborative ontology engineering methodologies that exist today [18], we have adopted DOGMA [15], which stands out for its groundings in linguistics. DOGMA relies on the fact that knowledge building blocks, expressed in natural language, are easily obtained and agreed upon (as inspired by database modelling methodologies such as NIAM [21] and ORM [10]), allowing domain experts and knowledge engineers to use natural language to communicate and capture knowledge. The knowledge building blocks - called lexons - in principle only need to express "plausible" facts (as perceived by the community of stakeholders) in order to be entered into the Lexon Base, a repository containing large sets of such lexons. A lexon is formally described as a 5-tuple `<G, head, role, co-role, tail>`, where `G` is an abstract context identifier (e.g., a document on the Web) and used to identify unambiguously (to human users at least) the concepts denoted by the term and role labels. Ontologies in DOGMA are selections of such lexons with constraints on their usage (e.g., "A person can have at most one Name").

It should be clear that DOGMA is only the method to reach agreements amongst different stakeholders. Ontologies in DOGMA are actually "representation agnostic": they can be implemented with other formalisms such as RDF(S) and OWL. DOGMA thus precedes the implementation of ontology, and can be repeated to incrementally grow and refine the ontology. Once agreement has been established and an appropriate mapping of (for instance GoodRelations) has been given, the concepts in the DOGMA ontology can be integrated with that particular schema by generating the necessary classes and properties.

Before building an ontology from scratch, one has to assess existing meta-models that describe products (not necessarily implemented with Semantic Web technologies). [12] We have analysed and compared four important product meta-models: eCl@ss[1], UNSPSC[2], EOTD[3] and RosettaNet Technical Dictionary[4]. Both eCl@ss and UNSPSC are broad: the first was created by and driven by the German industry and is thus a "de facto standard", whereas the United Nations Development Programme drives the development of latter. Both UNSPSC and eCl@ss provide very little detail for the travelling domain. The others were designed for more technical industries and did not fit the scope of this project.

Travel industry meta-models include Hi-Touch[5], OnTour[6], Harmonise [6] and the Open Travel Alliance specification[7]. Harmonise focuses on accommodation and events (e.g., sports and conferences), but its main aim is to transfer data between tourism industry partners. Hi-Touch is a commercial thesaurus implemented in OWL to align different vendor databases. OnTour, a recent initiative, mainly covers accommodation and activities. Open Travel Alliance provides a structure for

---

[1] `http://www.eclass-online.com/`
[2] `http://www.unspsc.org/`
[3] `http://www.eccma.org/`
[4] `http://www.rosettanet.org/`
[5] `http://www.mondeca.com/`
[6] `http://e-tourism.deri.at/ont/index.html`
[7] `http://www.opentravel.org/`

electronic messages, e.g., concerning flights, insurance, etc. Hi-touch and OnTour ontologies were developed based on international standards whereas Open Travel Alliance and Harmonise provide their own.

We bootstrapped the product ontology drawing inspiration from the existing meta-models and vendor applications, which were then refined and completed by several domain experts. We consulted domain experts with different views on the domain, such as tour operators for the vendor perspective and a community of skiers for the buyer perspective. Fig. 3 shows few of the hundreds of lexons created for the purpose of this project.

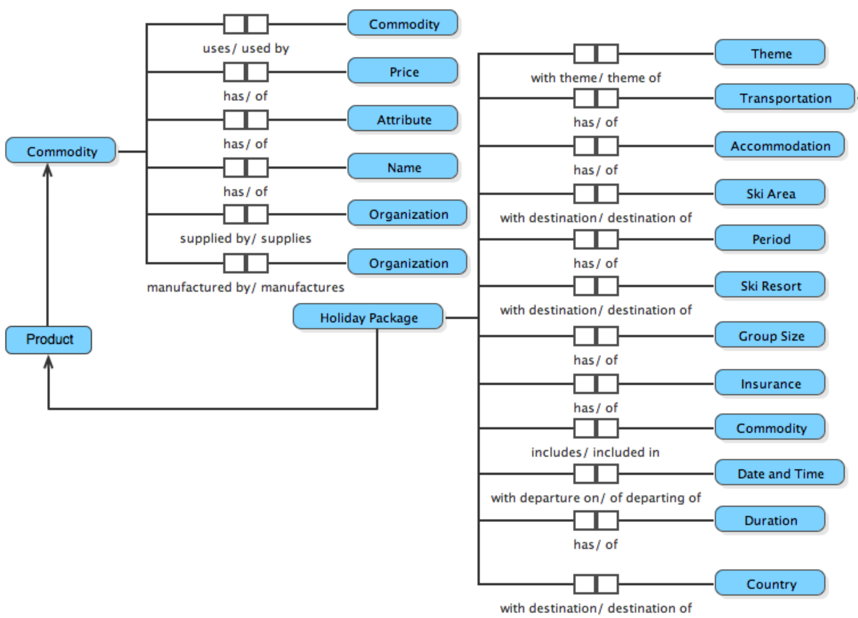


Fig. 3. Some lexons describing domain knowledge developed during the project. These lexons describe holiday packages and commodities, and how holiday packages inherit the properties of commodity by the is-a relationship denoted by the arrow.

The ontology was developed in a modular way. The *Upper Common Ontology* contains the conceptualizations and semantic constraints that are common to and accepted by a general domain, in this case `Product`. For instance, the lexon `<G, Product, with, of, Price>` is true for all applications of stakeholders within that domain and therefore belongs to that layer. The *Domain Application Ontology* contains lexons specific to a certain application domain. In the case of COMDRIVE RFP, these lexons will contain the terms `Holiday Package` and `Accommodation`. The *Lower Common Ontology* represents the interpretation of the domain from the perspective of an organization or community. For instance, the representation of a `Price` might change depending on the community: from a buyer's perspective it is represented by a `Range`, whereas from a Vendor's perspective it is represented as a `Value`. Whilst the ontology evolves, this layer contains the information that is going to be refined by a core domain expert to be integrated in the Upper Common Ontology. The different modules are then connected by matching context-term pairs. In the lexons shown in Fig. 3, the facts around Holiday Package would belong to the DOA (including the fact that a `Holiday Package` is a `Product`) and the facts and the facts around commodities and products to the UCO.

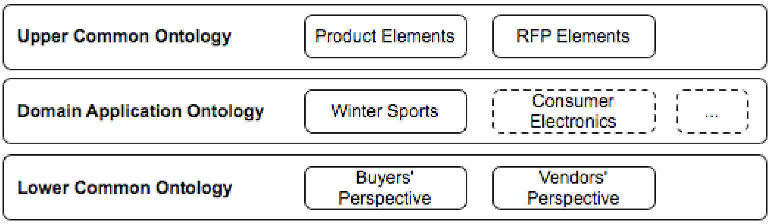


Fig. 4. The modular structure of ontologies modelled with DOGMA within the COMDRIVE architecture

Applications commit to the ontology by annotating the application symbols (e.g., fields in a database or tags in an XML schema) with concepts and relations in the ontology. An application commitment thus represents an explicit interpretation of an ontology for an application or a family of applications. In DOGMA, those annotations are made with Ω-RIDL [20]. It consists of a selection of lexons from the ontology, which are relevant for the application, the constraints to specify how that selection can be used (mandatory and uniqueness constraints, for example) and a set of mappings between the application symbols and the symbols used in the ontology. It also provides some scripting functionalities allowing database programmers to manipulate instances whilst accessing the data. Fig. 5 shows some examples of Ω-RIDL statements.

Constraining lexons

`Holiday Package is identified by Name.`
`Holiday Package has exactly 1 Name.`
`Holiday Package has at most 1 Name.`

Mapping application symbols

`Map "/items/item" on Holiday Package.`
`Map "/items/item/title" on Name of Holiday Package.`
`Map "/items/item/description" on Description of Holiday Package.`

Fig. 5. Example of how lexons within an application commitment can be constrained and mapped onto application symbols

This method was applied to construct the ontology that drives the COMDRIVE RFP platform. The matching engine exploits the application mappings expressed in Ω-RIDL to access the data by generating queries in vendor applications for comparison against the intent. As the interface of the platform is also driven by the concepts and relations described in the ontology, the COMDRIVE RFP platform is completely underpinned by semantics.

4 Evaluation

Milq Media[8] is the publisher of the *wintersporters.nl* platform and agreed to take part as pilot partner in our project. Wintersporters.nl's content is characterized by its actuality and a large amount of community generated information. With an average reach of more than 600.000 visitors per month, it is the largest winter sports platform in The Netherlands. The community of wintersporters.nl, with Milq Media acting as the community leader, agreed to be the pilot partner within this project. The pilot ran from Monday 4th October 2010 until Friday 29th October 2010. It was agreed with Milq Media to let the community test the interface to validate the assumptions and results of the project. Their forum provided feedback that enabled us to solve some of the initial bottlenecks in the pilot (e.g., suggest a starting point when users don't know where to start). 38% of the purchase intents were completed.

We presently focus on the pilot evaluation aspects relating to the ontology. These aspects include: result accuracy, result completeness, and concept coverage. The accuracy of results improved significantly during the pilot as we were able to tweak the parameters of the fuzzy matching engine according to the actual user queries to deliver better results.

Because of the heterogeneous nature of the vendor data streams, data that was present with one vendor, was sometimes not present with another. For example pictures were not present in all offers, which has implications for the uniformity of the results sheet.

The most important issue we have experienced was a serious discrepancy between what concepts and level of granularity the user thinks are important in defining an intent (the “ideal image”) and what is offered by vendors. Multiple accounts on the pilot forum (which had over 100 posts) related to the inability of users to express their intent fully. For example the size of the ski area, the height of the ski area, the presence of après-ski facilities, the calmness and/or cosiness of the environment, the presence of nice restaurants, the ability to buy travel insurance, ski bus distance, the grade of luxuriousness of the hotel or apartment, etc. This kind of data is not present in the vendor data streams, but constitutes *peripheral data* that is important in booking a ski holiday. Peripheral data is data that is used by the customer in the purchase decision process, but that is not part of the offering of the product or service provider. This kind of data directly relates to the function the user want to see performed by the service they book: to be relaxed, to have fun, to be satiated with good foods, to enjoy the scenery, to be safe, etc.

We have seen that merely building ontologies on top of the data that is provided by vendors does not solve the problem of finding products that match user needs. When given the liberty of defining one’s wish, users demonstrate the desire to involve peripheral data that pertains to the product application domain. Although some of these concepts were present in the ontology, there was no data to work with.

Committing to the ontology is not just a question of mapping existing data fields. The commitment will have to entail changes to the internal vendor data structure, for example added granularity or new peripheral concepts. As the business environment changes, vendors need to put on the hat of or work with their producers and innovate

[8] `http://www.milq.nl/`

products and their respective representations so that they better correspond with the purchase intent of users.

When a user searched for a product he or she had an ideal image in mind. Through queries and navigating the node cloud the user approximated his or her ideal image by mapping the offers to what was in mind. Eventually the user made a choice and purchase if the offer mapped satisfactorily to his or her ideal image (notwithstanding eventual ideal image transformations influenced by and during the search process). What was completely lost in this process is the original intent, the ideal image or the imagined product the user had. We just know that he or she purchased something that approximated it. This misunderstanding shown in the triangle in Fig. 6 means that no optimisations can be made in terms of product offering, categorisation, presentation or engineering, other than through exterior (and posterior) processes like customer satisfaction analysis and other market research methods.

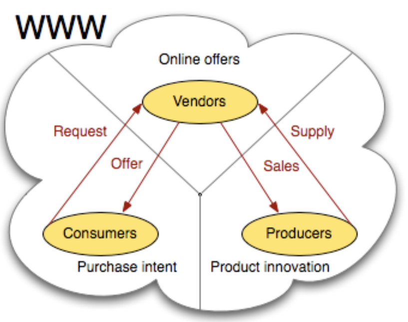


Fig. 6. Producers lack insight in evolving customer preferences and needs on the Web

We would like to argue that for product offers on the Web to improve in terms of matching user needs, the information expressed in purchase intents should correspond to structured data about the product application context. In other words, the entire domain in which the product is used needs to be modelled on the side of the vendor and producer.

In the winter sports pilot, there were several types of peripheral data that were entered by the users: structured data such as ski area size and height, but also concepts such as ‘cosiness’ and ‘luxuriousness’, which are user profile dependent.

Integrating peripheral data in one’s data structure allows producers to innovate directly based on user desires as perceived through intent definitions and vendors to better tailor their offers to user needs. This is the crux of the problem in product search today: producers fail to realize that users performing queries are a vast (potential) source of direct information for innovation processes.

5 Discussion

Information and knowledge are the most important objects in an innovation process. But it is not clear what information and what knowledge, in what form, at what stage, in which quantity, at what level of detail is needed. A major problem is that companies generally do not sufficiently grasp the concept of innovation to answer these questions. Innovation is still largely regarded as the result of a bright idea, or as just an organic improvement on an existing product. Generic innovation process models such as for example the Innovanet model [17] do not deal with the intricacies of knowledge, but treat it as a given or as a result. Clearly there are more dimensions to it.

In order to tackle the issue of knowledge intricacy in innovation processes one must:

1. Study and understand the concept of innovation and information in innovation;
2. Develop a conceptual framework to support this knowledge;
3. Embed the framework in a prototype for adaptation and validation purposes.

The type of information that is most important in an innovation process is how the user perceives a product and what it *will do* for him or her. This information is essential in moving the product, but is also exactly the type of information market input in innovation processes should consist of. Producers think in terms of properties or features, whilst consumers think in terms of functions, or the "job they want to get done" by using the product or service [2].

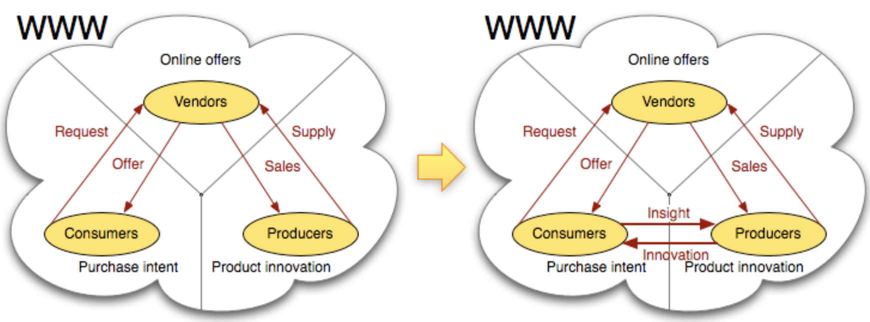


Fig. 7. Producers can use peripheral data to gain consumer insight and drive their product innovation, and (indirectly) supply better information to the customers through the vendors

Whilst the properties of a product are per definition stable and known, the context in which it is used or even how it will be used are usually quite difficult to predict. The notion that Thomas Edison thought of the phonograph as a device that would be used to "record the wishes of old men on their death beds" [7] is only one of many examples that the market frequently puts solutions to use for initially unexpected needs. If an organization wants to keep in touch with all the ways its products are used (and use that information to drive innovation), this information will have to come from the market. Our approach could offer a considerable advancement in bridging the ontology engineering lag. This conceptual dynamics bottleneck is essentially a lack of coverage of concepts by an ontology in a reality that is continually changing. When the concepts in a domain change, there is a "maintenance lag" in the ontology engineering process [11]. If we look at the causes of conceptual dynamics in corporate environments, these will often be changing market conditions.

The main problem that needs to be addressed is that of a semantic product data approach that is suitable for product innovation purposes. The problem of product data integration has been addressed by the Semantic Web community by developing product ontologies. Apart from the product classification standards discussed in Section 3, the most important product ontologies today are eClassOWL, PRONTO, SWOP and GoodRelations.

- eClassOWL is a product ontology that is derived from the eCl@ss product classification standard that is widely used in the manufacturing industry [11].
- PRONTO is geared towards product information regarding production, storage, sales and distribution [19].
- SWOP, a product ontology developed by the Semantic Web Open engineering Platform project, is an extremely powerful and very granular product modelling ontology for expressing product and component properties [1].
- GoodRelations is a popular lightweight product ontology for representing product information that is relevant to the e-commerce domain [14].

The above ontologies do a good job of representing product, component and related properties. Properties such as product shape, material, colour, family and other manufacturing and sales related concepts are covered, in some ontologies to very great detail. However, in the context of innovation the main shortcoming of these ontologies is that (product, component and property) functions and the application context of the products are not included. Hence they are to be regarded more as very important and valuable assets in the modelling, manufacturing and commerce process, rather than as tools to drive innovation. We argue that for ontologies to be valuable in innovation processes, we need to rework the entire idea of product ontology.

6 Future Work

Based on previous related research [16] in the leisure product domain we will create a product innovation ontology complementing the request and product ontology to include functions and more of the application domain. A major conceptual component of the extended ontology is the integration of product and service functions. Functions answer the question of why something is there. Because the existence of properties is driven by the function they perform, i.e. the fulfilment of initial requirements and in some cases posterior cost considerations, the inclusion of functions and their linkage to product and/or service properties is crucial in the context of ideation.

Another big difference between a regular product ontology and an ontology geared towards product innovation, is that the entire domain of product application is included in the product innovation ontology, as the usage of the product has a great influence on the reasoning behind feature introduction.

The product innovation ontology features a number of concepts for modelling a domain that have been validated in various industry contexts [4]. These concepts are:

- *Actor*: any actor using objects in any process in the domain. The actor is a superset of all possible user profiles. The profiles emerge from the preferences entered by the users. For example younger people tend to put more emphasis on budget and après-ski.
- *Object*: products and components used in processes by actors and objects.
- *Process*: any process in the domain, executed by an actor with the use of objects. The context of a particular process extends different parts of the ontology by introducing additional facts and constraints that support this process. Cfr. Lower Common Ontology in Section 3.
- *Quality*: concepts that define how, to what extent, when etc. something happens; properties and functions of objects, actors and processes. The quality concept also contains concepts pertaining to more intangible customer knowledge such as satisfaction and opinion, characterized by different levels of complexity [3].

For the travel domain we will add more peripheral concepts to the ontology and ensure there is data present for the matching engine to work with. Once the data is there, the effect will be twofold: users will be able to get offers tailored to their intent, and vendors will get insight into what the customer wants.

7 Conclusions

In the COMDRIVE RFP project, the initial outset was to improve product search on the Web by capturing the purchase intent of customers and match that to vendor offers using a request and product ontology and fuzzy matching engine. Although a success in terms of the ontology, we discovered an important gap between the intent descriptions of users and the available data in product descriptions. Users demonstrate the need to use *peripheral data* to describe their purchase intents. This has urged us to revisit the entire idea of product ontology and involve not only the customer and vendor side, but also and importantly the producer side of the equation, to establish a common understanding in the ménage à trois between these economic partners. Through capturing purchase intent including peripheral data (functions and application context), vendors are able to grasp evolving needs and preferences and innovate accordingly. The result is a method that is essentially a customer-driven innovation system that satisfies users with (increasingly) relevant results, and offers vendors a cost-effective way of gathering accurate insight into customer needs and preferences.

References

1. Böhms, M.: The SWOP Semantic Product Modelling Approach. In: SWOP project deliverable 23. SWOP (2006), `http://www.swop-project.eu/swop-solutions/semantic-product-modelling-approach/SWOP_D23_WP2_T2300_TNO_2008-04-15_v12.pdf` (accessed on June 14, 2010)
2. Christensen, C.M., Anthony, S.D., Berstell, G., Nitterhouse, D.: Finding the Right Job For Your Product. MIT Sloan Management Review 3, 38–47 (2007)
3. Cillo, P.: Market learning capability and innovation. An explorative empirical study in the fashion industry. In: Proceedings of the 3rd International Congress Marketing Trends, Venice (2003)

4. De Leenheer, P.: On Community-based Ontology Evolution. PhD Thesis, Vrije Universiteit Brussel (2008)
5. Debruyne, C., Meersman, D., Baert, M., Hansenne, R.: Community Driven Requests for Proposals: Applying Semantics to match a Customer's Intent to Annotated Vendor Data. In: Proceedings of the 7th International Conference on Web Information Systems and Technologies (to appear)
6. Dell'Erba, M., Fodor, O., Ricci, F., Werthner, H.H.: A solution for data interoperability. In: Monteiro, J.L., Swatman, P.M.C., Valadares Tavares, L. (eds.) I3E. IFIP Conference Proceedings, vol. 233. Kluwer (2002)
7. McDonald, S., Lamberton, D., Mandeville, T. (eds.): The Trouble with Technology. Explorations in the Process of Technological Change. Frances Printer Publishers Limited, London (1983)
8. Gruber, T.R.: Toward principles of the design of ontologies used for knowledge sharing. International Journal of Human and Computer (1995)
9. Guarino, N.: Formal ontology and information systems. In: Int. Conference On Formal Ontology In Information Systems, FOIS 1998, Trento (1998)
10. Halpin, T.: Information Modeling and Relational Databases. Morgan Kaufmann, San Francisco (2008)
11. Hepp, M.: The True Complexity of Product Representation in the Semantic Web. In: Proceedings of the 14th European Conference on Information System (ECIS 2006), Gothenburg, Sweden, pp. 1–12 (2006)
12. Hepp, M., Leukel, J., Schmitz, V.: A quantitative analysis of product categorization standards: content, coverage, and maintenance of eCL@SS, UNSPSC, EOTD, and the RosettaNet technical dictionary. Knowledge Information Systems (2007)
13. Hepp, M.: Possible Ontologies: How Reality Constrains the Development of Relevant Ontologies. IEEE Internet Computing 1, 90–96 (2007)
14. Hepp, M.: GoodRelations: An Ontology for Describing Products and Services Offers on the Web. In: Proceedings of the 16th International Conference on Knowledge Engineering and Knowledge Management, EKAW, pp. 332–347 (2008)
15. Jarrar, M., Meersman, R.: Ontology engineering – the DOGMA approach. In: Dillon, T.S., Chang, E., Chung, S., Sycara, K. (eds.) Advances in Web Semantics I. LNCS, vol. 4891. Springer, Heidelberg (2008)
16. Meersman, D., Dillon, T.S.: The Open Innovation Paradigm and the Semantic Web: An Ontology for Distributed Product Innovation. In: Proceedings of OTM Workshops 2010 (2010)
17. Paukert, M., Niederée, C., Muscogiuri, C., Bouquet, P., Hemmje, M.: Knowledge in the Innovation Process: An Empirical Study for Validating the Innovation Knowledge Life Cycle. In: Proceedings of the 4th European Conference on Knowledge Management, ECKM, pp. 725–737 (2003)
18. Sure, Y., Staab, S., Studer, R.: Ontology engineering methodology. In: Bernus, P., Blazewicz, J., Schmidt, G.J., Shaw, M.J., Staab, S., Studer, R. (eds.) Handbook on Ontologies, International Handbooks on Information Systems (2009)
19. Vegetti, M., Henning, G.P., Leone, H.P.: Product ontology: definition of an ontology for the complex product modelling domain. In: Proceedings of the Mercosur Congress on Process Systems Engineering (2005), `http://www.enpromer2005.eq.ufrj.br/lng/en/index.php` (accessed on: May 26, 2010)
20. Verheyden, P., De Bo, J., Meersman, R.: Semantically Unlocking Database Content Through Ontology-Based Mediation. In: Bussler, C.J., Tannen, V., Fundulaki, I. (eds.) SWDB 2004. LNCS, vol. 3372, pp. 109–126. Springer, Heidelberg (2005)
21. Wintraecken, J.: The NIAM Information Analysis Method, Theory and Practice. Kluwer Academic Publishers (1990)

A Concept of the Knowledge Strategic Resource Network (SknowNet) for SMEs

Justyna Patalas-Maliszewska[1,2] and Irene Krebs[3]

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

Abstract. This paper presents concept of the strategic knowledge resource network for SMEs. A strategic-knowledge resource in a company represents the knowledge, skills and capabilities of the individuals who make up the workforce of that company. In the paper we present a network for the transfer as well as storage of knowledge in enterprises in order to make sure every post is covered by an employee (seen as a strategic knowledge resource) who efficiently completes the respective tasks of an organisation. This is especially important in a network of SMEs easing the exchange and transfer of employees. The structure of the SknowNet consists of (1) sets of business processes in SME, (2) sets of values of personnel usefulness function (strategic knowledge resource), (3) a games theory approach, especially natural games, is used, which facilitates decision taking in hesitant situations.

Keywords: strategic knowledge resource, personnel usefulness function (value of a strategic knowledge resource.

1 Introduction

Small and medium sized enterprises have a decisive role in creating working places and, more generally, they are social stability and economic development factors. However, they often encounter difficulties in getting capital or credit, given the limited guarantees which they can offer, as well as limited access to information, concerning especially new technologies and possible markets. Companies, which have access to external resources for their development, may gain an competitive advantage in the new era of networked enterprises and flexible relationships, based on the Internet [Chesbrough, 2003].

Strategic knowledge management is understood as a relation between the overall organisation's strategy and human resources management strategy. Human resources are assumed as strategic knowledge resources and are defined as unique enterprise's potential in the form of knowledge and experience [Barney J., 1995]. In the process of knowledge resources management in a company on a strategic level, it is necessary to use appropriate methods and tools which support decision making. This is especially

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 144–153, 2011.

important in a network of SMEs easing the exchange and transfer of employees. In such a view a network of SMEs can also be seen as a means for transferring and storing knowledge in order to make sure that every position is covered by an employee (human resource) who efficiently completes tasks of the organisation.

In this context we aim to measure the value of strategic knowledge resource using a so-called personnel SME usefulness function W_{nm} (for the m-th employee in the n-the SME). It is defined based on following factors: GK - general knowledge of the m-th employee, PK - professional knowledge of the m-th employee, A - professional abilities of the m-th employee, E - experience of the m-th employee, P - patents of the m-th employee, C - clients of the m-th employee, P - personality of the m-th employee; $n, m \in N$ [Patalas-Maliszewska J., 2009].

The process of selecting the right persons is an important and strategic decision for SME that may determine the further development of the enterprise; thus, this decision has to be taken "correctly". A relevant framework to this issue is based on a database containing the values of the above defined usefulness function W_{nm} for each m-th employee. Given such a database one could this could feed an strategic knowledge resource allocation efficiency model for small and medium sized enterprises. The paper describes such an approach.

The structure of the paper is as follows: The next chapter presents the description of allocation of human resources in an organisation and the methods assessing the efficiency of decisions on acquiring knowledge, known from literature. The third chapter presents the author's network of strategic knowledge resource (SknowNet) for SMEs, consisting of a database of values of knowledge resource in SMEs. This is based on (1) a reference SME model, (2) employees (description of workplaces - business processes), (3) value of personnel usefulness function, and (4) the games theory (natural games) which facilitate taking a decision in hesitant situations. Using a real world case study we show how using this approach can be used in context of an SMEs network to maximize the knowledge value for an SME. Finally, the summary presents directions of further works.

2 Background and Related Work

Nowadays enterprises perceive knowledge as a strategic resource which may mark the competitive dominance of an enterprise. We may assume that the superior aim to manage an enterprise is a skilful use of strategic knowledge resource.

Small and medium enterprises in an especially flexible way adjust to the market requirements and clients' needs by changing activity's profile, products (services) assortment and by forming work time and activities' forms. The connection of resource of many co-operating enterprises makes it possible to concentrate on key skills (competences) of the company [Patalas-Maliszewska J., Krebs I., 2010].

Also, or especially, small and medium – sized enterprises have the necessity to planning and reporting about their capital value in the face of market globalization, strong and intensified competition. In this context they have to choose the right and the most appropriate method (or tool) for knowledge (or employee) allocation. However, evaluating the value of knowledge is complicated – and nearly impossible – task.

As regards research, there are methods of intellectual capital assessing based on investment of staff's development or transfer of knowledge of employees. However, there are no generally agreed methods assessing the efficiency of decisions on acquiring knowledge. The process of managing intellectual capital should consist of two stages: identifying and measuring. Literature distinguishes qualitative measures (e.g. Danish project of IC measurement, 'Scandia' navigator, intengible assets monitor, IC model –TM Rating, VCSTM, balanced result sheet, report by Saratoga Institute) and methods of valuating intellectual capital (e.g. MV/ MB, q-Tobin, CIV, KCE, VAICTM, economic added value, IAV model, Strassmann's method, IAMVTM, technology broker); see [Dudycz T., 2005], [Edvinsson L. and Malone S., 2001], [Fitz-enz J., 2001], [Kasiewicz S. et al., 2006], [Mikuła B., 2002]. Attempts are continuously made to find methods for knowledge resource allocation increasing the efficiency for the small and medium enterprises. Allocation of human resources in an organisation is a process of shaping employment by means of different personnel action, which include, e.g., winning employees, their internal and external selection as well as employees' resignations and dismissals. This is done to make sure that every post is covered by an employee who efficiently completes the respective tasks of an organisation [Pawlak, 2003]. Skilled personnel and versatile enterprise are key factors in the knowledge competition [Patalas-Maliszewska J., Krebs I, Stryjski R., 2010].

However, it is difficult to find a method allowing keeping all mentioned principles in the selection of employees. The recently used method of employees selection is personnel marketing – reliable informing candidates about a company, using clear principles and criteria in employees selection, which ensures subjective and identical treatment of all candidates [Król H. and Ludwiczyński A., 2007]. However, there is no method found, which enables keeping objectivity and uniformity principles at the employees selection.

Since there is no overall agreed method to allocate human resources in an organisation and for assessing the efficiency of decisions on acquiring knowledge, in the following.

We present our approach to choose an outside employee for an enterprise (SknowNet for SME) at the condition of maximalisation of the knowledge value for SME; it consists of elements: (1) a database of strategic values of knowledge resource in SMEs, (2) the games theory (natural games) which facilitate taking a decision in hesitant situations.

3 SknowNet for SME

For the SME to meet different demands and stand up to the international and local market competition it is necessary to proceed with the oncoming tasks in an efficient and reasonable way, which can be achieved by the systematic analysis of the undertaken tasks in relation to the company organizational and economical ability of their completion. In economy practice making a decision in enterprise is conditioned by competitors' action, changing factors of environments, eg. technical progress and results of the research works [Haas-Edersheim, 2007]. Added value for SME can be determine as knowledge, employees' skills and abilities, social relation, know-how, and particularly effective investing in intellectual capital. The enterprises which invest in human capital and systems of work are achieved competitive advantage because of theirs workers' readiness to learning and qualifying themselves and also thanks to effective information and communication transfers.

3.1 The Database of Values of Strategic Knowledge Resources in SMEs

The personnel function in enterprises encompasses all the issues concerning employees, e.g. recruiting workers, managing them, their professional development etc. Forming a personnel function is a simplified picture of a given part of reality, in which features, relations and other unimportant elements for a given period are eliminated [Król H., Ludwiczyńsk A., 2007]. Authors attempt a trial of defining the SME database of values of a personnel usefulness function W_{nm} (the value of strategic knowledge resources for m-th employee in the n-the SME) in the aspect of formulating a strategy of knowledge management.

In this context we define the personnel SME usefulness function, W_{nm} for the m-th employee in the n-th SME [Patalas-Maliszewska, 2009]:

W_{nm} = f(GK, PK, A, E, P, C, P), where n, m$\in$N, and the following parameters are received as the result of tests for employee, which was evaluated within the range from 1 to 5, where 1 is a bad and 5 is a very good level and:

- GK - General knowledge of the m-th employee.
- PK - Professional knowledge of the m-th employee.
- A - Professional abilities of the m-th employee.
- E - Experience of the m-th employee.
- P - Patents of the m-th employee.
- C - Clients of the m-th employee.
- P - Personality of the m-th employee.

The linear form of this function W_{nm} is chosen because all elements are independent and equally important to assess the effectiveness and efficiency of investment in knowledge.

It is possible to receive indispensable data for account of value personnel SME usefulness function for the m-th employee in the companies belonged to reference model of SME (see Fig. 1). The survey has done by interview in 10 companies (in 10 selected SMEs) in line with the reference model. Each employee completed the questionnaire. On the basis of an algorithm for testing solutions for each employee it was possible to receive a specific value of the personnel usefulness function and each of the parameters of this function. Based on the result research in SMEs (the research group consisted of 10 companies, conformed to concrete model of enterprise – see Fig. 1) the values of personnel usefulness function for the two employees (m=2), who can realize the defined business processes in SME, in the each SMEs of 10 (Table 1) are created.

The structure of database is given (see Fig. 1). Please note that this reference model an SME is seen as a set of processes, and modeled accordingly.

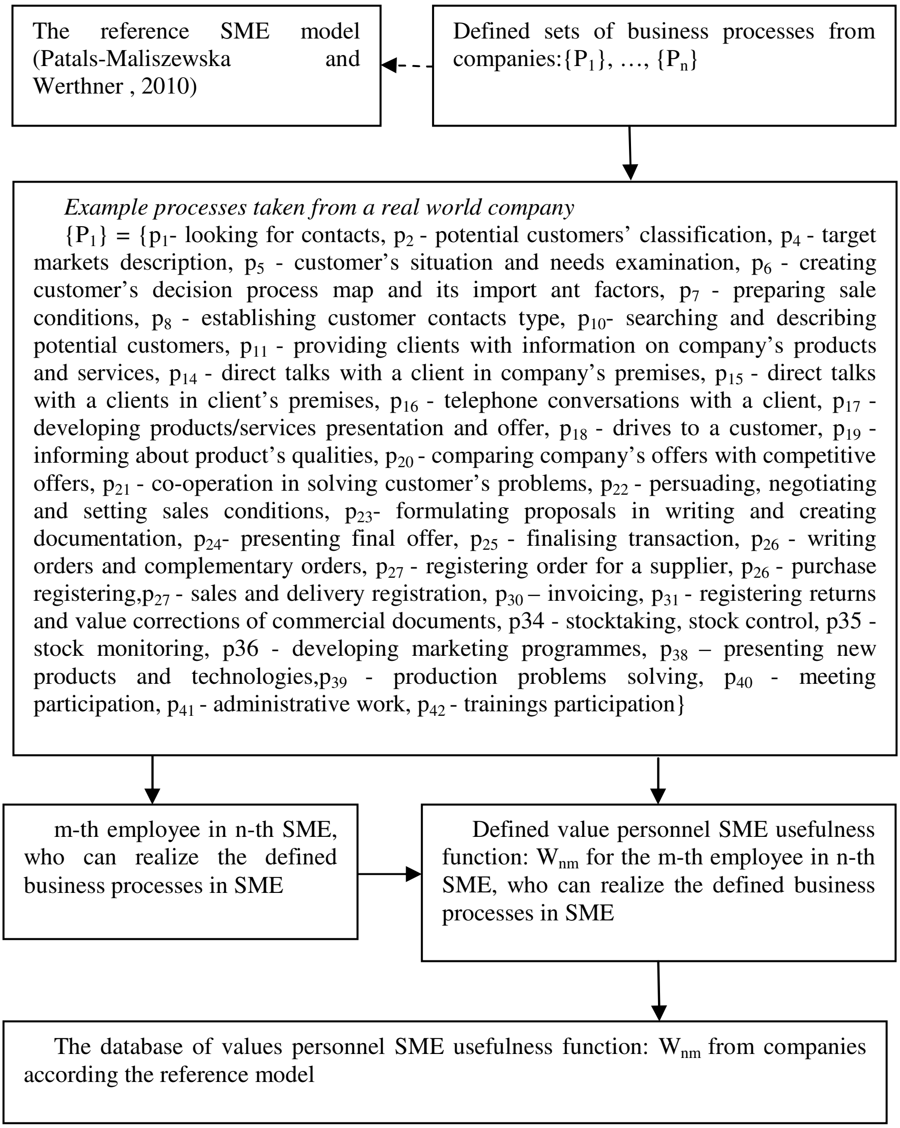


Fig. 1. The structure of the database of strategic values of knowledge resources in SMEs

Table 1. The database of values of strategic knowledge resources in SMEs, (Patalas-Maliszewska and Werthner , 2010)

Companies	Values of personnel SME usefulness function for first employee, who can realize the defined business processes in SME (see Fig. 1)	Values of personnel SME usefulness function for second employee, who can realize the defined business processes in SME (see Fig. 1)
SME_1	$W_{11} = 25$	$W_{12} = 4$
SME_2	$W_{21/=} 19$	$W_{22} = 13$
SME_3	$W_{31} = 21$	$W_{32} = 15$
SME_4	$W_{44} = 15$	$W_{42} = 12$
SME_5	$W_{51} = 12$	$W_{52} = 17$
SME_6	$W_{61} = 17$	$W_{62} = 9$
SME_7	$W_{71} = 21$	$W_{72} = 13$
SME_8	$W_{81} = 21$	$W_{82} = 18$
SME_9	$W_{91} = 15$	$W_{92} = 12$
SME_{10}	$W_{10\,1} = 23$	$W_{10\,2} = 19$

Figure 2 represents the structure, concepts and sequence of actions of SknowNet.

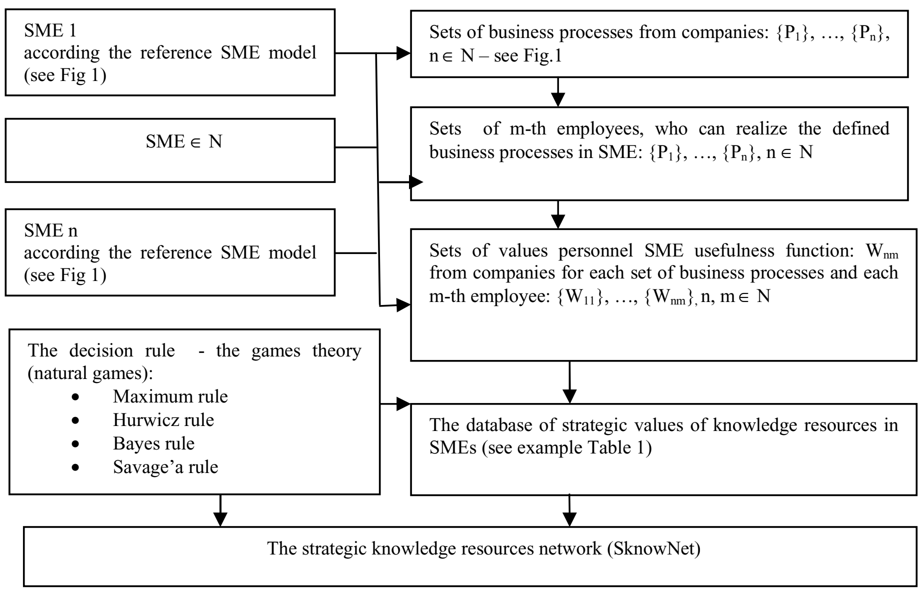


Fig. 2. The structure of SknowNet

The games theory (natural games) creates a model class, which the task facilitate taking a decision in hesitant situations [Luce R. D. and Raiffa H, 1964]. Using such a decision rule would enable SME choosing of strategic knowledge resource ascribed to company's needs. They view game theory as providing an explanatory account of strategic reasoning. It is possible to transfer or choice of the proper knowledge resource (a right employee) for SME from the database of strategic values of knowledge resources in SMEs (see Table 1) based on the rules:

- **Maximum rule:** $v = \max_{i} \{ \min_{j} a_{ij} \}$, where a_{ij} – element of a matrix (element of the database of strategic values of knowledge resources in SMEs - see example Table 1)
- **Hurwicz rule:** $v = \lambda \min_{j} (a_{ij}) + (1-\lambda) \max_{j} (a_{ij})$, where $\lambda (0 \leq \lambda \leq 1)$, where a_{ij} – element of a matrix (element of the database of strategic values of knowledge resources in SMEs - see example Table 1)
- **Bayes rule:** $v = \frac{1}{n} \sum_{j=1}^{n} a_{ij}$, where a_{ij} – element of a matrix (element of the database of strategic values of knowledge resources in SMEs - see example Table 1)
- **Savage'a rule:** $v = \min_{i} \{ \max_{j} \alpha_{ij} \}$, where $\alpha_{ij} = \max_{i} a_{ij} - a_{ij}$, where a_{ij} – element of a matrix (element of the database of strategic values of knowledge resources in SMEs - see example Table 1)

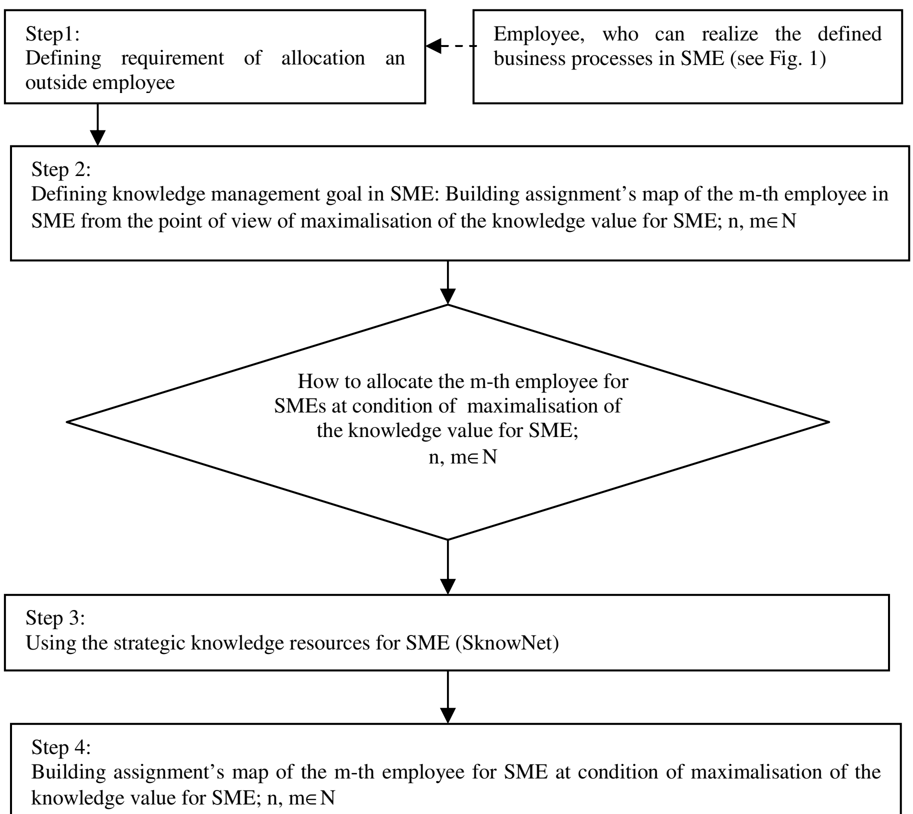


Fig. 3. The procedure of supporting a decision maker about allocation of human resources in an organization

So, the company A, that is described following reference SME model (see Fig. 1), can receive (transfer) the knowledge resource described to its needs and budget from SKnowNet using the decision rules. The procedure of supporting a decision maker about allocating the knowledge resources (i.e., employee) in an organisation is proposed in Fig. 3.

3.2 Case Study

In order to illustrate the procedure supporting a decision maker (see Fig. 2) let me consider the SME that is about to make decision concerning the allocation of an outside employee.

Consider the company A (the object), that is described following reference SME model (see Fig. 1), that is about to make decision concerning the allocation of an outside employee in the sale area. The main areas of the company correspond to following functions supporting: the sale, the supply, orders scheduling, the service, the accounting, human resources management, export/import transactions. The firm employs 16 employees; in the sales area are 5 employees. All four employees in the sales area, without the manager, can realize the same business processes.

The allocation of an outside employee should be done in such a way that the company will receive (transfer) the human resource described to its needs and budget. The finally decision is connected with the allocation of an outside employee (strategic knowledge resource). The presented SME network (see Fig. 3) enables us to make the decision about chosen the strategic knowledge resource ascribed to the company's needs. In order to illustrate this problem let us consider a particular step of the procedure supporting a decision maker.

1. Defining the company A: the sets of business processes (see Fig. 1)
2. Defining requirement of transfer an outside employee:
   (employee, who can realize the defined business processes in SME – see Fig. 1)
3. Using the SknowNet (see Fig. 2) to select the right persons for company A and aimed at maximalisation of the knowledge value for company A (value personnel SME usefulness function): W_{Am}, for the m-th employee.

So, we have 10 SMEs, that are described following reference SME model (see Fig. 1) and the database of strategic values of knowledge resources in SMEs (see example Table 1). The values of personnel SME usefulness function (see Table 1) are treated as hesitant situations in the natural games.

So, we can make decision about transfer or chosen external strategic knowledge resource (company A can receive the employee described to its needs and budget) based on result of decision rule:

- Maximum rule: v = 19 (SME 10)
- Hurwicz rule: for $\lambda = 0{,}8$, v = 19,8 (SME 10)
- Bayes rule: v = 21 (SME 10)
- Savage'a rule: v = 6 (the smallest loss for SME 10)

Company A should make decision about chosen external strategic knowledge resource from SknowNet:

- **based on the maximum rule:** transfer an employee from SME10 to company A will provide at least knowledge value for the new employee in the company A: $W_{A1} = 19$.
- **based on the Hurwicz rule:** transfer an employee from SME10 will provide at average knowledge value for the new employee in the company A: $W_{A1} = 19{,}8$.
- **based on the Bayes rule:** transfer an employee from SME10 will provide at average knowledge value for the new employee in the company A: $W_{A1} = 21$.
- **based on the Savage'a rule:** transfer an employee from SME10 will provide at the smallest loss average knowledge value for the new employee in the company A: $W_{A1} = 6$ relatively to highest knowledge resource value in company.

4 Concluding Remarks

The procedure which has been introduced in this paper gives more possibilities in the area of allocating external knowledge resources. A decision model of employee's assignment to a given task is suggested, from a point of view of maximization of the knowledge value in a company.

A universal character of the suggested SME network is worth emphasizing. Depending on defining business processes in the enterprise and on precise verification of constituent elements of SknowNet, this allows to receive more detailed planning results as basis for the strategy of knowledge resources management.

The research in progress is focused on the development of the SME database of strategic values of knowledge resources in SMEs, which would be similar in terms of business processes as well as defined the reference SME model.

Acknowledgement. The research leading to these results has received funding from the European Community's, Seventh Framework Programme [FP7/2007-2013; FP7-PEOPLE-IEF-2008], under grant agreement n°[235585] "SKnowInnov".

References

1. Byrne, J.A., Brandt, R.: The Virtual Corporation, Business Week, 02 (1993)
2. Chesbrough, H.: Open Innovation: the new imperative for creating and profiting from technology. Harvard Business School Press (2003)
3. Dudycz, T.: Enterprises value management, PTE, Warsaw, Poland (2005) (in polish)
4. Edvinsson, L., Malone, M.: Intellectual Capital: realizing your company's true value by finding its hidden brainpower. HarperBusiness, New York (1997)
5. Fitz-enz, J.: Investment in human capital, Krakow, OE, Poland (2001) (in polish)
6. Kasiewicz, S., Rogowski, W., Kicińska, M.: Intellectual capital, Krakow, OE, Poland (2006) (in polish)
7. Król, H., Ludwiczyński, A.: Human Resource Management, PWN, Warszawa Poland (2007) (in polish)
8. Luce, R.D., Raiffa, H.: Games and decisions, PWN, Warsaw (1964) (in polish)

9. Mikuła, B.(i in.): Enterprises management in XXI. Concepts and methods, Difin, Warsaw, Poland (2002) (in polish)
10. Patalas-Maliszewska, J.: Virtual net of innovative SME. e-mentor (3), pp. 59–67 (2009)
11. Patalas-Maliszewska, J., Krebs, I.: Model of innovation transfer is small and medium enterprises (SME). In: Advances in Production Engineering APE 2010: Proceedings of the 5th International Conference. Warsaw University of Technology, Faculty of Production Engineering (2010)
12. Patalas-Maliszewska, J., Krebs, I., Stryjski, R.: Grenzüberschreitendes, innovatives Lehrmodell in den Bereichen ”eBusiness” und Produktionsmanagement. trafo Wissenschaftsverlag, Berlin (2010)
13. Patalas-Maliszewska, J., Werthner, H.: The method for assessing and forecasting value of knowledge in SMEs - research results. Współczesna Ekonomia (2010)
14. Pawlak, Z.: Company’s personnel function – processes, PWN, Warsaw (2003) (in polish)

Advanced Resource Selection for Federated Enterprise Search

Matthias Wauer, Daniel Schuster, and Alexander Schill

TU Dresden, Faculty of Computer Science, Chair for Computer Networks,
Helmholtzstraße 10, 01062 Dresden, Germany
{matthias.wauer,daniel.schuster,alexander.schill}@tu-dresden.de

Abstract. Distributed information retrieval is a well-known approach for accessing heterogeneous, highly autonomous sources of unstructured information. Selecting and querying only a number of relevant sources can help improve its performance, but most resource selection algorithms are limited to syntactic comparisons.

We present a framework for applying resource selection in the context of a semantic federated product information system, and evaluate the performance of the well-known CORI resource selection algorithm in this context.

Keywords: Resource selection, distributed information retrieval, federated search, enterprise search, enterprise information systems.

1 Introduction

Product information of companies are stored in many different sources, typically due to organizational and technical requirements. This diversification prevents a comprehensive view and seamless access to this information. Furthermore, most of the relevant product information is only available in unstructured form.

We are therefore developing a federated product information system (FPIS) [1] with regards to innovative application scenarios of five collaborating industry partners (ABB, BMW, Deutsche Post DHL, Otto, SAP). It connects federated heterogeneous information providers using semantic middleware. Structured information can already be integrated by semantically mapping their respective schemas to an ontology using existing tools. One of our goals is to extract information from documents as needed in order to associate them with an ontology.

The sheer number of documents produced in companies today can barely be managed in a central document database. Distributed management of documents allows for greater autonomy of the individual *collections* but requires a more sophisticated approach to search. Centrally indexing these documents can only be done if all collections are collaborative, i.e., if they provide immediate access to all of their documents. An alternative is distributed information retrieval [2], which issues a query to the search engines of each collection and merges all results. It can be applied to all collections that provide a search interface.

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 154–159, 2011.

Drawbacks of distributed information retrieval are the potential processing and communication overhead for a large number of collections and increased response times if the response of a collection is delayed. Hence, a resource selection algorithm is supposed to reduce the number of queried collections, typically to those that are estimated to be relevant with regards to the query.

We evaluate the applicability of an existing well-known resource selection algorithm for an FPIS on a corpus of industrial service documents, and propose a framework which utilizes the available semantic information for improving resource selection performance.

2 Related Work

With regards to resource selection, CORI [3] is one of the most popular algorithms. It uses *per collection* statistical features to estimate the relevance of collections, based on inference network document ranking. Queries are expected to be a simple set of terms.

The actual computation estimates two components for each term: a term-based measure $T_{i,t}$ which uprates a term that occurs frequently in collection i w.r.t. average and collection-specific number of different terms, and a collection-evaluative measure which increases the impact of highly distinctive terms, e.g., terms that only occur in few collections. Each term in query $Q = \{t_1, t_2, \ldots, t_n\}$ is weighted equally.

Some drawbacks of CORI have been identified [4] and addressed by other approaches. ReDDE [5] is less prone to disregard large collections if the collections are skewed, i.e., the collections vary considerably in size. For similarly sized collections, results improve marginally. CRCS [6] and SUSHI [7] find that these algorithms barely use the document samples of each collection and their scores for each query, although they are valuable for assessing a collection's relevance. They also determine how many collections should be selected, whereas CORI usually selects a fixed amount of them.

Collections are typically assumed to be independent, so relationships between them are typically not taken into account by these algorithms. Hong et al. [8] present a model that classifies resources not only on singular features for each resource, but also on joint similarity between resources as an additional feature. They estimate the importance of detecting the similarity by applying different algorithms, and conclude that a similarity metric based on relevance for each query performs better than a language model based Kullback-Leibler metric, which performs worse than the common independent approach. The differences are fairly small with TREC testbeds. However, the performance increases significantly for a real-world testbed, in particular for high precision values, i.e., the topmost source ranking results.

Arguello and colleagues [9] extend the document-based selection with both an estimated query topic and query click-through data. These three evidences, namely corpus-based, query-categorical, and click-through features, are combined using a machine learning algorithm, which is initialized with automatically generated training data. Evaluation of this approach shows that the categorization of a query can improve the accuracy significantly if the collection sample is small.

3 Resource Selection Concept

The resource selection is part of our current FPIS, the Aletheia prototype [1]. Similar to Arguello et al. [9], the proposed solution should be able to combine several features for the final collection relevance assessment, but in a much more extensible way as shown in Figure 1. The processing of such features is wrapped as *plugin* components that can be applied flexibly depending on the actual scenario. Connector components are the actual mediators communicating with the federated information providers.

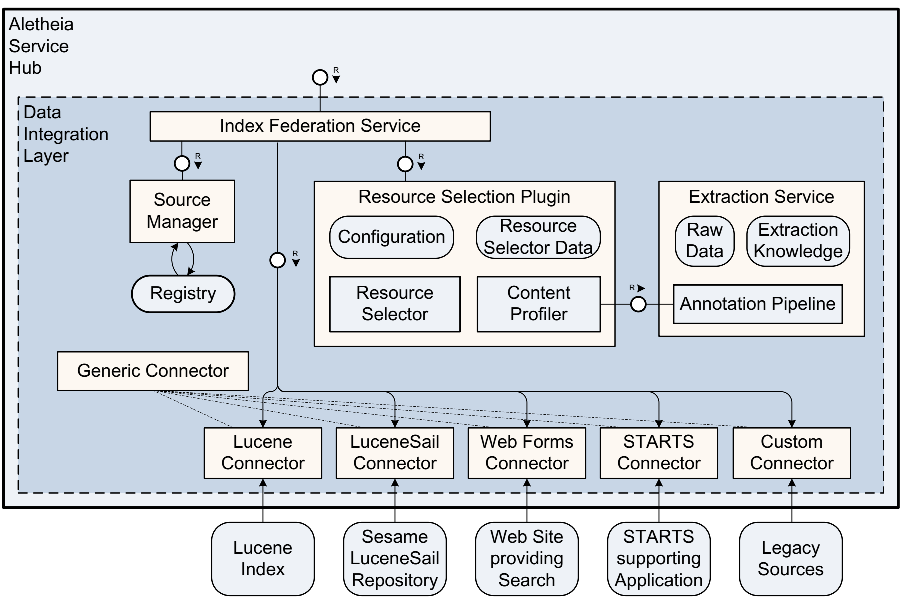


Fig. 1. Architecture of a resource selection framework, in FMC notation

Knowledge based resource selection can be applied by using the Extraction Service provided by the FPIS, which annotates the sampled documents semantically. A semantic resource selection plugin may adapt this component by applying custom UIMA [10] annotators.

The integration of this federated query processing with other components of the Aletheia Service Hub (not shown here) enables many other features, e.g., adding and modifying semantic tags for documents by the user.

4 Preliminary Evaluation

In order to find out how existing algorithms perform, a CORI resource selection plugin indexed different test sets before evaluating a range of queries.

4.1 Test Sets

The evaluation was executed on multiple test sets in order to find out how the framework and algorithms perform. All test sets were derived from a collection of real industrial documents, a subset of a project partner's digital library compiled for offline use. This library consists of

- A topical structure T (tree.xml) containing links to
- Node files (*nodeId*.xml) describing a set of documents D related to the node's topic, and
- 3.624 folders, each containing one of the documents d in D, with a total of 2.89GB of files.

This library is analysed and split into appropriately sized collections, assuming that the content of a sub-tree's referenced documents are related to a limited set of topics as in, e.g., files of a certain workgroup. can As a first attempt, the sum s_{node} of the number of documents in its own node file and all sub-node's files is appended to each node in T. Then, an XPath [11] expression can be applied to find all nodes having a defined minimum and maximum collection size. This approach, however, does not result in the expected collections because of similar product's node files often reference the same documents. Hence, the sum of *unique* documents $s_{u_{node}}$ is usually much less than expected.

A second algorithm therefore not only counts the number of documents, but traverses through T computing the list of unique documents for each node, not including the documents of sub-nodes that form a collection themselves.

For some collections, the sum $s_{u_{node}}$ can still exceed the expected maximum collection size. If they are composed of documents from multiple nodes in T they may be split, but for the evaluated collections it is not reasonable to do so due to the topical clustering. Using this algorithm, three test sets have been generated as shown in Table 1.

Table 1. Test sets generated from the document samples

Test set	#Collections	Expected size range	Overflow of $s_{u_{node}}$
TS_{small}	≈ 230	20–50 documents	35 collections (4 > 100 documents)
TS_{large}	9	250–500 documents	1 collection (557 documents)
TS_{skew}	≈ 50	manually compiled from TS_{small} and TS_{large} aiming for low overlap, to analyse shortcomings w.r.t. collection skew	

The queries have partly been taken from a developing gold standard of the FPIS. As an exceptionality of the FPIS, they are typically *hybrid*, i.e., they consist of semantic elements identifying concepts or instances of the ontology and literals which resemble keywords.

The query intents classified by Broder [12] for Web search (informational, navigational, transactional) can not be applied directly, but the queries can roughly be distinguished between:

– *immediate informational:* the query should return one document and ideally answer the information need in the first document snippet
– *composed informational:* the information need can't be answered by a single document, but several relevant documents need to be studied for an answer

Navigational queries can be considered similar to immediate informational queries in that they focus on a single document (instead of a certain Web site), whereas transactional queries are inapplicable here.

4.2 Results

The CORI algorithm has been modified to select a variable set of collections, based on a fixed threshold. It produced mixed but consistent results for a set of 12 queries. With short queries identifying a certain product, CORI typically selected very few selections and ranked the most relevant with a high accuracy.

For *immediate informational* queries, such as "[product] error 3", performance dropped significantly, apparently because the discriminating first query part was suppressed. The algorithm failed to rank the most relevant resource topmost for about half the queries, but it always remained above the threshold.

Composed informational queries showed a worse performance, with a distinct uncertainty in the selection results indicated by a low precision. For example, the query "'sensor drive' fitting procedure" yields some documents explaining how to install such a product option, but CORI fails to accurately distinguish collections using these barely specific terms. Furthermore, applying clustering to the distribution of CORI scores would not clearly discern relevant sources.

5 Conclusions

The preliminary evaluation shows that the performance of existing syntactic algorithms varies considerably regarding the kind of query. For more ambiguous queries, the syntactic approach is blatantly limited. Resource selection performance will probably benefit from a more thorough knowledge based analysis. The envisioned federated product information system supports this extension by providing semantic annotation services and an integrated hybrid query processing. Thus, users are encouraged to explicitly define the intended query terms in order to improve precision.

Future research will evaluate the existing algorithm quantitatively, based on an extended set of queries, and propose an index structure for efficient matchmaking of semantic query terms. We expect that an independent feature model and algorithms like Naive Bayes can be applied to combine the individual plugins' results.

Acknowledgement. This work was funded by the German Ministry of Education and Research under the research grant number 01IA08001F.

References

1. Wauer, M., Schuster, D., Meinecke, J.: Aletheia: an architecture for semantic federation of product information from structured and unstructured sources. In: Proceedings of the 12th International Conference on Information Integration and Web-based Applications & Services, iiWAS 2010, pp. 325–332. ACM, New York (2010)
2. Callan, J.: Distributed information retrieval. In: Advances in Information Retrieval, pp. 127–150. Kluwer Academic Publishers (2000)
3. Callan, J.P., Lu, Z., Croft, W.B.: Searching distributed collections with inference networks. In: Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1995, pp. 21–28. ACM, New York (1995)
4. Si, L., Lu, J., Callan, J.: Distributed information retrieval with skewed database size distributions. In: Proceedings of the 2003 Annual National Conference on Digital Government Research. dg.o 2003, pp. 1–6. Digital Government Society of North America (2003)
5. Si, L., Callan, J.: Relevant document distribution estimation method for resource selection. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2003, pp. 298–305. ACM, New York (2003)
6. Shokouhi, M.: Central-rank-based collection selection in uncooperative distributed information retrieval. In: Amati, G., Carpineto, C., Romano, G. (eds.) ECIR 2007. LNCS, vol. 4425, pp. 160–172. Springer, Heidelberg (2007)
7. Thomas, P., Shokouhi, M.: SUSHI: scoring scaled samples for server selection. In: SIGIR 2009: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 419–426. ACM, New York (2009)
8. Hong, D., Si, L., Bracke, P., Witt, M., Juchcinski, T.: A joint probabilistic classification model for resource selection. In: Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2010, pp. 98–105. ACM, New York (2010)
9. Arguello, J., Callan, J., Diaz, F.: Classification-based resource selection. In: Proceeding of the 18th ACM Conference on Information and Knowledge Management, CIKM 2009, pp. 1277–1286. ACM, New York (2009)
10. Ferrucci, D., Lally, A.: UIMA: an architectural approach to unstructured information processing in the corporate research environment. Nat. Lang. Eng. 10(3-4), 327–348 (2004)
11. Clark, J., DeRose, S.: XML Path Language (XPath) version 1.0. Recommendation, World Wide Web Consortium (November 1999), http://www.w3.org/TR/xpath.html
12. Broder, A.: A taxonomy of web search. SIGIR Forum 36, 3–10 (2002)

Knowledge-Based MDA Approach

Audrius Lopata[1,2] and Martas Ambraziunas[1]

[1] Vilnius University, Kaunas Faculty of Humanities,
Muitines St. 8, Kaunas, Lithuania
[2] Kaunas University of Technology, Information Systems Department,
Studentu St. 50, Kaunas, Lithuania
{Audrius.Lopata,Martas.Ambraziunas}@khf.vu.lt

Abstract. Knowledge-Based MDA is IS development approach created using core MDA components (models and transformations) as well as the basic elements of Knowledge Based IS engineering (Enterprise Model, Enterprise Meta- Model). This approach enhances basic MDA with Enterprise Model and two transformations in order to improve user requirements acquisition and enterprise modeling stages of IS development process.

Keywords: Knowledge-Based IS Engineering, Model Driven Engineering, Model Driven Architecture, Enterprise Model, Enterprise Meta- Model.

1 Introduction

There are various advanced IS development approaches. Traditional methods are constantly updated, but despite this, some fundamental IS development issues are still relevant. One of the most important issues is software validation against user requirements [2]. The expert plays a pivotal role in the problem domain knowledge acquisition process, and few formalized methods of knowledge acquisition control are taken into consideration, therefore user requirement analysis and enterprise modelling process heavily depend on the expertise of system analyst and the user.

One of the most perspective software development approaches is called Model Driven Engineering (MDE). This approach emphasizes heavily the usage of domain models and provides the following benefits: increases short and long term productivity [10]; simplifies the process of requirements acquisition and system design; improves understanding and communication among stakeholders; positively influences requirements of engineering and enterprise modelling stages.

The most known and widely supported MDE's subset is Model Driven Architecture (MDA) proposed by Object Management Group (OMG). The main goal of MDA is to provide a framework which integrates the existing OMG standards including UML, MOF, XMI, and QTV in order to improve collaboration among participants of IS engineering process, including users, system analysts, system architects, developers, etc. Due to abstract MDA nature this approach can be put into practise as standalone Information Systems Engineering (ISE) methodology or as the

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 160–165, 2011.

basis for development of various ISE related methodologies (e.g. RUP MDA, Agile MDA etc.). However, MDA doesn't specify a detailed users' requirements validation technique [1].

In order to address this issue, the basic MDA approach can be enhanced by best practices of Knowledge-Based ISE. The main idea of Knowledge-Based ISE is as follows: problem domain knowledge necessary for ISE process is validated against formal criteria and stored in Knowledge-Based subsystem. The Knowledge-Based subsystem consists of Enterprise Model (EM), Enterprise Meta-Model (EMM) and is the main problem domain knowledge source for ISE stages like: design, development, deployment. MDA enhanced by Knowledge-Based subsystem is called Knowledge Based – MDA.

Knowledge-Based MDA approach is presented in the article. This approach combines the main components (modelling language, models, mappings among models, and process [11]) of traditional MDA and best practices of Knowledge-Based ISE. Partial transformation process among Knowledge-Based MDA internal models is presented in this article as well. Knowledge-Based MDA combines two software development approaches: Knowledge-Based IS Engineering (KBE) and Model Driven IS Engineering (MDE). Core element of KBE is Enterprise meta-model. There exist many different enterprise modelling standards including: CEN EN 12204 [5], CEN EN 40003 (CIMOSA) [6], UEML [7] which are the basis for EMM's internal structure.

2 MDE Approach

Model Driven Engineering is a branch of software development methodologies that heavily uses domain models during all software development life cycle stages including requirements acquisition, enterprise modelling, system design and development, and deployment stages. Depending on particular MDE methodology, domain models can be used as data and activities'_representation or/and as the main data source for programming code generation. Currently most MDE based tools support [4] direct programming code generation from domain models. "A model driven engineering approach must specify the modelling languages, models, translations between models and languages, and the process used to coordinate the construction and evolution of the models" [12]. In order to put MDE into practise all these components should be defined. Modelling languages can be classified into two major categories: Domain Specific Modelling Languages (DSML) or general purpose modelling languages. Mostly DSMLs are trivial and have a small number of notation elements. They are created for specific domain. DSML can support a particular domain very well, but usually these modelling languages lack possibilities to depict large and complex systems. In order to model such kind of systems general purpose modelling languages are used. One of the most popular standard for modelling complex IS systems is UML. UML's internal structure is created in a way that allows a constant enhancement and development. This language supports OMG's provided modelling and data transferring standards like MOF [4] and XMI [4].

Model is a "representation of the structure, activities, processes, information, resources, people, behaviour, goals, and constraints of a business, government, or other enterprises” [17]. In software development there are two major types of models: business oriented models and technical oriented models. Usually there is a logical gap between these models because business oriented models is created by system analyst and user; technical oriented models are created by system architects and developers. Models are used for better understanding between IS project stakeholders or for software development directly.

Model transformation is a process during which a particular model element is depicted as another model element. This process can be performed if there are mapping specifications between meta-models only. There are two types of model transformations: horizontal and vertical. Horizontal transformation deals with the same level of abstraction but of different view or notation. Vertical transformation deals with different level of abstraction. There exist mapping languages/frameworks (e.g. OMG’s QVT [4]) that provide specifications for mapping process and constructs.

MDE process defines actions, actions implementation sequence and actors in order to perform vertical and horizontal models transformations. In the ideal scenario MDE process describes a full IS development life cycle from requirements acquisition to IS deployment.

In 2001 OMG presented MDA approach which specifies the appliance of domain models in the software development life cycle. The domain model is a description or specification of system and its environment for some certain purpose. A model is often presented as a combination of drawings and text [4]. The main concept of MDA is to separate the specification of system functionality from the specification of the implementation of that functionality on a specific technology platform [4] (“What” to do from “How” to do).

3 Knowledge-Based MDA Approach

Knowledge-Based MDA approach combines the basic MDE and KBE principles. The main components of this approach are presented in figure 1.

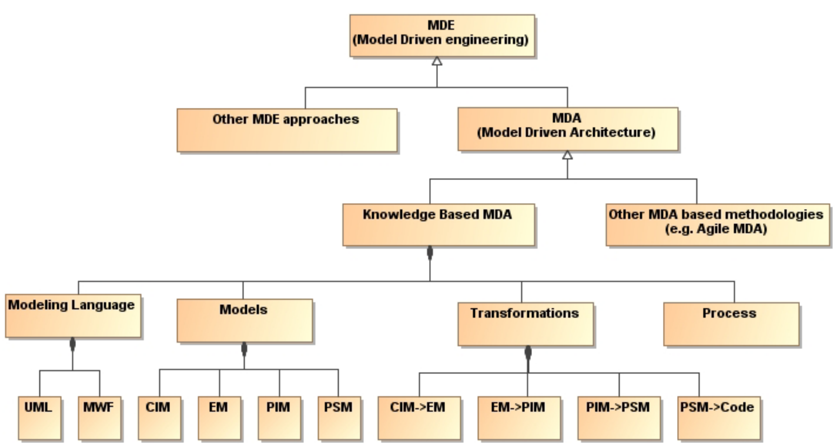


Fig. 1. The main Knowledge-Based MDA elements

MDE defines the main principles of domain models usage in IS development process. MDE is the basis for MDA approach. This MDE branch is defined by OMG using standards like MOF, XMI, and UML. Although MDA is less abstract approach than MDE but it still provides only guidelines (and standards) for IS development, but not a real working solution. The next abstraction level is methods based on MDA e.g. Agile MDA [3] as well as Knowledge-Based MDA belongs to this level. Basically Knowledge-Based MDA extends MDA with additional model (Enterprise Model) and two transformations (CIM to EM and EM to PIM). A more detailed description of Knowledge-Based MDA specific elements are presented in the chapters listed below.

3.1 Modeling Languages

The following modelling languages are selected for implementation of Knowledge-Based MDA: Modified Work Flow models (WFM) [9] and UML. WFM is used for problem domain analysis stage and UML - for IS design and implementation stages. Modified Work Flow models are successfully used in problem domain knowledge acquisition and specification into EM surveys [1]. This technique was proposed by the same scientific group that created EMM (that is one of the main elements in KB MDA).

3.2 Models

As presented KB MDA defines four types of models: CIM, EM, PIM, and PSM. CIM specifies system requirements of a particular problem domain (it can also be named as Business Model). Usually this model is created by a system analyst and user, therefore CIM's nature is empirical. EM is enterprise model based on EMM. PIM provides system's architecture and functionality without platform specific information and thus, technical details are used in the IS design stage. PSM is constructed on the basis of PIM enhancing it with platform specific details e. g. information about IS implementation and deployment. PSM is the main data source for the code generation process. All these models consist of the following components [13] CIM contains Modified Workflows Models and in some cases SysML models. Knowledge acquisition questionnaires can be used as well. EM is a part of element called Knowledge-Based subsystem. This subsystem also contains EMM and transformation algorithm that handles CIM to EM and EM to PIM transformations. EM is implemented as database logical scheme. PIM consists of three UML models: Class model, Sequence model, and Use Case model. PSM includes the same model types as PIM but is appended with platform specific data. Application element represents three layers software architecture that can be generated using fully specified PSM model. EM and PIM structures are described in a more detailed way[13]. When transformations are performed from EM to PIM, transformations are performed with these models.

3.3 Transformations

KB MDA defines four basic transformations CIM to EM, EM to PIM, PIM to PSM, and PSM to Code. Transformations are defined in Table 1.

Table 1. Knowledge- Based MDA transformations

Transformation	Description
CIM to EM	CIM is used for problem domain data acquisition and representation. In a usual case the data are not validated against formal criteria. Directly CIM cannot be validated automatically due to complexity and different internal CIM models notations. To perform this validation CIM must be transformed into EM which is created on Control Theories [8] principles.
EM to PIM	EM to PIM transformation is performed when EM is fully validated against Control Theory based formal criteria. If these criteria are met, PIM model can be created and next stages of MDA can be performed.
PIM to PSM	PIM to PSM transformation is a part of traditional MDA transformations (along with CIM to PIM and PSM to Code).
PSM to Code	During PSM to Code transformation, PSM elements are translated to a particular programming language.

Transformations described in the table posses more complex internal structure and processes [13].

3.4 Process

Knowledge-Based MDA method can support forward IS engineering as well as backward. Forward engineering phase starts from CIM construction. This process is performed by System Analyst using numerous techniques [11] to capture problem domain knowledge. Next step is EM creation from CIM. This process is performed by System Analyst and partly Knowledge-Based Subsystem. Constructed EM is validated against EMM. If EM conforms to EMM provided constrains, EM to PIM transformation process is launched. This process is performed by Knowledge-Based Subsystem. If validation fails, System Analyst revises CIM according to validation report. After revision and CIM updating, EM is partly or fully recreated and validation is launched again. The cycle *CIM->EM->EM validation* can have several iterations. Eventually when EM is consistent with EMM internal structure, next transformations can be performed (EM to PIM, PIM to PSM, and PSM to Code).

Reverse engineering starts as usual from Code (working software) transformation to PSM. This process is performed by transformation tool. Particular MDA compatible tool performs PSM to PIM transformation process, removing or transforming platform related constructs to higher abstraction (PIM) level in next step. Knowledge-Based Subsystem handles PIM transformation to EM process. The final reverse engineering result is EM which is consistent with analyzed IS. At this point EM can be used for two main purposes: specification and analysis of information system architecture from Control Theory [8] view or improvement of existing IS by updating problem domain knowledge that will start forward engineering process.

4 Conclusions

Traditional MDA approach enhanced by Knowledge-Based subsystem is called Knowledge-Based MDA. This approach consists of the following components:

modelling languages (UML, WFM), models (CIM, EM, PIM, and PSM), mappings/transformations among models (CIM to EM, EM to PIM, PIM to PSM, and PSM to Code) and process.

The Knowledge-Based subsystem (consisting of EM and EMM) is an intermediate component between CIM and PIM that stores validated enterprise knowledge necessary for IS engineering process. The mapping specifications among PIM's internal models (Class model, Use Case model, and Sequence model) and EM are presented in this article as well as the algorithm specifying EM knowledge-based generation of Class model.

The Knowledge-Based MDA approach ensures the problem domain knowledge validation against formal criteria, thus improving the consistency of software artefacts and reducing IT projects dependency on empirical processes in problem domain knowledge acquisition and user requirements specification stages.

Practical implementation of approach described above is in progress using the following technologies: UML, .Net, MS SQL server.

References

1. Asadi, M., Ramsin, R.: MDA-based Methodologies: An Analytic Survey. In: Proceedings of the 4th ECMDA: Foundations and Applications, Berlin, pp. 419–431 (2008)
2. Ellis, K.: The Impact of Business Requirements on the Success of Technology Projects. IAG Consulting, Benchmark (2008)
3. Ambler, S.W.: Agile Modeling, `http://www.agilemodeling.com/essays/inclusiveModels.htm`
4. OMG, `http://www.omg.com`
5. ENV 12 204. Advanced Manufacturing Technology Systems Architecture - Constructs for Enterprise Modelling. CEN TC 310/WG1 (1996)
6. ENV 40 003. Computer Integrated Manufacturing Systems Architecture - Framework for Enterprise Modelling, CEN/CENELEC (1990)
7. Vernadat, F.: UEML: Towards a Unified Enterprise modeling language. Proceedings of MOSIM 2001, Troyes, France, 2001-04-25/27, `http://www.univ-troyes.fr/mosim01`
8. Gupta, M.M., Sinha, N.K.: Intelligent Control Systems: Theory and Applications. The Institute of Electrical and Electronic Engineers Inc., New York (1996)
9. Lopata, A., Gudas, S.: Workflow- Based Acquisition and Specification of Functional Requirements. In: Proceedings of 15th International Conference on Information and Software Technologies IT 2009, Kaunas, Lithuania, April 23-24, pp. 417–426 (2009) ISSN 2029-0020
10. Haan, J.: Model Driven Engineering, `http://www.theenterprisearchitect.eu/archive/2008/03/14/model_driven_engineering`
11. Kent, S.: Model Driven Engineering. In: Butler, M., Petre, L., Sere, K. (eds.) IFM 2002. LNCS, vol. 2335, pp. 286–298. Springer, Heidelberg (2002)
12. Fox, M.S., Gruninger, M.: Enterprise Modeling. American Association for Artificial Intelligence (1998), `http://www.eil.utoronto.ca/enterprise-modelling/papers/fox-aimag98.pdf`
13. Lopata, A., Ambraziunas, M.: MDA Compatible Knowledge– Based IS Engineering Approach. In: Wang, F.L., Deng, H., Gao, Y., Lei, J. (eds.) AICI 2010. LNCS, vol. 6319, pp. 230–238. Springer, Heidelberg (2010)

Automated Ontology Evolution for an E-Commerce Recommender*

Elmar P. Wach[1,2]

[1] STI Innsbruck, University of Innsbruck,
Technikerstraße 21a, 6020 Innsbruck, Austria
elmar.wach@sti2.at
[2] Elmar/P/Wach eCommerce Consulting,
Hummelsbüttler Hauptstraße 43,
22339 Hamburg, Germany
wach@elmarpwach.com

Abstract. This research proposes a completely automated OWL product domain ontology (PDO) evolution by enhancing an existing ontology evolution concept. Its manual activities are eliminated by formulating an adaptation strategy for the conceptual aspects of an automated PDO evolution and establishing a feedback cycle. This strategy decides when and how to evolve by evaluating the impact of the evolution in the precedent feedback cycle and is implemented in a new adaptation layer. The adaptation strategy was validated/ firstly "instantiated" by applying it to a real-world conversational content-based e-commerce recommender as use case.

Keywords: Ontology Evolution, Ontology Versioning, Recommender Systems, Self-Adapting Information Systems, Heuristics.

1 Introduction

Recommender systems in e-commerce applications have become business relevant in filtering the vast information available in e-shops (and the Internet) to present useful product recommendations to the user. As the range of products and customer needs and preferences change, it is necessary to adapt the recommendation process. Doing that manually is inefficient and usually very expensive. Recommenders based on product domain ontologies[1] (PDO) modelling the products offered in the e-commerce application can extract questions about the product characteristics and features to investigate the user preference and eventually recommend products that match the needs of the user. By changing the PDO, such a recommender generates different questions and/ or their order. Hence, an automated adaptation of the recommendation

* The research presented in this paper is funded by the Austrian Research Promotion Agency (FFG) and the Federal Ministry of Transport, Innovation, and Technology (BMVIT) under the FIT-IT "Semantic Systems" program (contract number *825061*).

[1] A product domain ontology (PDO) is defined as the formal, explicit specification of a shared conceptualisation of a product description based on OWL DL; this definition is derived from [Gruber, T. R. 1993].

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 166–177, 2011.

process can be realised by automatically evolving the PDO[2]. The high cost of the manual adaptation of the recommendation process and the underlying PDO can herewith be minimised.

This research proposes a completely automated OWL PDO evolution (without a human inspection) based on given user feedbacks[3] and enhancing an existing ontology evolution concept. Its manual activities are eliminated by formulating an adaptation strategy for the conceptual aspects of an automated PDO evolution and establishing a feedback cycle. Automatically evolving the PDO is more efficient and less expensive than manually doing it. The present research tackles an automated process for the first time (to the best knowledge of the author).

2 Related Work

Previous approaches to the topic of this research can be found in concepts for ontology evolution like formulated frameworks for ontology evolution, e.g. [Haase, P. et al. 2005], [Klein, M. and Noy N. F. 2003], [Konstantinidis, G. et al. 2007], [Noy, N. F. et al. 2006], [Stojanovic, L. et al. 2002], [Stojanovic, N. et al. 2003], [Zablith, F. 2009]. Due to the specific challenges of the present research like the automated ontology evolution process, none of the identified frameworks can be completely used as basis, e.g. all of the frameworks include a step for the human inspection of the ontology changes before they are executed. The closest work to the research in this paper is [Stojanovic, L. et al. 2002] – in the six phase evolution process, two steps include manual activities, namely (i) "Implementation" in which the implications of an ontology change are presented to the user and have to be approved by her before execution, and (ii) "Validation" in which performed changes can get manually validated. This research aims at eliminating both manual steps in [Stojanovic, L. et al. 2002] with the adaptation strategy and its implementation. To automate (i), the ontology evolution is conceptualised and implemented as a complete feedback cycle [Bennett, K. H. and Rajlich, V. T. 2000]. An insufficient ontology change is indicated by decreased metrics and gets revised according to the evolution strategy chosen. Hence, the ontology changes do not have to get manually approved before execution. To automate (ii), the PDO changes are predefined and application-oriented. Hence, only valid changes are executed, and nobody has to manually validate them. This approach is addressed with the adaptation strategy and its implementation as a new adaptation layer consisting of two components [Broy, M. et al. 2009].

3 Adaptation Strategy

The adaptation strategy addresses when a change has to be executed and how the changes will be executed in the PDO by evaluating the impact of the evolution in the

---

[2] Ontology evolution is defined as the timely adaptation of a PDO by preserving its consistency (a PDO is consistent if and only if it preserves the OWL DL constraints); this definition is derived from [Haase, P. and Stojanovic, L. 2005] and [Suárez-Figueroa, M. C. and Gómez-Pérez, A. 2008].

[3] In order to focus this research on developing an automated ontology evolution, the feedback is assumed to be given.

precedent feedback cycle. The first question defines the (temporal and causal) trigger initiating the PDO change. This is addressed with the feedback transformation strategy (confer section 3.1) which is implemented in the Feedback Transformer.

The second question defines the changing of the PDO with annotated instances (i.e. products). This is evolving the PDO and will be addressed with the PDO evolution strategy (confer section 3.2) which is implemented in the Adaptation Manager.

By following the principles of adaptive systems [Broy, M. et al. 2009], the strategies are implemented in a new adaptation layer (confer figure 1) consisting of components in which the user feedback gets transformed (i.e. Feedback Transformer) and the respective actions are decided and initiated (i.e. Adaptation Manager).

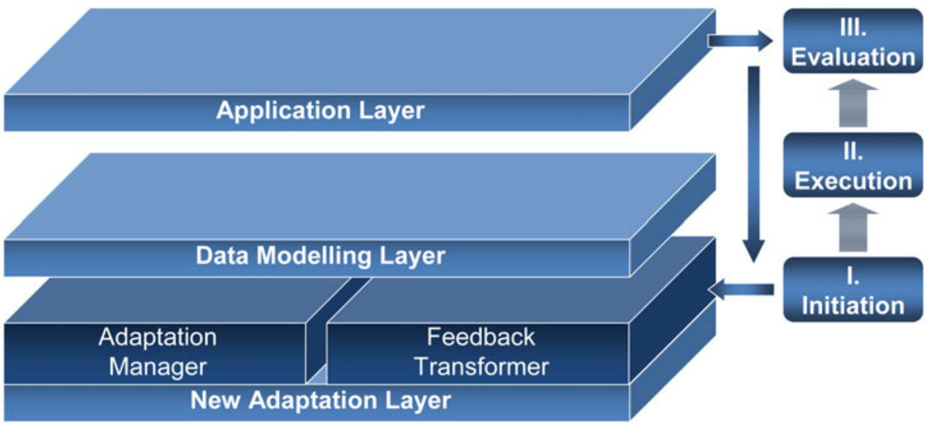


Fig. 1. Evolution cycle with a new adaptation layer

3.1 Feedback Transformation Strategy

The feedback transformation strategy defines when the PDO change. It transforms different kinds of user feedback (e.g. implicit, explicit) into ontology input (i.e. calculating Success Trends ST). This strategy is implemented in the Feedback Transformer where the user feedback channels and the PDO affected by the feedback reported are identified, the feedback is analysed and gathered, and eventually transformed.

The strategy comprises the following steps:

1. Identify the user feedback channels
2. Analyse and gather the user feedback
3. Transform the user feedback

Ad 1. Identify the user feedback channels

In this step the application setup is analysed with regard to the available user feedback. In order to focus this research on developing an automated PDO evolution, the feedback is assumed to be given, and thus extracting the information is out of scope. The application can provide two kinds of user feedback to get a complete view of the user: Internal data sources from the application layer like the KPI[4] or statistical evaluations of the usage. As the application is based on PDO, PDO changes influence the application behaviour, and KPI and statistical evaluations of the usage of the application layer are a valid feedback for the impact of the PDO evolution.

The application setup can also provide external data sources like data and information extractions from the Web, databases, or ontologies. E.g., discussions in Blogs and portals, official and unofficial product and product feature ratings, and appearances of new features and product aspects are valuable PDO information.

The two kinds of user feedback are delivered via different feedback channels that have to be identified and analysed with regard to the feedback representation, its accessibility, and the PDO affected. As the PDO is the backbone of a semantic application, the feedback is assumed to be RDF data. In case it is not, it is recommended to convert it to RDF[5]. A crucial aspect is the accessibility of the user feedback – can it be programmatically retrieved by the Feedback Transformer, e.g. via an API or from a SPARQL endpoint?

Ad 2. Analyse and gather the user feedback

In this step the user feedback channels and the feedback delivered are analysed with regard to the feedback content, structure, and meaning. In case the feedback is in RDF, e.g. it can be dynamically queried with SPARQL SELECT statements.[6] In order to adequately interpret the feedback, the metrics delivered have to be identified as well as their meaning have to be clear. Generally, there can be two types of feedback: Explicit user feedback could be provided by answering questions about the user satisfaction with the application. As this effort cannot be expected from a user, an alternative is to extract feedback from the Web that could also deliver new information and aspects about the products offered. Implicit user feedback is given by the user as a side-effect of the usage behaviour, e.g. by clicking on the product recommended.

Currently, two feedback channels with two types of feedbacks are defined:

- Implicit feedback channel (user feedback derived from user interactions in the application layer) “KPI trend”: The implicit feedback mainly evaluates the success/ usage/ usability of the e-commerce recommender; it is PDO-based
- Explicit feedback channel (user feedback extracted from the Web) “Feature relevance”: The explicit feedback gathers information about products based on PDO extractions and is represented as an annotation property; it is PDO- and property-based.

The RDF of the KPI trend feedback includes a value indicating a positive or negative trend of the defined KPI between two PDO versions (i.e. relating the currently evaluated feedback to the precedent one based on the previous changes). It is defined as KPI(t) with the range [-1…∞]. In rare cases, the value can calculatory be larger than +1. This feedback repository is queried with SPARQL SELECT statements to retrieve the KPI for each latest PDO version that represent a valid PDO version test.

[4] Key Performance Indicator, e.g. click-out rate (i.e. clicks-to-recommendations ratio).
[5] Generating the RDF data is out of scope.
[6] Due to space limitations, SPARQL statements are omitted.

The RDF of the feature relevance feedback includes the property (i.e. feature name) and its relevance, e.g. based on the count of appearances in the Web over a period of time. It is defined as Feat(t) with the range [0…+100]. After having retrieved feedback from the first feedback channel, this feedback repository is queried with SPARQL SELECT statements to retrieve the relevance for each latest PDO version to be changed.

Ad 3. Transform the user feedback

In this step the different types of feedback are transformed to ontology input, and thus a PDO change is requested. The impact of the change is measured by calculating adequate metrics for the new user feedback from the application layer and external data sources reported to the adaptation layer and for each feedback channel, defined as Success Trends ST_{ch}, e.g. with an algorithm, formula, or transformation. In case the feedback includes information extracted from the PDO, the transformed feedback has to be in the same representation as before (e.g. ontological entity, range).

In case several feedback channels deliver analogue feedbacks, the respective channels have to be weighted separately. The channel weight is a factor that expresses the relative importance of either feedback channel for the PDO evolution. It can be changed between two feedback cycles, though it is recommended to observe the quality of the feedbacks over time before tuning it. The weights for the corresponding analogue feedback channels sum up to 100%. Additionally, the ST calculation can respect the certainty of a feedback channel. The certainty expresses the probability of the correctness of the reported feedback as a percentage value.

The KPI trend KPI(t) is converted by a simple value transformation to the ST with the range [-1…+1] relating the currently transformed feedback to the precedent one. In the rare case of a KPI value larger than +1, it will be normalised to +1.

The feature relevance Feat(t) is converted by calculating the new relevance of the properties with the relative frequencies of the properties in the feature relevance feedback. The ST with the range [0…+100] is calculated by determining classes correlated to that range and based on the interval of the relative frequencies of the properties. To the classes the corresponding properties (i.e. the relative frequency of the property in the feature relevance feedback is within the bounds of the respective class) as well as the respective relevance are assigned. The new relevance is represented as before (i.e. as an annotation property).

After having transformed the different feedback types, the calculated ST are reported to the next component, i.e. the Adaptation Manager.

3.2 PDO Evolution Strategy

The PDO evolution strategy defines how the PDO change. It associates an evolution action to the ST and ensures a consistent new PDO version. This strategy is implemented in the Adaptation Manager where the structure of the respective PDO gets queried with SPARQL SELECT statements and the PDO changes are executed with SPARQL CONSTRUCT rules or programmatically according to an evolution heuristic and predefined evolution strategies. Alternatively, a statistical analysis of the user feedback and its history can be conducted.

The strategy comprises the following steps:

4. Define the representation of PDO changes
5. Define the analysis of the transformed feedback
6. Ensure a consistent ontology evolution and versioning.

Ad 4. Define the representation of PDO changes

In this step options for the representation of PDO changes are defined, e.g. reusing an existing representation. The change representation defines the possible and allowed PDO changes.

For deciding whether an existing representation of ontology changes should be reused, adequate evolution criteria have to be defined. An existing representation has to be investigated with regard to the PDO representation language (e.g. OWL 1, OWL 2) and the PDO changes (e.g. switching a specific individual, switching the range of a specific property) offered – they have to constitute the types of PDO changes[7] needed by the application and to be executed and evaluated in the next feedback cycle, i.e. PDO evolution cycle. In case the application utilises a specific PDO representation, this is the preferred basis for the representation of its changes as well – in this research the PDO are based on GoodRelations[8]. In case the necessary evolution criteria are not met by an existing or application-oriented representation, a customised one has to be developed, e.g. a specific ontology of changes.

As the PDO model the knowledge queried by the user, it is helpful to describe probable user scenarios to predefine the types of PDO changes needed.

Ad 5. Define the analysis of the transformed feedback

In this step options for the analysis of the transformed feedback are defined, e.g. statistical means or utilising a heuristic, and the adequate PDO evolution is decided. The impact of the PDO change is measured in the Feedback Transformer by calculating the ST for the new user feedback from the application layer and external data sources reported to the adaptation layer and for each feedback channel. The ST can be analysed by statistical means. The method as well as the relevant metrics has to be defined and the calculations formulated. By programmatically calculating the relevant metrics, a complete automation of the analysis as well as the derived evolution actions can be achieved.

Another option is to formulate and utilise a heuristic that defines the PDO change to be executed. A heuristic is a strategy that uses accessible and loosely applicable information to solve a problem of a human being or a machine [Pearl, J. 1983] and leads to a solution of a complex problem with simplified conceptual aspects or reduced computation power. [Glover, F. W. 1986] mentioned first the term metaheuristic for a computational method that makes few or no assumptions about the problem being optimised and introduced the tabu search metaheuristic [Glover, F. W. and Laguna, M. 1997] which is utilised in this research with the philosophy that the highest precedent ST (“greedy”) defines the next PDO change to always choose the best evolution.

---

[7] Currently defined are switching individuals, switching datatype property ranges, switching annotation properties label and comment, and changing annotation property priority.

[8] `www.purl.org/goodrelations`

The relevant characteristics of the heuristic have initially to be defined (confer section 4.). This manual effort is rewarded with a greater conceptual flexibility resulting in a more specific application-oriented evolution behaviour with regard to its impact on the application. The relevant metrics have to be defined and the calculations formulated.

Regardless of the analysis method chosen, the PDO evolution is decided based on the ST. In case the feedback includes information extracted from the PDO, the subsequent evolution (i.e. type of PDO change) is defined by implementing the ST in the same representation as before (e.g. ontological entity, range), and neither statistical means nor a heuristic has to be applied.

In case a heuristic is chosen, this research proposes to additionally formulate evolution strategies that decide the general evolution behaviour (e.g. executing the same type of PDO change or a rollback) by correlating the types of PDO changes needed to the ST calculated. Additionally, the path for determining the initial ST has to be defined, e.g. the order of the different types of PDO changes and for which PDO they are executed (i.e. ramp-up of the evolution strategies). The philosophy should be that the development (and its strength) of the precedent ST defines the next type of PDO change to distinguish different evolution impacts.

The predefined evolution strategies summarised in table 1 are considered as basic categories. They can be fine-tuned with regard to the associated types of PDO changes as well as the threshold defining the trend significance.

Table 1. Evolution strategy, Success Trend ST, and associated type of PDO change

Evolution Strategy	Decision Criteria	Type of PDO Change
Risky Evolution ("always evolve differently")	$-1 \leq ST \leq 1$	Different than before
Progressive Evolution ("learn from the past")	$0{,}2^* \leq ST \leq 1$ $0 \leq ST < 0{,}2^*$ $-1 \leq ST < 0$	Same as before Different than before Different than before or Rollback
Safe Evolution ("only revert negative trends")	$0 \leq ST \leq 1$ $-1 \leq ST < 0$	None Rollback
Rollback ("undo the ontology changes")	Manually	Rollback

* Threshold trend significance: Increase of the ST by 20 basis points between the precedent and the current feedback cycle.

Each evolution strategy besides Rollback ensures an adaptive change of the PDO. By selecting a strategy in the administration interface, the business manager decides how fundamental the evolution will be.

Ad 6. Ensure a consistent ontology evolution and versioning

After having chosen the PDO changes to be executed, the PDO has to evolve depending on rules and by retaining its consistency to eventually provide its knowledge to the application layer. This is done by executing SPARQL CONSTRUCT rules or programmatically. Due to space limitations, the rules are omitted.

When evolving the PDO, it has to be clear how the PDO has been evolved over time, i.e. the different PDO evolutions have to be versioned. By versioning a PDO, its changes get documented, and the historical path of evolution gets traceable. In the context of this research this is of paramount importance for deciding the next PDO change to be executed and reverting the changes executed in the precedent feedback cycle, i.e. a rollback.

The preferred concept of ontology versioning is change-based versioning (i.e. each state gets its own version number and additionally stores information about the changes made), because it facilitates change detection, integration, conflict management [Mädche, A. et al. 2003], and it allows the interpretation how PDO changes influence the metrics. A change-based versioning can be best realised by tracking the PDO changes in a semantic log [Mädche, A. et al. 2002].

4 Evaluation and Validation

The adaptation strategy has been validated/ “instantiated” by applying it to the use case which is a real-world conversational content-based e-commerce recommender system based on PDO that semantically describe the products offered in e-commerce applications according to GoodRelations. Implicit user feedback is derived from user interactions in the application layer and gathered by unobtrusively monitoring user needs. Explicit user feedback is gathered by extracting information from various websites. Both feedback channels deliver RDF data via separate SPARQL endpoints programmatically accessible. Four types of PDO changes are defined, i.e. switching individuals, switching datatype property ranges, switching annotation properties label and comment, and changing annotation property priority.

Applying the adaptation strategy could be done quite smoothly. Only minor aspects of the strategy were clarified, restructured, and reformulated. After having applied the strategy, the use case was concisely described and conceived by the ontology engineer. Moreover, the result formed the basis of the technical specification and thus the development of the adaptation layer.

Due to space limitations the “instantiation” of the adaptation strategy is not completely elaborated in this paper. In the following the evolution heuristic based on tabu search is introduced in extracts (excluding its ramp-up, for instance). The “taboos” are defined as follows:

- **General tabu criterion gt:**
  - To avoid an uniform optimisation and cycles, the PDO changes within the same type of PDO change are consecutively executed only as often as there are different types T of PDO changes not induced by a feedback based on a PDO extraction
  - Exception: In case a type of PDO change has less than T PDO changes, the general tabu criterion is met when all PDO changes within the respective type of PDO change have been executed
  - The general tabu criterion gt is calculated by multiplying the two specific tabu criteria defined below; result is the number of allowed PDO changes gt; the PDO changes are sequentially executed and added to the tabu list

- After the ramp-up phase and in case the general tabu criterion gt or T is met, the PDO change with the highest ST in another type of PDO change is going to be executed and ST(t+1) calculated.

– **Specific tabu criteria (specifically calculated for each type of PDO change):**

- "Allowed number of horizontal switches" sw: With sw one (set of) ontological entity of a PDO within the same type of PDO change is switched, e.g. a PDO change of one (set of) property or (set of) individual – most of times there is only one switch possible like changing the individual, the property range, or the annotation properties label and comment, and the next change would be reverting that change. This tabu is defined as follows:

$$\mathrm{sw} = \begin{cases} 0, \text{ case: } p = 1 \wedge c_{fix} = 0 \\ 2 + c_{fix}^2 / 2 - c_{fix}, \text{ case: } p = 1 \wedge c_{fix} = 2*k,\ c_{fix}, k \in \mathbb{N} \setminus \{0\} \\ 1 + c_{fix} * (c_{fix} - 1) / 2, \text{ case: } p = 1 \wedge c_{fix} = 2*k - 1,\ k \in \mathbb{N} \setminus \{0\} \\ 1 + p^2 / 2 - p, \text{ case: } p > 1 \wedge p = 2*k,\ p \in \mathbb{N} \setminus \{0,1\},\ k \in \mathbb{N} \setminus \{0\} \\ p * (p - 1)/2, \text{ case: } p > 1 \wedge p = 2*k - 1,\ p \in \mathbb{N} \setminus \{0,1\},\ k \in \mathbb{N} \setminus \{0\} \end{cases} \tag{1}$$

(c_{fix} being the number of fixed candidates within a type of PDO change (i.e. to these candidates can be switched), p being the number of pools of sets of entities (e.g. each source for the properties is a pool like string ranges, Boolean ranges, DBpedia, or WordNet; p can be changed for each type of PDO change in the administration interface); a pool p can be switched on the level of ontological entity (s') or completely (s), i.e. all sets of ontological entities are switched at once (can be changed for each type of PDO change in the administration interface, in case of more than one data pool p), k being a natural number to indicate an even ($c_{fix} = 2 * k$, $p = 2 * k$) or odd ($c_{fix} = 2 * k - 1$, $p = 2 * k - 1$) number of fixed candidates or pools: The case for the even c_{fix} or p equates to an Eulerian trail, the case for the odd c_{fix} or p to an Eulerian circuit).

Result is the number of allowed switches sw. In case s is already connected to c_{fix} (e.g. s - c_{fix} = 1), the second and third case in (**1**) are lessen by this one "impossible" switch (i.e. sw_{fix} = sw - 1). In case sw is met, the PDO change with the second highest ST within the same type of PDO change is going to be executed and ST(t+1) calculated.

- "Allowed number of vertical PDO change iterations" ch: With ch successive sw switches within the same type of PDO change are executed, i.e. the next (sets of) ontological entities are going to be switched. This tabu is defined as follows:

$$\mathrm{ch} = \begin{cases} (s - \mathrm{ch}_{\mathrm{fix}})\,/\,n;\ \text{case: } p = 1,\ n \in \mathbb{N} \setminus \{0\},\ s,\ \mathrm{ch}_{\mathrm{fix}} \in \mathbb{N},\ s \geq \mathrm{ch}_{\mathrm{fix}} \\ \mathrm{s'}/\,\mathrm{n},\ \text{case: } p > 1 \wedge \mathrm{s'} \subset \mathrm{s} \text{ (i.e. single sets)},\ n \in \mathbb{N} \setminus \{0\},\ \mathrm{s'} \in \mathbb{N} \\ \text{Not applicable, case: } p > 1 \wedge \mathrm{s'} \equiv \mathrm{s} \text{ (i.e. all sets at once)} \end{cases} \tag{2}$$

ch is truncated to the natural number.

(s being all sets of ontological entities within a type of PDO change (e.g. all sets of individuals, all sets of properties, all sets of annotation properties label and comment), s' being a single set of ontological entities within a type of PDO change (e.g. specific properties) to be switched to another pool, n being the fraction of the sets of entities within a type of PDO change allowed to be switched (e.g. n = 1: All sets of entities, n = 2: Half of the sets, etc.; n can be changed for each type of PDO change in the administration interface)).

Result is the number of allowed PDO change iterations ch. Analogous to the case distinction of the horizontal switches sw and sw_{fix}, ch is splitted in the first case in (**2**) into s is not connected to c_{fix} before switching (ch), and s is already connected to c_{fix} before switching (ch_{fix}).

– In case another type of PDO change is executed, the oldest tabu of the precedent type of PDO change is deleted from the tabu list.

In addition to this validation, the adaptation strategy is going to be evaluated by conducting an experiment with approximately thirty ontology experts who analyse and formulate ontology evolution characteristics. These are then aligned with the adaptation strategy and adopted accordingly where applicable.

The adaptation layer is going to be evaluated by conducting an experiment with approximately thirty ontology experts who evaluate the ontology evolution. The automatically evolved PDO is going to be compared with a manually evolved one by setting up and evaluating an experiment with ontology experts who analyse the feedbacks delivered and decide the PDO changes to be executed. Eventually, the PDO resulted from this manual evolution is compared with the automatically evolved one regarding the evaluation criteria consistency, completeness, conciseness, expandability, and sensitiveness [Gómez-Pérez, A. 2001].

The adaptation layer is going to be validated by programming the layer and measuring the effects in the e-commerce recommender system. Its success is defined by the click-out rate (i.e. clicks-to-recommendations) that measures the impact of the PDO evolution induced by the implicit and explicit user feedback.

The intended results are a highly adaptive system and eventually better recommendations given to the customer leading to an increase of the defined KPI. The expected business impacts are a higher customer satisfaction and loyalty and eventually increased revenue for the provider of the e-commerce application (and the recommender system).

5 Conclusion

The need for automatically updating and evolving ontologies is urging in today's usage scenarios. The present research tackles an automated process for the first time (to the best knowledge of the author). The reason for that can be found in the ontology definition "formal, explicit specification of a shared conceptualisation" [Gruber, T. R. 1993]. "Shared" means the knowledge contained in an ontology is consensual, i.e. it has been accepted by a group of people. Entailed from that, one can argue that by processing feedback in an ontology and evolving it, it is no longer a shared conceptualisation but an application-specific data model. On the other hand, it is still shared by the group of people who are using the application. It may even be argued that the ontology has been optimised for the usage of that group (in a specific context or application) and thus is a new way of interpreting ontologies: They can also be a specifically tailored and usage-based knowledge representation derived from an initial ontology – an ontology view, preserving most of the advantages like the support of automatically processing information. Thus, this changed way of conceiving ontologies could facilitate the adoption and spread of using this powerful representation mechanism in the real world, as it is easier to accomplish consensus within a smaller group of people than a larger one.

In this research the PDO are based on GoodRelations and evolve within that upper ontology. This ontology as well as the "subsumed" PDO conforms to the ontology definition by [Gruber, T. R. 1993]. The PDO are application-specific and evolve according to the needs of their users. Hence, they offer the advantages of both worlds.

References

[Bennett, K. H. and Rajlich, V. T. 2000] Software maintenance and evolution: A roadmap. In: Proceedings of the Conference on the Future of Software Engineering, pp. 73–87

[Broy, M. et al. 2009] Formalizing the notion of adaptive system behavior. In: Proceedings of the 2009 ACM Symposium on Applied Computing (SAC 2009), pp. 1029–1033 (2009)

[Glover, F. 1986] Future paths for integer programming and links to artificial intelligence. Comput. Oper. Res. 13, 533–549

[Glover, F. and Laguna, M. 1997] Tabu Search. Kluwer Academic Publishers

[Gómez-Pérez, A. 2001] Evaluation of ontologies. International Journal of Intelligent Systems 16, 391–409

[Gruber, T. R. 1993] Toward principles for the design of ontologies used for knowledge sharing. In: Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer Academic Publishers

[Haase, P. and Stojanovic, L. 2005] Haase, P., Stojanovic, L.: Consistent evolution of OWL ontologies. In: Gómez-Pérez, A., Euzenat, J. (eds.) ESWC 2005. LNCS, vol. 3532, pp. 182–197. Springer, Heidelberg (2005)

[Haase, P. et al. 2005] Haase, P., van Harmelen, F., Huang, Z., Stuckenschmidt, H., Sure, Y.: A framework for handling inconsistency in changing ontologies. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) ISWC 2005. LNCS, vol. 3729, pp. 353–367. Springer, Heidelberg (2005)

[Klein, M. and Noy N. F. 2003] A component-based framework for ontology evolution. In: Proceedings of the IJCAI 2003 Workshop on Ontologies and Distributed Systems (2003)

[Konstantinidis, G. et al. 2007] Ontology evolution: A framework and its application to RDF. In: Proceedings of the Joint ODBIS & SWDB Workshop on Semantic Web, Ontologies, Databases

[Mädche, A. et al. 2002] Managing multiple ontologies and ontology evolution in Ontologging. In: Proceedings of the IFIP 17th World Computer Congress – TC12 Stream on Intelligent Information Processing, pp. 51–63

[Mädche, A. et al. 2003] Managing multiple and distributed Ontologies on the Semantic Web. The VLDB Journal – The International Journal on Very Large Data Bases 12(4), 286–302

[Noy, N. F. et al. 2006] A framework for ontology evolution in collaborative environments. In: Proceedings of the 2005 International Semantic Web Conference (ISWC 2005), pp. 544–558 (2005)

[Pearl, J. 1983] Heuristics: Intelligent search strategies for computer problem solving. Addison-Wesley

[Stojanovic, L. et al. 2002] Stojanovic, L., Maedche, A., Motik, B., Stojanovic, N.: User-driven ontology evolution management. In: Gómez-Pérez, A., Benjamins, V.R. (eds.) EKAW 2002. LNCS (LNAI), vol. 2473, pp. 285–300. Springer, Heidelberg (2002)

[Stojanovic, N. et al. 2003] Stojanovic, L., Stojanovic, N., González, J., Studer, R.: OntoManager – A system for the usage-based ontology management. In: Chung, S., Schmidt, D.C. (eds.) CoopIS 2003, DOA 2003, and ODBASE 2003. LNCS, vol. 2888, pp. 858–875. Springer, Heidelberg (2003)

[Suárez-Figueroa, M. C. and Gómez-Pérez, A. 2008] Towards a glossary of activities in the ontology engineering field. In: Proceedings of the Sixth International Conference on Language Resources and Evaluation, LREC 2008 (2008)

[Zablith, F. 2009] Zablith, F.: Evolva: A comprehensive approach to ontology evolution. In: Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvönen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E. (eds.) ESWC 2009. LNCS, vol. 5554, pp. 944–948. Springer, Heidelberg (2009)

Modeling the Economic Saturation

Stasys Girdzijauskas, Andžela Mialik, Erika Jociūtė, and Neringa Zujytė

Vilnius University, Kaunas Faculty of Humanities, Department of Informatics,
Muitines st. 8, LT-44280, Kaunas, Lithuania
{stasys.girdzijauskas,neringa.zujyte}@vukhf.lt,
{andzelina.m,jociute.erika}@gmail.com

Abstract. The analysis of cyclic processes in the financial markets has drawn attention to the possibility of market saturation. It should be noted that overproduction is carried out in open markets, and saturation – in closed markets. Due to saturation, **profitability** of investing in closed markets increases and thus price bubbles are formed. The increase in profitability and emergence of price bubbles are forms of expression of a new economic paradox. The analysis of saturated economic systems started only in the last decade. This article contains modeling of saturated economic systems, supported by the newly-created "Logistic analysis tool". The tool is based on the model of logistic interest, developed in the last decade (S. Girdzijauskas, 2002).

Keywords: Market capacity, market gap, real market, market saturation, saturation, paradox, logistical model, economic growth, bubbles.

1 Introduction

Modern world agrees on pursuing economic welfare. Most experts (but not all) agree that economic welfare is not possible without economic growth. But it is impossible to achieve constant, unstoppable in time (endless) economic growth. “Everything that rises must converge” was rightly pointed out by Flannery O’Connor [14]. It is not hard to realize, that nothing can grow endlessly. Thus, if economy is growing, and endless growth is impossible, then growth must stop at some point of the time and this automatically means a crisis [3,8,12]. From this a conclusion can be derived that the main cause of economic crises lies in its growth.

Relevance of the topic. In the economic growth, as in every part of nature or society, thought, the externally visible chaos and chains of coincidence, the economic patterns manifest. We must agree that individual economic processes as well as the whole economic growth are generally controlled by the inner laws. Everything has its own cause and everyone’s responsibility is to try to find it, wrote Ralph Nelson Elliot (1871-1948) a researcher of economic waves. [9,12]

Limited growth is defined by a logistic function. Such growth is widely used in researching saturation of populations (biological and etc.) [14].

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 178–183, 2011.

So far we have modelled the economic situations by using MS Excel or "Loglet Lab" programme that was created and presented by New York's Rockefeller University's students in 1999. Since the "Loglet Lab" tool uses a logistic function which is oriented for the analysis of several biological processes, it was decided to create and design the economic logistic analysis tool [2]. The core of the new tool is – a mathematical-economical logistic interest model (S.Girdzijauskas, 2002) [6]. It represented the economic saturation by using MS Excel and the new tool, of which the advantages, possibilities and functionality are described in the article.

The purpose of the topic is to introduce a logistic software tool for simulating the growth of the economic processes by using logistic interest function.

The main goals of the article are:

- Present the logistic capital growth model;
- Specify its importance for the market economy;
- Represent the advantages and opportunities of the "Logistic analysis tool".

2 Capital Growth Model Market Saturation

In order to find out the main causes that drive the growth, we have to understand the logistic growth model before simulated the growth [3,6]:

$$K = \frac{K_p \cdot K_0 (1+i)^n}{(K_p - K_0) + K_0 (1+i)^n} \tag{1}$$

In these formulas: K – population (capital size) after n periods; K_p – maximal (potential) capital value (market capacity); K_0 – initial capital value (initial investment); i – interest rate.

Market saturation can be understood as an amount of capital that could be effectively absorbed in investments. The main characteristic of market saturation is the potential capital. Market saturation is the biggest theoretically possible sales amount of products or services that could be achieved by all companies at a certain time. Very often, producing and selling more and more goods or services of the same market, the market is gradually filled and market saturation is reached. The term '*saturated'* in most cases means *fullness*, *filling up to a certain limit*, *absolute loading* [4,7].

All the market capacity could be understood as consisting of two parts: a) the part, which is already filled and b) the part which is not filled yet. Let us call the already filled market the real market, and the unfilled market – a market niche. Then, the market capacity will be equal to the real market plus market niche, i.e.

In such a way, semi-closed and especially closed markets can become saturated increasing investments. Meanwhile open markets cannot be saturated (it is only possible to have an overproduction). Hints of saturation are increasingly common in the world [5,10,14]. Unfortunately, they cover only the outside – a superficial half of the saturation.

During the last decade, deeper research in Lithuania has exceeded its expectations – it has become clear that saturation does not only exist in financial areas but it occurs in a paradoxical way and thus, it is one of the major factors that cause economic cataclysms.

3 "Logistic Analysis Tool" – The Price Bubble Simulation

There is a lot of different software tools used for economic analysis, such as: *Matlab, Loglet Lab, Excel, MS* and others, by their functionality and user interface but none of them meets the requirements for the smooth and logistic function-based modeling. For this reason, it was decided to propose the "logistic analysis tool" to solve this problem.

The purpose of designing the "Logistic analysis tool" is to create a software tool for the analysis of the selected logistic function. This tool is intended to provide the user with the possibility to import and export processed data. The user is also able to apply the "Logistic analysis tool" in a virtual area and the results are presented in a graphical form. With the help of this tool, the time distributed data is aimed at to be analyze: the first step is to divide the main growth process into subprocesses (with the help of a software code) in order to determine the limits of growth, then, according to the graphic form, allow the definition of the entire system's behavior.

3.1 The Algorithm and Dynamics of the Software Tool

The algorithm of the created logistic tool is presented in figure 1. The main steps for simulation of economical situation are [13]:

1. The user enters the data;
2. The data are processed by the system;
3. The economic logistic model is being simulated;
4. The graph of simulation is presented for the user;
5. The user has the option to adjust the growth at any time of simulation.

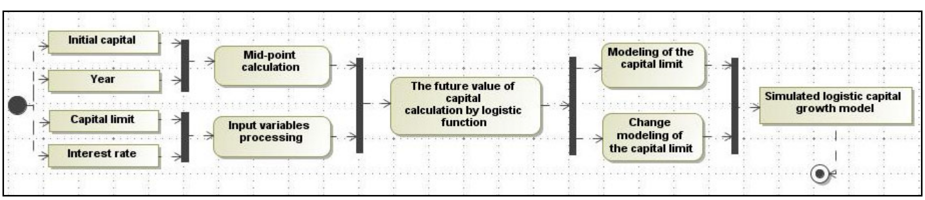


Fig. 1. The algorithm of "Logistic analysis tool"

In order to design the logistic economical model, we need to define four variables of the logistic function. These variables are: initial capital, period, capital limit and interest rate. After the introduction of these data, the mid-point is calculated and, with the help of the economic-logistic function, the output variables are processed. The type of analysis of variables can be chosen by the user of a logistic tool. The user can decide whether he wants to simulate only the future capital value (threshold) or the scope of change of this threshold [13].

Final result of the tool – a simulated logistic capital growth model, which is reproduced graphically (figure 2).

3.2 The Realization of the “Logistic Analysis Tool”

It is decided to use software realization tools for “Logistic analysis tool” [13]:

- PHP 5.2.17. programming language;
- Control Panel "DirectAdmin", based on Linux (tool can be reached at: http://lai.tefno.net/);
- Notepad + +, for PHP programming language;
- Internet Explorer 8.0, Opera, Mozilla Firefox browsers for review of “Logistic analysis tool” outcome.

Consider the simplified case where all the cash flow consists of only two members: the amount of investment and the only income member. In addition, the investment should be equal to one cash unit (c.u.), and the income received after one period - equal to 1.1 c.u.

Then let us make a table when $K = 1.1$ (which corresponds to an annual interest rate equal to 10%), and $K_p > 1.1$. Express the degree of market filling by ratio of K / Kp.

Table 1.

K_p	10^8	11	5.5	3.67	2.75	2.20	1.83	1.57	1.38	1.22	1.10
K/K_p	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99

The table shows only the standard points. Actually, it can be taken more in the function's intensive change area.

The graph in the 2nd “Logistic analysis tool” picture shows the tool's operation, capabilities and functionality. There are modeling a clear dependence on the profitability of market saturation rates: **with increasing saturation – increased profitability**. This is the paradoxical behavior of the market. Taking deeper consideration, we notice that the overproduction is accompanied by a decline in profitability only to open markets, where there is no saturation. Meanwhile, in partially or fully closed (and saturated) markets, overproduction is accompanied by growth in profitability, which in turn triggers the stock market. The latter further enhances overproduction. In this way price bubbles occur. When the bubble bursts, recession or crisis is expected. [1,2,3,11].

The user can easily understand the dynamics of the “Logistic analysis tool” from the input of data to graph drawing. With the help of this tool, each user can analyze and simulate the economic paradoxes at any time.

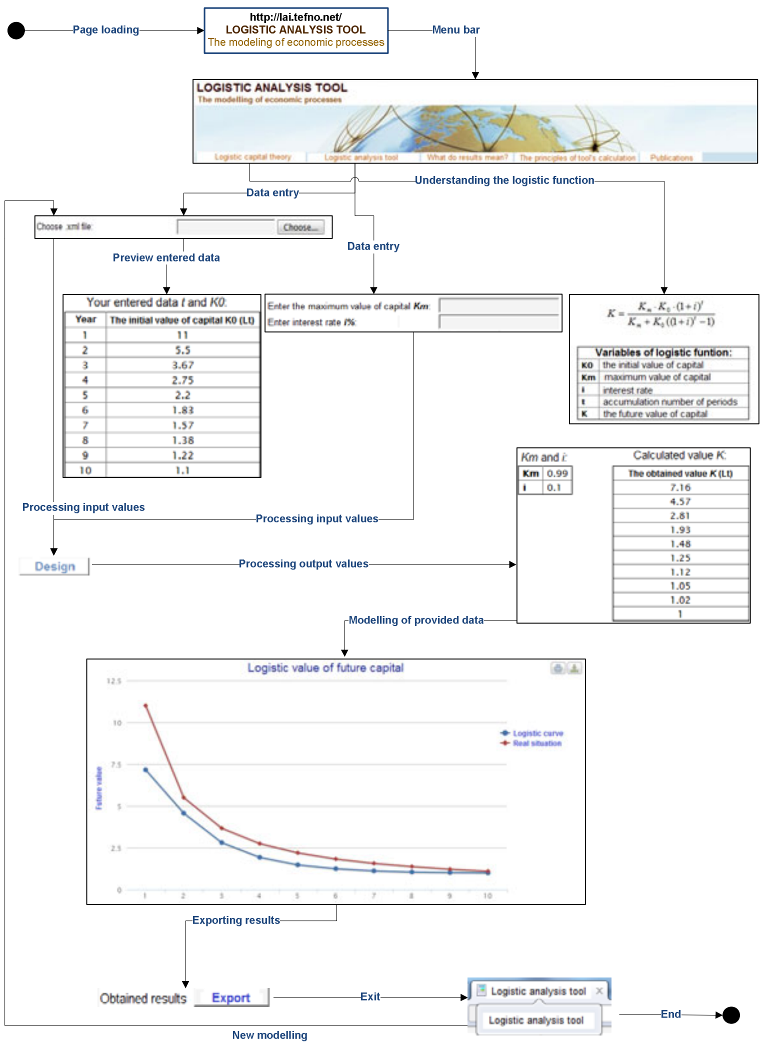


Fig. 2. "Logistic analysis tool"

4 Conclusions

- Market capacity, saturation effects which influence modern economic development ere presented. The analyzed example allows the reader to understand the advantages and opportunities of logistic growth. Since that time, it is particularly important to examine the details of these paradoxes and their linkages with market capacity, excess liquidity, economic bubbles' formation, cyclical evolution of economic development, overproduction, inflation, stagnation, unemployment and other economic phenomena.

- Price modeling has been presented with the help of the "Logistic analysis tool". The algorithm of the tool shows the modeling of economic processes in a clear way and is an opportunity for data analysis.

References

1. Dubnikovas, M., Moskaliova, V., Girdzijauskas, S.: Analysis of the Share Price Burbles in the Baltic Countries. In: van der Aalst, W., Mylopoulos, J., Sadeh, N.M., Shaw, M.J., Szyperski, C. (eds.). LNBIP, pp. 119–129. Springer, Heidelberg (2009) ISSN: 1865-1348
2. Girdzijauskas, S.A., Štreimikienė, D., Dubnikovas, M.: Analyzing Banking Capital with LOGLET Lab Software Package. Transformations in Business and Economics 8(2-17), 45–56 (2009)
3. Girdzijauskas, S., Dubnikovas, M.: Logistic Analysis of Price Bubble and Current Situation in USA Real Estate. In: Abramowicz, W., Tolksdorf, R., Węcel, K. (eds.) BIS 2010. LNBIP, vol. 57, pp. 13–18. Springer, Heidelberg (2010)
4. Knyvienė, I., Girdzijauskas, S., Grundey, D.: Market Capacity from the Viewpoint of Logistic Analysis. Technological and Economic Development of Economy 16(4), 690–702 (2010) ISSN 2029-4913
5. Sornette, D.: Why Stock Markets Crash: Critical Events in Complex Financial Systems. Princeton University Press (2003)
6. Girdzijauskas, S.: Logistinis kritinių situacijų valdymas; burbulų fenomenas arba antrasis kvėpavimas. Ekonomika ir vadyba; Vilnius: Vilniaus universiteto leidykla 1(7), 53–59 (2004) ISSN 1392-8732
7. Burns, A., Mitchell, W.: Measuring Business Cycles. National of Economic Research, N.Y (1946)
8. Juchem Neto, P., Claeyssen, J.C.R., Ritelli, D.: G. Mingari Scarpello. Migration in a Solow Growth Model (2009), `http://ssrn.com/abstract=1578565` (Žiūrėta: 2011 sausio 15d.)
9. Girdzijauskas, S., Štreimikienė, D.: Application of Logistic Models for Stock Market Bubbles Analysis. Journal of Business Economics and Management 10(1), 45–51 (2009) ISSN 1611-1699
10. Modis, T.: The End of the Internet Rush. Technological Forecasting & Social Change 72, 940–941 (2005)
11. Girdzijauskas, S., Štreimikienė, D., Čepinskis, J., Moskaliova, V., Jurkonytė, E., Mackevičius, R.: Formation of Economic Bubbles: Causes and Possible Preventions. Technological and Economic Development of Economy; Baltic Joulnar on Sustainability 15(2), 267–280 (2009) ISSN 1392-8619
12. Poser, S.W.: Applying Elliott Wave Theory Profitably, p. 256. John Wiley & Sons, Ltd. (2003) ISBN 0-471-42007-7
13. Zujytė, N., Jociūtė, E., Mialik, A., Slanys, S.: Logistinės analizės įrankio projektavimas. In: 16-oji tarpuniversitetinė magistrantų ir doktorantų konferencija. Technologija, Kaunas, p. 140 (2011) ISSN 2029-249X
14. Sterman, J.D.: Business Dynamics: Systems Thinking and Modeling for a Complex World, p. 950 (2000) ISBN 0-07-231135-5

Product-Service System Configuration in SOA-Based Environment

Nikolay Shilov

SPIIRAS, 39, 14 line,
199178 St.Petersburg, Russia
nick@iias.spb.su

Abstract. Requirements from rapidly developing global markets force companies to develop new paradigms of business, such as product-service systems. Such paradigms, in turn, require new solutions from Information and Communication Technologies (ICT), and modern ICT do make it possible to develop such new approaches. The paper presents an approach based on efficient management of information services in the open information environment oriented to product-service system configuration. The approach is based on the technologies of ontology and context management. The standards of Web-services are used to provide for interoperability between information services. Application of constraints for knowledge representation makes it possible to integrate with existing services.

Keywords: Service-oriented architecture, ontology, context, product-service system.

1 Introduction

Current trends in the worldwide economy require companies to implement new production and marketing paradigms. This determines major trends of knowledge-dominated economy: (i) shift from “capital-intensive business environment” to “intelligence-intensive business environment” – an “e” mindset – and (ii) shift from “product push” strategies to a “consumer pull” management – mass customisation approach [1].

Product-Service Systems (PSS) assumes orientation on combination of products and services (often supporting the products) instead of focusing only on products. This is a relatively new paradigm that fits well, for example, industrial equipment manufacturers, for which the equipment maintenance is a considerable part of the business. Therefore, tight relationships with customers are of high importance in such cases.

A strategy that brings companies and their customers in a closer collaboration is innovation democratisation. This is a relatively new term standing for involvement of customers into the process of designing and creating new products and services. This makes it possible for companies to better meet the needs of their customers. [2]

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 184–195, 2011.

For companies with wide assortments of products (more than 30 000 – 40 000 products of approx. 700 types, with various configuration possibilities), it is very important to ensure that customers can easily navigate among them to define needed services. One possible solution is to provide a codification system that can produce easily recognizable and at the same time relatively short codes. This is an important task for customer communication management because well defined and understandable product identification is mandatory for successful collaboration with customers and for ensuring a good corporate look for the company [3-5].

New information technologies open new boundaries for researchers. One of new possibilities is remote usage of information services [6]. The service-oriented architecture (SOA) is a step towards information-driven collaboration. This term today is closely related to other terms such as ubiquitous computing, pervasive computing, smart space and similar, which significantly overlap each other [7].

This paper presents an approach to efficient management of information services in the open information environment for PSS configuration. For this purpose the information is actualized in accordance with the current situation. An ontological model is used in the approach to solve the problem of service heterogeneity. This model makes it possible to enable interoperability between heterogeneous information services due to provision of their common semantics [8]. Application of the context model makes it possible to reduce the amount of information to be processed. This model enables management of information relevant for the current situation [9]. The access to the services, information acquisition, transfer, and processing (including integration) are performed via usage of the technology of Web-services.

Fig. 1 represents the generic scheme of the approach. The main idea of the approach is to represent the product's components by sets of services provided by them. This makes it possible to replace the configuration of PSS with that of distributed services. For the purpose of interoperability the services are represented by Web-services using the common notation described by the application ontology (AO). Depending on the problem considered the relevant part of AO is selected forming the abstract context that, in turn, is filled with values from the sources resulting in the operational context. The operational context represents the constraint satisfaction problem that is used during self-configuration of services for problem solving.

Some elements of the presented approach have been implemented in an industrial company that has more than 300 000 customers in 176 countries supported by more than 52 companies worldwide with more than 250 branch offices and authorised agencies in further 36 countries.

The remainder of the paper is organised as follows. Section 2 describes use cases for the PSS-oriented company. The major components of the approach are described in sections 3 (ontological knowledge representation and formalism of object-oriented constraint networks) and 4 (abstract and operational contexts). The procedure of alignment of Web-service descriptions and the application is explained in section 5. Section 6 summarises main research features of the approach.

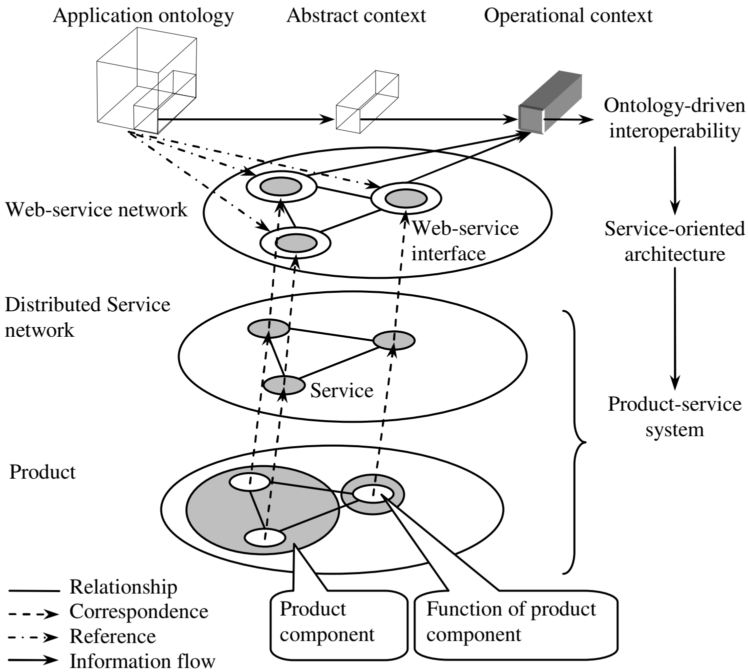


Fig. 1. Generic scheme of the approach

2 Ontology-Driven Use Cases for Industrial Equipment Manufacturer

Ontologies have shown their usability for this type of tasks [e.g., 10-12]. On the basis of the formal description of possible products within the common ontology it is possible now to design new applications which offer customers better ways to find and choose the right products and services.

A simple but always necessary kind of relationship between properties and values describes the consistency of a complex product. This is mainly done by constraints restricting the set of all possible combinations to those which are possible in real-life. The reasons for applying constraints can be different – the most common is the technical possibility of a certain combination.

Furthermore it is possible to add dependant technical data to a certain configuration (which is a set of selected properties and values). For example, a product's weight can be calculated based on the properties / values selected by customer. Another common use case is to configure a CAD 3D model by sending its constructive relevant information from the order code. Practically a lot of data can be made dependant on the current configuration of a modular product. This provides a possibility to provide data which is similarly exact to data of discrete products (for example with a fixed weight).

Even more challenging are inter-product relationships. The most common use case is the relationship between a main product and an accessory product. While both products are derived from a different complex modular product model there are dependencies which assign a correct accessory to a configured main product. Those dependencies are related to the products individual properties and values. The depth of product-accessory relationships is basically not limited, so accessory-of-accessory combinations have to be taken into account, too. Certain problems have to be eliminated like circular relationships which lead back to main product. The relationships can be very complex when it comes to define the actual location / orientation of interfaces and mounting points between products.

A more complex scenario is solution-oriented. The idea is to solve a certain real-life problem with modular-products and their inter-product relationships. The result of such a solution is basically a system of products working together. Used formalism of Object-Oriented Constraint Networks (OOCN, see sec. 3) makes it possible to develop services performing automatic definition of configurable complex products based on the required functions and other constraints specified by the customer.

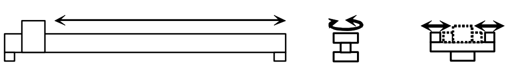
A handling module offers a good example for this problem. For this purpose the ontology is extended with an extra attribute “Function” for classes.

For illustrative purposes a simple example consisting of three simple products is considered:

- simple product function: movement (Fig. 2a)
- simple product function: rotation (Fig. 2b)
- simple product function: gripping (Fig. 2c)

The goal of the example is to provide for a service to configure complex products that can perform predefined functions. For instance, if two functions (movement and gripping) are required, then the resulting product will consist of two simple products (Fig. 3). If all three functions are required, the resulting product will be as shown in Fig. 4. Such a configuration can be calculated automatically by a constraint solver. Principles of the configuration of more complex constructions are planned to be researched.

The compatibility table for these products is presented in Table 1.



a) b) c)

Fig. 2. Simple products: (a) movement function, (b) rotation function, (c) gripping function

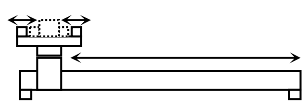


Fig. 3. Complex product performing two functions (movement and gripping)

Table 1. Compatibility table

	Product 1	Product 2	Product 3
Product 1	-	+	+
Product 2	-	-	+
Product 3	-	-	-

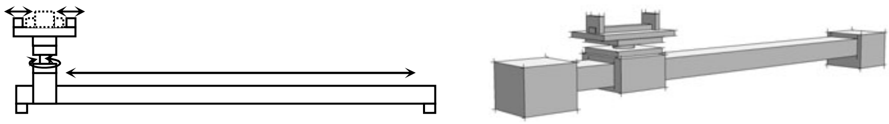


Fig. 4. Complex product performing three functions (movement, rotation and gripping)

3 Ontological Knowledge Representation

In the approach the ontological model is described using the formalism of Object-Oriented Constraint Networks (OOCN). Application of constraint networks allows simplifying the formulation and interpretation of real-world problems which in the areas of management, engineering, manufacturing, etc. are usually presented as constraint satisfaction problems [13]. This formalism supports declarative representation, efficiency of dynamic constraint solving, as well as problem modelling capability, maintainability, reusability, and extensibility of the object-oriented technology.

In the presented methodology the PSS is supposed to be interpreted as a dynamic constraint satisfaction problem (CSP). OOCN provides compatibility of ontology model for knowledge representation and internal solver representations. As a result, ontology-based problem model is described by a set of constraints and can be directly mapped into the constraint solver. A result of CSP solving is one or more satisfactory solutions for the problem modelled.

Compatibility of CSP, ontology, and OOCN models is achieved through identification of correspondences between primitives of these models. CSP model consists of three parts: (i) a set of variables; (ii) a set of possible values for each variable (its domain); and (iii) a set of constraints restricting the values that the variables can simultaneously take. Typical ontology modelling primitives are classes, relations, functions, and axioms. The formalism of OOCN describes knowledge by sets of classes, class attributes, attribute domains, and constraints. Concept "class" in OOCN notation is introduced instead of concept "object" in the way object-oriented languages suggest.

The OOCN paradigm (the detailed description can be found in [14]) defines the common ontology notation used in the system. According to this representation an ontology (A) is defined as: $A = (O, Q, D, C)$ where: O – a set of *object classes* ("*classes*"); each of the entities in a class is considered as an *instance* of the class. Q – a set of class attributes ("*attributes*"). D – a set of attribute domains ("*domains*"). C – a set of *constraints*.

For the chosen notation the following six types of constraints have been defined $C = C^I \cup C^{II} \cup C^{III} \cup C^{IV} \cup C^V \cup C^{VI}$: $C^I = \{c^I\}$, $c^I = (o, q)$, $o \in O$, $q \in Q$ – accessory of attributes to classes; $C^{II} = \{c^{II}\}$, $c^{II} = (o, q, d)$, $o \in O$, $q \in Q$, $d \in D$ – accessory of domains to attributes;

$C^{III} = \{c^{III}\}$, $c^{III} = (\{o\}, True \vee False)$, $|\{o\}| \geq 2$, $o \in O$ – classes compatibility (compatibility structural constraints); $C^{IV} = \{c^{IV}\}$, $c^{IV} = \langle o', o'', type \rangle$, $o' \in O$, $o'' \in O$, $o' \neq o''$ – hierarchical relationships (hierarchical structural constraints) "is a" defining class taxonomy ($type$=0), and "has part"/"part of" defining class hierarchy ($type$=1);$C^V = \{c^V\}$, $c^V = (\{o\})$, $|\{o\}| \geq 2$, $o \in O$ – associative relationships ("one-level" structural constraints); $C^{VI} = \{c^{VI}\}$, $c^{VI} = f(\{o\}, \{o, q\}) = True \vee False$, $|\{o\}| \geq 0$, $|\{q\}| \geq 0$, $o \in O$, $q \in Q$ – functional constraints referring to the names of classes and attributes.

Correspondences between the primitives of ontology model, OOCN, and CSP is shown in Table 2.

Table 2. Correspondence between ontology model, OOCN, and CSP model

Ontology Model	OOCN	CSP
Class	Object	Set of variables
Attribute	Variable	
Attribute domain (range)	Domain	Domain
Axiom / relation	Constraint	Constraint

Below, some example constraints are given:

- the attribute *Locking in end positions* (q_1) belongs to the class *Series C (pneumatic drive)* (o_1): $c^I_1 = (o_1, q_1)$;
- the attribute *Locking in end positions* (q_1) belonging to the class *Series C* (o_1) may take the values *Without (Standard)*, *Extend / Retract*, *Extend*, and *Retract* (the explanation of the values is given in sec. 4): $c^{II}_1 = (o_1, q_1,$ {*Without (Standard)*; *Extend / Retract*; *Extend*; and *Retract*});
- the class *Valve* (o_2) is compatible with the class *Series C* (o_1): $c^{III}_1 = (\{o_1, o_2\}, True)$;
- an instance of the class *Valve* (o_2) can be a part of an instance of the class *Valve terminal* (o_3): $c^{IV}_1 = \langle o_2, o_3, 1 \rangle$;
- the Series C (o_3) is a Pneumatic Drive (o_4): $c^{IV}_1 = \langle o_3, o_4, 0 \rangle$;
- an instance of the class *Valve* (o_2) can be connected to an instance of the class *Series C* (o_1): $c^V_1 = (o_2, o_1)$;
- the value of the attribute *cost* (q_2) of an instance of the class *solution* (o_5) depends on the values of the attribute *cost* (q_2) of instances of the class *component* (o_6) connected to that instance of the class *solution* and on the number of such instances: $c^{VI}_1 = f(\{o_6\}, \{(o_5, q_2), (o_6, q_2)\})$.

To summarize, the general three-level scheme of the approach is shown in Fig. 5. The technological base is provided by Web-services. The problem is described via object-oriented constraint networks and the semantics is provided for by using ontologies. Semantic level provides for knowledge sharing and exchange in the PSS model.

Semantic level	Ontology
Problem description level	Constraints
Technological level	Web services

Fig. 5. Three-level scheme of the proposed approach

4 Contexts

For modelling the current state of PSS (current situation) two types of contexts are used: abstract and operational. The abstract context is an ontological model of the current situation build on the basis of selection of knowledge relevant to the current situation. The operational context is a specification of the abstract context for the particular real world situation.

In accordance with the chosen formalism the modelling of the current situation can be defined as the following task. Based on the formalized knowledge about the problem domain $A = (O,Q,D,C)$, build two-level model of the current situation S, presented via the abstract context $Context_a(T_a) = (S|O_a,Q_a,D_a,C_a,R_a,T_a)$ ($O_a \subseteq O$, $Q_a \subseteq Q$, $D_a \subseteq D$, $C_a \subseteq C$, $R_a \subseteq R$, where T_a is the time of the model adequacy) and operational context $Context_p(t) = (S|O_p,Q_p,D_p,C_p,R_p,t)$ ($O_p \subseteq O_a$, $Q_p \subseteq Q_a$, $D_p \subseteq D_a$, $C_p \subseteq C_a$, $R_p \subseteq R_a$, where t is the current time), and find the set of the object instances $I(t) = \{v_{q_i}^o | v \in D_p, o \in O_p, q \in Q_p, i = 1,\ldots,N_q\}$, such that all constraints C_p hold, where N_q is the number of attributes of class o, with the parameter values being acquired from the services R of the open information environment at time *t*.

The ontological model of the abstract context is defined as:

$$Context_a(T_a) = (S|O_a,Q_a,D_a,C_a,WS_a,T_a),$$

where S is the time of the situation modelled;

O_a is a set of classes, generally required for modelling the situation S;

Q_a is a set of attributes of the classes O_a;

D_a is a set of domains of the attributes Q_a;

C_a is a set of constraints included into the abstract context;

WS_a is a set of Web-services representing services of the open information environment assigning the values to the attributes Q_a, $WS_a \subseteq WS$, where WS is a set of registered Web-services.

T_a is an estimated time of the model adequacy.

When the information of the open information environment becomes available from the services, references to which are stored in the abstract context, the appropriate values are assigned to the attributes of the classes of the abstract context.

Thus, the operational context is getting built. The operational context $Context_p$ is the model of the current situation in the notation of the object-oriented constraint networks with values assigned to the variables. This model is interpreted as the constraint satisfaction task. The model of the operational context is represented as:

$$Context_p(t) = \left(S \middle| O_p, Q_p, D_p, C_p, WS_p, T_a, \Delta T\right),$$

where t is the current time,

O_p is a set of classes used for the modelling of the situation S in particular conditions,

Q_p, D_p, C_p, WS_p are used sets of attributes, domains, constraints and Web-services respectively;

$\Delta T = t \quad t_0$ is the current time of the operational context life span, where t_0 is the creation time of the abstract context.

The constraint satisfaction task is a triple of sets: $CSP = (V_{CSP}, D_{CSP}, C_{CSP})$,

where V_{CSP} is a set of variables,

D_{CSP} is a set of corresponding domains of the values of the variable,

C_{CSP} is a set of constraints.

The solution of the constraint satisfaction task is a set of values of the variables $v \in V_{CSP}$, such that all the constraints hold. This set contains sets of object instances $I(t) = \left\{ v^o_{q_i} \middle| v \in D_p, o \in O_p, q \in Q_p, i = 1,\dots,N_q \right\}$.

To provide for the information exchange between heterogeneous services as well as its processing, the service model has been developed that is compatible with the ontological model of the knowledge representation. The compatibility is achieved via usage of Web-service interfaces for the services of the open information environment.

5 Alignment of Web-Service Descriptions and the Application Ontology

Illustration of the approach starts with a preliminary phase at that the Web-service descriptions and the AO are aligned. The alignment operation is based on discovering attributes occurred in Web-service descriptions, values of which can assign values to properties of the AO classes. It is supposed that an AO-property can take on a value provided by an attribute the name of which is semantically close to the name of this property. For discovering semantically close names a measure of semantic distance is used.

For the purpose of measuring the semantic distances between concepts containing in the Web-service descriptions and in the AO a machine readable dictionary [15] extracted from Wiktionary [16] is used. Wiktionary was chosen by reasons of its free use, its multilingual support, and keeping, besides lexical relations, definitions of words. The extracted machine-readable dictionary includes 1) a set of words defined in Wiktionary along with for each word 2) definitions given for this word, 3) a set of synonyms, if any, and 4) a set of associated words. Words associated to a word are considered the hyperlinked words occurring in the Wiktionary definition given for this word.

The AO is represented as a semantic network where names of classes and properties specified in the AO constitute nodes of the network. The nodes corresponding to the AO concepts are linked to nodes representing their synonyms and associated words as this is given in the machine-readable dictionary. The links between the nodes are labelled by the weights of relations specified between the concepts represented by these nodes in the machine-readable dictionary. Weight w of a relation specified between two concepts t_i and t_j is assigned as:

$$w = \begin{cases} 0{,}5 & -t_i, t_j \text{ are synonyms} \\ 0{,}3 & -t_i, t_j \text{ are associated words} \\ \infty & -t_i, t_j \text{ are the same word} \end{cases}$$

The values for the weights were evaluated based on the following principles.

Weights for the synonyms are assumed to be greater than weights for the associated words;

Semantic distance is proposed to be calculated as inversely proportional to weights raised to a power. The power corresponds to the path between the compared words. The longer the path the greater the semantic distance for the two different words is expected to be. To meet this expectation with reference to the way of the semantic distance calculation, a weight of the relation between two different words should be in the range (0, 1). Taken into account the first principle the weights 0,5 for the relation between the synonyms and 0,3 for the relation between the associated words are chosen empirically;

The semantic distance between the same words is equal to 0. Correspondingly, ∞ is assigned to the weight of the relation between the same words.

The first step in the alignment operation is parsing a Web-service description represented by the Web Service Definition Language (WSDL). The result of the parsing is a set of meaningful words found in the attribute values of WSDL-tags. If this set contains words differing from the nodes of the semantic network built for the AO, the semantic network is extended with the nodes representing the words extracted from the WSDL-file, synonyms for these words from the machine-readable dictionary, words associated in the machine-readable dictionary with the extracted words, and appropriate links. Only those synonyms and associated words for the WSDL-words are added in the semantic network, which differ from the concept the network already represents. If the semantic network built for the AO contains words found in the WSDL-file, the same words in the AO and in the WSDL file are linked by the relation labelled "∞".

Next, nodes representing WSDL words are checked for their similarity to nodes representing AO-concepts. As a measure of similarity semantic distance *Dist* (2) is used.

$$Dist(t_i, t_j) = \frac{1}{\sum_{S} \prod_{k=s_i}^{s_j} w_k}$$

where t_i - WSDL word, t_j – AO-concept; w – weight of lexical relation existing between t_i and t_j; S – a set of paths from t_i to t_j, where a path s is formed by any number of links that connect t_i and t_j passing through any number of nodes.

After the semantic distances between the words from the WSDL-file and the names specified in the AO are calculated, experts are provided with a ranked list of semantically similar words for each word found in the WSDL-file. Based on this list the experts align related, in their judgment, attribute values in the Web-service descriptions and class properties that can take these values.

Let's consider the following example. The service for finding equipment that meets certain processing requirements has (in its WSDL description) an attribute "*force*" that describes how heavy the processed part can be. In the AO there are two attributes that might correspond to it: "*size*" and "*weight*". An illustrative piece of the semantic network built based on Wiktionary processing and the formula for calculating weights of relationships is represented in Fig. 6.

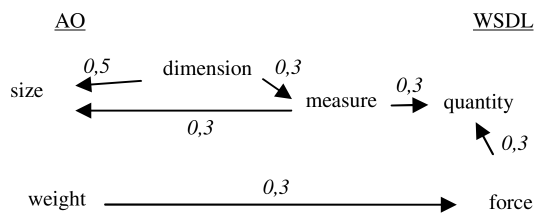


Fig. 6. A piece of the semantic network relevant to WSDL attribute "*force*"

The path from the WSDL word "*force*" to the AO word "*weight"* is "*weight*" → "*force*" (weight 0.3). Thus, semantic distance between the two words is calculated as:

$$Dist(\, force, weight\,) = \frac{1}{0.3} = 3.33\,.$$

The set of paths from the WSDL word "*force*" to the AO word "*size"* in Fig. 6 comprises of two paths and the semantic distance is

$$Dist(\, force, size\,) = \frac{1}{0.3^3 + 0.3^3 \bullet 0{,}5} = 24.69\,.$$

It can be seen that the distance between the concepts "*force*" and "*weight"* is much shorter than between the concepts "*force*" and "*size"*. Hence, the attribute "*weight"* of the AO and the attribute "*force*" contained in the Web-service description can be aligned.

6 Conclusions

The paper presents an approach to integration of efficient management of information services in the open information environment for PSS configuration. The major idea of the approach is product representation via a set of services provided by the product's components. The formalism of object-oriented constraint networks used in the approach makes it possible to represent the problem domain in the approach is described via an ontology with links to the services that provide required information. For each particular situation a fragment of the ontology relevant to the situation is built (the abstract context) and complemented with particular values from the services of the open information environment (the operational context). The operational context, in turn, can be used for problem solving as a constraint satisfaction problem.

The presented codification system significantly simplifies and speeds up the process of product code development in the company producing industrial equipment. Generating a new order code for a new product with the help of the developed system takes approximately one day. The technical options presented by the product manager and developer are converted into order-relevant options. As most of the characteristics can be used again, only new options must be discussed and entered in the system. Without the system this process would need several days. Besides this, the error risk would be very large. It could happen that for the same option another code letter is used for example.

The major advantages of the developed system are:

- Systematic order codes for all products;
- Machine readability;
- Quick orientation for selecting right products and services;
- Security when selecting and ordering products and services.

One other advantage is the reusability of the data. The structured data are used in other processes such as:

- Automatically creating master data in SAP models;
- Automatically creating data for the configuration models and services;
- Automatically generating an ordering sheet for the print documentation (this ordering sheet was generated earlier with high expenditure manually);
- Automatically generating a product and service list which is needed in the complete process implementing new products.

Acknowledgments. The presented work is a result of the joint project between SPIIRAS and Festo Ag&Co KG. Some parts of the research were carried out under projects "COBIT" funded by The Swedish Foundation for International Cooperation in Research and Higher Education, # 09-07-00436 and # 10-07-00368 funded by the Russian Foundation for Basic Research, and project #213 of the research program "Intelligent information technologies, mathematical modelling, system analysis and automation" of the Russian Academy of Sciences.

References

1. Smirnov, A., Pashkin, M., Chilov, N., Levashova, T.: Knowledge Fusion in the Business Information Environment for e-Manufacturing Pursuing Mass Customisation. In: Rautenstrauch, C., Seelmann-Eggebert, R., Turowski, K. (eds.) Moving into Mass Customization. Information Systems and Management Principles, pp. 153–175. Springer, Heidelberg (2002)

2. von Hippel, E.: Democratizing innovation. The MIT Press, Cambridge (2006)
3. Baumeister, H.: Customer relationship management for SMEs. In: Proceedings of the 2nd Annual Conference eBusiness and eWork e2002, Prague, Czech Republic (2002)
4. Fjermestad, J., Romano Jr., N.C.: An Integrative Implementation Framework for Electronic Customer Relationship Management: Revisiting the General Principles of Usability and Resistance. In: Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS 2003), Big Island, HI, USA (2003)
5. Piller, F., Schaller, C.: Individualization Based Collaborative Customer Relationship Management: Motives, Structures, and Modes of Collaboration for Mass Customization and CRM, Working Paper No. 29 of the Dept. of General and Industrial Management, Technische Universität München (2004)
6. Johanesson, P., Andersson, B., Bergoltz, M., Weigand, H.: Enterprise Modelling for Value Based Service Analysis. In: Stirna, J., Persson, A. (eds.) The Practice of Enterprise Modelling, First IFIP WG 8.1 Working Conference, PoEM 2008. LNBIP, vol. 15, pp. 153–167. Springer, Heidelberg (2008)
7. Balandin, S., Moltchanov, D., Koucheryavy, Y. (eds.): ruSMART 2009. LNCS, vol. 5764. Springer, Heidelberg (2009)
8. Uschold, M., Grüninger, M.: Ontologies: Principles, methods and applications. Knowledge Engineering Review 11(2), 93–155 (1996)
9. Dey, A.K.: Understanding and Using Context. Personal and Ubiquitous Computing J. 5(1), 4–7 (2001)
10. Bradfield, D.J., Gao, J.X., Soltan, H.: A Metaknowledge Approach to Facilitate Knowledge Sharing in the Global Product Development Process. Computer-Aided Design & Applications 4(1-4), 519–528 (2007)
11. Chan, E.C.K., Yu, K.M.: A framework of ontology-enabled product knowledge management. International Journal of Product Development 4(3-4), 241–254 (2007)
12. Patil, L., Dutta, D., Sriram, R.: Ontology-based exchange of product data semantics. IEEE Transactions on Automation Science and Engineering 2(3), 213–225 (2005)
13. Baumgaertel, H.: Distributed constraint processing for production logistics. IEEE Intelligent Systems 15(1), 40–48 (2000)
14. Smirnov, A., Levashova, T., Shilov, N.: Semantic-oriented support of interoperability between production information systems. International Journal of Product Development, Inderscience Enterprises Ltd. 4(3/4), 225–240 (2007)
15. Krizhanovsky, A., Lin, F.: Related terms search based on WordNet / Wiktionary and its application in Ontology Matching. In: Digital Libraries: Advanced Methods and Technologies, Digital Collections (RCDL 2009), September 17-21, Petrozavodsk, Russia (2009) (accepted for publication)
16. Wiktionary, the free dictionary (2008), `http://en.wiktionary.org/`

Operational and Structural Business IT Alignment

Mats-Åke Hugoson[1] and Kalevi Pessi[2]

[1] Jönköping International Business School,
P.O.B. 1026, SE-551 11 Jönköping, Sweden
[2] University of Gothenburg, Department of Applied IT,
P.O.B. 8718, SE-412 96 Göteborg, Sweden
Mats-Ake.Hugoson@ihh.hj.se, pessi@ituniv.se

Abstract. In recent years, Enterprise Architecture has gradually emerged as the preeminent means to change and transform large organizations. By employing Architectural Principles, organizations strive to master the complexity inherent in business processes and IT system and their harmonious alignment. Unfortunately, very rarely has the coevolutionary and emergent nature of alignment been taken into consideration in IS research. Even if different approaches focus on business IT alignment, most of them have a tendency to focus on alignment as a state or an outcome. In this paper we argue that a dynamic approach is necessary in order to achieve business IT alignment in a long-term perspective. Furthermore we demonstrate that the choice of architectural principles has an impact on the ability to achieve and maintain operational as well as structural alignment. A case study is used as a basis for the analysis. The conclusion is that an Enterprise-centric architecture can create freedom of action for dynamic operational alignment, and that Business-oriented IT management can keep the IS Architecture aligned with the Business Architecture in a long term perspective.

Keywords: Enterprise Architecture, Operational Alignment, Structural Alignment, Architectural Principles, Business Changes.

1 Introduction

Contemporary organizations invest more and more money in new IT systems and in managing legacy systems. At the same time organizations are finding it more and more difficult to keep these IT systems in alignment with business demands. Furthermore, the role of IT systems has changed during the last two decades, from automation of routine administrative tasks to a strategic and competitive weapon. In recent years, Enterprise Architecture has gradually emerged as the preeminent means to change and transform large organizations. In this sense Enterprise Architecture is playing an increasingly important role in improving IT management practice.

By employing architectural principles, one strives to master the complexity inherent in business processes and IT systems and their harmonious alignment [1]. One driving force is the vast legacy of IT systems that has accumulated over several decades. Yet another impetus to employ architectural principles is an increasingly

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 196–207, 2011.

dynamic and interconnected world that forces organizations to rely heavily on support from integrated IT systems to realize new ways of doing business. The challenge is essentially one of having the IT systems and their interaction allow for changes in business activities as well as changes in relations to the environment rather than discourage such changes [2][3]. It is therefore of paramount importance to carefully choose the principles by which one seeks to realize one's goals.

Architectural principles should be chosen based on the effects that are to be achieved in one's own business operations. They should however not be purely introspective, but rather consistent with changes in the business environment. This applies to the harmonious operation of existing IT systems as well as acquisition of new IT systems, thus creating desired effects in business operations which enable a long-term sustainable architecture that can be developed gradually. The latter makes great demands on change management. Unfortunately, very rarely has the coevolutionary and emergent nature of alignment been taken into consideration in IS research and practice [2]. Even if different approaches focus on business IT alignment, most of them have a tendency to focus on alignment as a state or an outcome.

In this paper we argue that an enterprise-centric approach instead of an IT-centric approach is more favorable in order to achieve business IT alignment in a long-term perspective. Furthermore we demonstrate that the choice of architectural principles has an impact on the ability to achieve and maintain operational as well as structural alignment. A case study is used as a basis for the analysis.

The research has been based on collaborative practice research [4] and its inside/outside perspectives. One author of this paper has been working with the case companies for several years and provides an "inside perspective". The other author is a full-time academic researcher and provides an "outside perspective", which allows for deeper assessment and reflection. The research methodology is essentially an interpretive case study [5], [6]. We carried out collection of data primarily through observations, semi-structured interviews, and workshops with CEO, CIO, Process Managers, Enterprise Architects, Project managers and Users. The main objective in the case study is to increase the understanding of the coevolutionary and emergent nature of business IT alignment.

2 Architectural Principles and Alignment

Enterprise Architectures have, for a number of years, been the focus of much academic attention due to the successful experiences of some companies and less successful experiences of others. Enterprise Architecture is usually divided into different categories or domains or architecture types. For example Aerts et al. [1] identify three domains in which architecture matters: (1) The business architecture, (2) The application architecture (or information systems architecture), (3) ICT platform architecture (or IT architecture). The developments in the various domains influence each other which makes agility that much harder to attain by businesses. Coping with change is, however, facilitated by architectures supporting reflectivity. Alignment practices must take into consideration the relation between the various architectures. Thus, there is a need for more knowledge about architectural principles that maintains a harmony between 1) the ever changing nature of business and 2) the capabilities provided by IT systems to respond both quickly and in a long term perspective to these changes [7][8]. Enterprise architecture design and management should therefore be guided by principles that contribute to a suitable and sustainable alignment. Architectural principles are statements that express how your enterprise needs to design and deploy IT systems across the enterprise in order to connect, share and structure information [9]. In this paper we focus on principles for delineation (differentiation) of IT systems and principles for integration of IT systems [10] [11] [12].

Architecture matters are more and more critical for the creation and maintaining of operational (short term) alignment as well as structural and strategic (long term) alignment between the business and its information systems. Henderson and Venkatraman proposed a Strategic Alignment Model (SAM) As early as 1993[13]. Still, despite years of cumulative research and practice, business IT alignment remains one of the leading areas of concern for business executives.

One common definition of alignment is "as the degree to which the information technology mission, objectives, and plans support and are supported by the business mission, objectives and plans" [14] [15]. This definition focuses on alignment as a state or an outcome. Another research perspective focuses on the process of alignment and suggests that business IT alignment is a continuous coevolutionary process [2] [16]. In this research perspective business IT alignment is considered a series of adjustments at several levels of analysis. There are different views on which levels to consider. For the purpose of this paper, we define two levels of analysis: Structural and Operational.

Structural alignment (or architectural alignment) means the degree of "structural fit" between IS-Architecture and the Business Architecture.

Operational alignment means alignment between a particular information system and the related business domain.

Considering business IT alignment as a continuous coevolutionary process emphasizes changes in organizations. In the management literature there are many theories of change in organizations [17]. A dynamic view of change focuses on the organizational ability to change rapidly and continuously [18]. Changes can occur at different levels of the organization and be of different magnitude. For example Bartunek and Moch [19] discuss different orders of change. In this paper we define three levels of changes with respect to their range:

First-order changes occur within a particular business domain and do not affect other business domains.

Second-order changes involve changes in business structures and affect relationships between domains, but also the definition and differentiation of business domains. Second-order changes are mainly structural changes within the organization.

Third-order changes involve changes in relationships with the business environment and are mainly inter-organizational changes.

3 Case Study

In this chapter, we introduce and discuss a case from Swedish industry in order to illustrate the coevolutionary and emergent nature of alignment. The case illustrates how architectural principles regarding IT systems and their interaction allow for changes in business activities as well as changes in relations to the environment rather than discourage such change. Even though the case is entirely based on an actual company, we will for the purpose of this paper refer to the company as 'BCC'.

3.1 The Business Structure

The mission for the BCC is to produce and sell liquid bleaching chemicals for the paper industry. It is a process industry producing two main products, with a limited number of variants. Product development is carried out together with customers, in order to improve bleaching in the pulp process, using fewer amounts of chemicals. So far the company operates only in one single site, with two process lines, serving a number of pulp and paper mills. Consequences for running out of essential chemicals in the paper mills are severe, and therefore each paper mill keeps high stock of bleaching chemicals, and calls for high delivery capacity and delivery precision from the supplier. Also, the chemical plant must keep a high amount of ready-made products in stock, as disturbances occur in both of the two production lines for bleaching chemicals.

Changes in the business structure

The delivery process. In order to reduce the amount of chemicals in the supply chain, the logistics manager has proposed and received customers' interest for Vendor Managed Inventory (VMI), which means that BCC will guarantee that bleaching chemicals are available at the paper mills. Customers' tanks will be filled up without orders. The total amount of chemicals of each type will serve as safety stock. VMI calls for accurate information on customers' actual situation, and a forecast for future demands. The proposal is to install meters connected to the delivery process at BCC in tanks of the customers. VMI-customers production plans should be available for BCC. The delivery process at BCC must make calculations and coordinate deliveries in an efficient way. There will be no major changes in other business processes at BCC. Customers that do not apply VMI will send their orders in the conventional manner through the sales process. The VMI delivery process can be established in a rather short period of time.

The maintenance process. So far all types of deliveries from suppliers (also components and spare parts for maintenance) are handled by the central purchasing department. The proposal from the maintenance manager is to define a new process with full responsibility for maintenance of the production lines, including supply of material, tools and resources. Future maintenance should be based on detailed up-keeping programs with components and spare parts available in order to reduce disturbances. Down-time due to emergency repair should be shortened through agreements with suppliers on direct deliveries that cover heavy equipment. When the maintenance process is reorganized and tested, outsourcing to trusted maintenance vendors could be an option. As the supply of raw material is seen as a part of production planning, the role for the purchasing department will be changed to more strategic procurement.

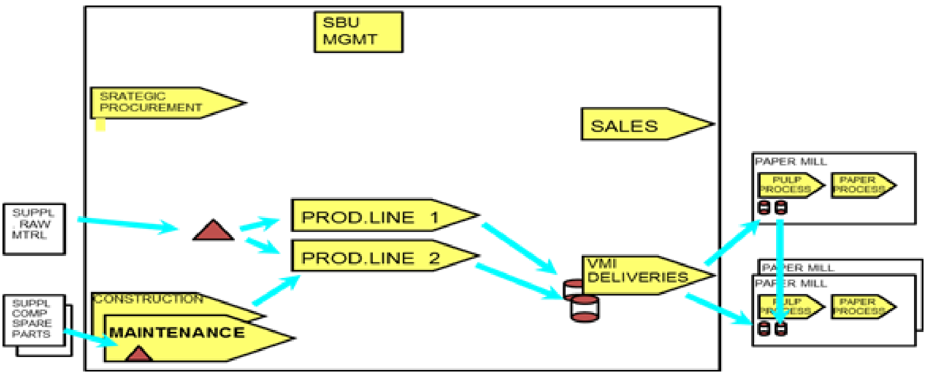


Fig. 1. The revised business structure

Business scenario

In a long term scenario the BCC management sees an expanding market. As the capacity in the current production lines is limited, new production sites must be established, probably through acquisitions in other countries.

Another business scenario is a closer partnership with customers. Product development and the sales process will merge with customer experts to a Customer Support process, operating at the customers' site with improvements in the bleaching process. The central product development process in BCC will provide knowledge and experts for Customer Support.

3.2 Strategic IT Planning

The current IT system cannot possibly be altered to support the changed business structure. The system is old, technically out of date, and there are strong dependencies between different parts of the system. An ERP vendor made an offer for a new integrated system that could handle two separate purchase and supply channels, and could be adapted to VMI deliveries. The CIO at BCC evaluated the offer and found that the proposed system was flexible enough to support different business processes. Costs and time delay for adaptation to VMI could be a problem. A main objection from CIO was that the integrated system could not match the strategic business plan in a timely manner. It would take a rather long time to implement the complete system, and as the new delivery process should be in operation as soon as possible and the maintenance process should be changed later, the system was not sufficiently flexible. The main reason for turning down the offer was however the risk for a ´lock in´ situation that would be a hindrance for future structural changes in BCC, such as Outsourcing, Customer Support and Acquisitions.

The Executive group at BCC assigned the CIO together with a strategy consultant to design the future IS architecture with the following directives:

- Development of the future IT support must be carried out step by step, coordinated with changes in the business structure.
- ERP solutions, or other standardized software products should be used.
- The IS Architecture shall support future changes in the business structure such as outsourcing, establishment of new production sites or new Customer Support processes.

The future IS Architecture

An objective for the strategic IT planning was to create a structure of interacting IT systems with a high degree of independence in order to enable IT development step-by-step, coordinated with business development. Both systems delineation and systems interaction must take demands from the business structure into account.

Systems are delineated according to main business responsibility: (1) For the new Delivery process a dedicated VMI system is delineated. (2) The changed Maintenance process is to be supported by a separate Maintenance System, allowing future outsourcing without major changes in other systems. (3) For the production processes a Material and Production planning system (MPS) was proposed. A separate system for each production process was discussed, but as the processes are partly integrated, and no structural changes are foreseen for the current production processes, a common system was considered more suitable. (4) A separate CRM system is delineated. The current sales process can use a rather simple Order handling system that can be replaced when the sales process is re-organized to Customer Support. (5) A central executive system is proposed for general accounting and performance follow up.

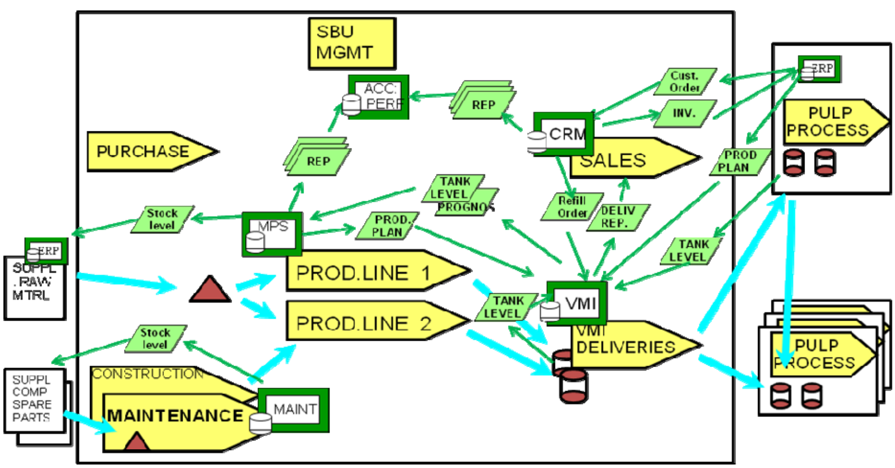


Fig. 2. The Future Strategic IS Architecture

Systems interaction through exchange of business based messages

In the future IS Architecture, interaction between different systems should reflect normal information exchange between different business processes, as shown in fig 2.

Interaction between the VMI system and customers is performed through incoming messages for Actual Tank Level and Production Plan as an EDI message from customers' ERP systems. For customers that have not (yet) implemented VMI support, Refill Orders are sent as a message from the CRM system. The VMI system sends Delivery report to the CRM system. The MPS system receives updates on the actual situation in one's own tanks, and forecasts from VMI as a basis for production planning, and reports production plans for each production line back to VMI. The detailed information exchange has to be specified for each system before implementation.

3.3 The Step by Step Implementation Procedure

The VMI system

The first step in implementation of the Strategic IS Architecture is to support Vendor Managed Inventory with a VMI system. The Logistics Manager is responsible for the project, and a contract with the specialized VMI software vendor is signed. The VMI system has to fulfill defined computerized interaction with external and internal systems. Actual situation in customer tanks and BCC tanks are picked up and an EDI connection is established with the customers' ERP system. Even if the new CRM system is not yet implemented, the existing system can be adapted to send Delivery Order and receive Delivery reports as specified. Messages on planned production are received from the central system.

The CIO controls the IT structure and supports the infrastructure for EDI communication with customers and an internal Message Handling system. The new system could be implemented and VMI deliveries established within a short period of time, as there were no major changes in other systems or other parts of the organization.

The Maintenance system

A process development project was carried out in order to outline the new maintenance principles for Quick maintenance and Zero emergency stops. As a part of the project, demands on the information system to support the renewed process were specified, together with the chosen software vendor. CIO gave directives to the vendor that the new Maintenance system must interact with the existing system through exchange of messages as specified in the Future IS structure, in order to allow outsourcing of the maintenance process.

CRM, MPS and ACCOUNTING Systems

The intention was to use ERP software products for the three remaining systems in the Strategic IS Architecture in order to finally replace the old system. The ERP vendor proposed (again) an integrated solution with a Customer module to support the whole organization with information on customers and orders, and a Product module with information on products and the situation in the production processes. The system should allow "drilling down" into single order information and detailed information on each type of chemicals along the flow.

The CIO clearly pointed out that the solution must be in compliance with the Strategic IS Architecture, and that interaction between the three systems should be carried out through information exchange, as specified in the Strategic IS

Architecture. The CRM should handle all information for support to the sales process, both customer information and product information, and MPS should give total support the production process. CIO explained to the ERP-vendor that BCC could still choose a CRM system form some other vendor, if the ERP vendor was not willing to comply with the Strategic IS Architecture.

Finally the ERP vendor accepted the structure and the three systems CRM, MPS and ACCOUTING could be implemented with the specified interaction. Functionality for the CRM system was defined by the sales manager to support the current Sales process. As the Production processes after reorganization of the purchasing department is also responsible for supply of raw material, MPS must include functionality for material planning and warehousing.

Future changes in the Strategic IS Architecture

Of special interest is the possibility to adapt the Strategic IS Architecture to inter-organizational changes in the business structure. A joint venture with the customer to control the bleaching process at customer´s site calls for customer specific information in a separate system, which can work in parallel with the CRM system.

An acquisition of a chemical plant, which is in operation with a suitable information system, will result in an extension of the Strategic IS Architecture. Interaction with some of the existing systems, especially the Accounting system, must be changed, but the information system for the additional plant must not be replaced.

4 Analysis of the Case

In this chapter, the case will be analyzed in terms of (1) Demands on IT systems to enable alignment, (2) Architectural principles, and (3) IT management issues.

4.1 Demands on IT Systems to Enable Alignment

Flexibility

It is evident that alignment in a long-term perspective calls for flexibility in IT support, but different kinds of flexibility must be analyzed.

The possibility to change a system. In the BCC-case the vendor proposed an integrated system, which should be adapted to different business processes in order to yield good operational alignment all over the company. ERP systems are generally very flexible. By defining a large number of parameters, the system can be well adapted to different demands from different users. A problem is however that once the system is implemented, it is more difficult to change the system and a change will have an impact on many users – even those who do not want change. As the integrated system is to be implemented in full at one single occasion, maintaining operational alignment in a long-term perspective will be problematic, as all changes in the enterprise do not occur at the same time.

The possibility to change one part of the system (one system in a structure of systems) without changes in other parts of the structure. It is often argued from ERP vendors that modern integrated ERP systems offer this possibility through modularization. Even so, in a long term perspective one runs the risk of a ´lock in´ situation if every part of the organization must use the same system. When the system finally must be replaced there will be a new ´big bang´. In a strategic perspective the integrated system will not support outsourcing or other structural changes. In BCC the risk for misalignment in the long-term perspective was the main reason to reject the proposal from the ERP vendor.

Possibility to replace a part of the structure with a system of different origin, without changes in the remaining parts of the structure. This kind of flexibility enables operational alignment in a long-term perspective, but is generally not available in integrated ERP solutions. Other architectural principles must be applied.

Possibility to change the structure through insertion, deletion, or changed relations between different systems. This type of flexibility is closely connected to structural alignment and will be further discussed later on.

The Principle of Independence and Interdependence

The analysis so far has pointed out that operational alignment can be maintained if the IT support for a certain areas of operation (for instance the Delivery process) can be changed or replaced without changes in other parts of the total structure. Independence is the main condition for this freedom for change, but interaction with other parts of the total structure must be satisfied.

The conditions can be expressed as The Principle of Independence and Interdependence:

- A change in the inner structure of a system should not change the relations to other systems in the structure. (First order change)
- Relations between different systems in the structure must be defined and described independent of the inner structure in each system.
- Changes in relations between some systems in the structure (Second and Third order change) may have an impact on the inner structure of these systems, but only on the inner structure in these systems.
- Systems can be added to or deleted from the structure, which may change relations to some systems in the structure. Remaining systems in the structure should not be affected.

These basic principles must be considered when architectural principles for operational and strategic alignment are analyzed.

4.2 Architectural Principles

IT centric versus enterprise centric architectures

The IS Architecture in the proposal from the ERP vendor in the case can be characterized as **IT-centric.** The design of different parts of the system is based on a generic information model, as the integrated system cannot possibly in advance reflect the structure of the different enterprises where it will be implemented. Generic information processes to support different business processes are designed for flexibility in enterprises where the system is to be installed. When business processes in the actual enterprise differs too much, the actual business processes must be adapted to the system [20]. Both delineation of different parts (modules) in the system, and interaction between different parts is based on the generic information model. A module generally handles a specific type of information for the whole enterprise (the Customer module supports everyone with customer information) and therefore a change in one module will affect many parts of the organization. It is generally possible to exclude modules that are not relevant for the individual enterprise (if there is no production the Production Planning module is omitted), but it is not possible to insert a module of different origin. Concepts like Service oriented architecture for ERP systems can increase the ability to create and maintain IT systems fit to structures, processes and information needs of the specific enterprise, but the interaction between different parts of the system is still based on the generic information model.

The conclusion must be that an IT centric architecture cannot satisfy the Principle of Independence and Interdependence. Even if operational IT alignment can be achieved when the IT centric architecture is implemented, it will be difficult to maintain operational alignment in a long term perspective. IT structure will be less and less aligned with the business structure as different parts of the enterprise are changed. Finally the whole structure must be replaced.

An **Enterprise centric IS-architecture** represents a quite different approach – both for delineation and for interaction. The business structure forms the basis for the IS structure, and each part (system) of the IS Architecture is delineated to support a specific area of responsibility in the actual enterprise. The computerized interaction between different systems is based on ordinary information exchange between different areas of responsibility (inside and outside the enterprise), and can therefore be specified independent of the inner structure in each system. Functionality, data storage, technology can be specified and implemented for each system (such as ERP software) fully independent of all other systems in the structure. The only condition is that each system must comply with the computerized systems interaction specified in the enterprise specific IS architecture. Local changes in one area of operations can be supported by changes in (or replacement of) the system that supports that area, without changes in any other system. Different systems of the same or of different origin can interoperate in the IS Architecture. The Enterprise centric IS Architecture must be supported by an infrastructure for systems interaction.

The conclusion drawn from the case study is that an Enterprise centric IS Architecture can satisfy the Principle of Independence and Interdependence, and also enable both Structural and Operational IT alignment in a long-term perspective. It is shown possible to coordinate changes in IT support with different types of changes in the enterprise. A sustainable IS Architecture can be kept aligned with the business structure even after structural changes of the second and third orders in the enterprise.

4.3 IT Management Issues

In BCC the CIO plays a vital role both in strategic IT planning and during the implementation process. The executive group gave directives, and the CIO found that enterprise centric IS Architecture is an appropriate tool to satisfy these directives. Responsibility for the Strategic IS Architecture is seen as an enterprise issue. The case study indicates that this is a condition for enterprise-oriented architectures, as vendors generally propose IT-centric solutions. Strategic choice of vendor is seen as a central responsibility, and the CIO must control compliance with the chosen architectural principles.

On the other hand we have the decentralized approach used during the implementation process. CIO controls the structure, and gives authorization for each development project, but the manager for each area of operations is responsible for functionality in the system, for evaluation of solutions and for coordination of business development and IT development.

It is shown in the presented case that an enterprise centric architecture enables central responsibility for structural IT alignment, and local responsibility for operational alignment.

5 Summary and Conclusion

The case shows that both operational and structural alignment can be achieved through an Enterprise centric architecture, in combination with Business oriented IT management. Independence between different systems in the IS Architecture is a key issue, that enables operational alignment through coordination of gradual local changes in different areas of operations with changes in corresponding IT support for each area, without major changes in other parts of the IS Architecture. A decentralized management can be applied for this coordination. Responsibility for the IS Architecture is on the other hand seen as a central responsibility within the enterprise that cannot be given to IT specialists, inside or outside the organization. Each step in IT development must be in compliance with the IS Architecture. IT management decisions on changes in the IS Architecture corresponding to gradual changes in the business structure can maintain structural alignment in a long term perspective.

References

1. Aerts, A.T.M., Goossenaerts, J.B.M., Hammer, D.K., Wortmann, J.C.: Architectures in context: on the evolution of business, application software, and ICT platform architectures. Information & Management 41, 781–794 (2004)
2. Benbya, H., McKelvey, B.: Using coevolutionary and complexity theories to imrpove IS alignment: a multi-level approach. Journal of Information Technology (21), 284–298 (2006)
3. Chan, Y.E.: Why haven't we mastered alignment? The importance of the informal organization structure. MIS Quarterly Executive 1(2) (June 2002)
4. Mathiassen, L.: Collaborative Practice Research. Scandinavian Journal of Information Systems (14), 57–76 (2002)
5. Walsham, G.: Interpretive Case Studies in IS Research: Nature and Method. European Journal of Information Systems (4), 74–81 (1995)
6. Klein, H.K., Myers, M.D.: A set of principles for conducting and evaluating interpretive field studies in information systems. MIS Quarterly 23(1), 67–94 (1999)
7. Beimborn, D., Franke, J., Wagner, H.-T., Weitzel, T.: The Impact of Operational Alignment on IT Flexibility – Empirical Evidence from a Survey in the German Banking Industry. In: Proceedings of Americas Conference on Information Systems. Paper 131 (2007)

8. Wagner, H.-T., Weitel, T.: Operational IT Business Alignment as the Missing Link from IT Strategy to Firm Success. In: Proceedings of Americas Conference on Information Systems. Paper 74 (2006)
9. Hugoson, M.-Å., Magoulas, T., Pessi, K.: Interoperability Strategies for Business Agility. In: Dietz, J.L.G., et al. (eds.) CIAO! 2008 and EOMAS 2008. LNBIP, vol. 10, pp. 108–121. Springer, Heidelberg (2008)
10. Allen, B.R., Och Boynton, A.C.: Information Architecture: In Search of Efficient Flexibility. MIS Quarterly 15(4) (December 1991)
11. Solotruk, M., Kristofic, M.: Increasing the Degree of Information System Integration and Developing an Integrated Information System. Information & Management 3(3) (1980)
12. Hugoson, M.-Å., Magoulas, T., Pessi, K.: Architectural principles for alignment within the context of agile enterprises. In: Proceedings of the 3rd European Conference on Information Management and Evaluation (2009)
13. Hendersson, J.C., Venkatraman, N.: Strategic alignment: Leveraging information technology for transforming organizations. IBM Systems Journal 36(2&3) (1999)
14. Reich, B.H., Benbasat, I.: Measuring the Linkage Between Business and Information Technology Objectives. MIS Quarterly 20(1), 55–81 (1996)
15. Reich, B.H., Benbasat, I.: Measuring the Information Systems - Business Strategy Relationship. In: Leidner, D.E., Galliers, R.D. (eds.) Strategic Information Management: Challenges and Strategies in Managing Information Systems, pp. 265–310. Butterworth-Heinemann, Oxford (2003)
16. Maes, R., Rijsenrij, D., Truijens, O., Goedvolk, H.: Redefining business IT alignment through a unified framework. PrimaVera Working Paper 2000-19. Universiteit van Amsterdam (2000)
17. van de Ven, A.H., Scott Poole, M.: Explaining Development and change in organizations. The Academy of Management Review 20(3), 510–540 (1995)
18. Brown, S.L., Eisenhardt, K.M.: The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. Administrative Science Quarterly 42(1), 1–34 (1997)
19. Bartunek, J.M., Moch, M.K.: First-Order, Second-Order, and Third-Order Change and Organization Development Interventions: A Cognitive Approach. The Journal of Applied Behavioral Science 23(4) (1987)
20. Davenport, T.H.: Putting the Enterprise into the Enterprise System. Harvard Business Review (July-August 1998)

On the Role of Competence Models for Business and IT Alignment in Network Organizations

Vladimir Tarasov[1] and Kurt Sandkuhl[2]

[1] School of Engineering, Jönköping University, P.O. Box 1026, 55111 Jönköping, Sweden
vladimir.tarasov@jth.hj.se

[2] Institute of Computer Science, Rostock University, 18055 Rostock, Germany
kurt.sandkuhl@uni-rostock.de

Abstract. Customer-motivated changes in network organizations lead to alignment of business and IT-infrastructure. This paper investigates the use of competence models in network organizations for more efficient alignment of IT-infrastructure to changed business demands. The paper shows that a competence model is pertinent in a process of aligning an IT-infrastructure to changed business demands. Then, a way is presented of extracting alignment needs from competence information in enterprise models. Finally, our approach for competence profile management is extended with alignment operations. These operations together with other operations on competence profiles form a lifecycle model for competence profile management. Such operationalization allows for support of aligning network organizations to changing business demands.

Keywords: Continuous engineering, continuous business engineering, business and IT alignment, network organizations, competence modelling, competence profile management.

1 Introduction

Competence modelling and enterprise modelling are tightly related subject areas. In general terms, enterprise modelling is addressing the systematic analysis and modelling of processes, organization structures, products structures, IT-systems or any other perspective relevant for a given modelling purpose [1]. The knowledge about enterprises captured in enterprise models is usually a starting point for organizational change or innovation activities, or is used in information system development. In this context, modelling of organization structures including organizational roles is part of most enterprise modelling projects. Competence modelling contributes an additional dimension to enterprise modelling by capturing what competences, skills and abilities are required at what level in order to perform a task, both from organizational and from individual perspective. However, using competence information for business and IT alignment purposes so far has not received much attention.

This paper investigates the use of competence models in network organizations for more efficient alignment of IT-infrastructure to changed business demands.

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 208–219, 2011.

Based on a lifecycle model for competence profile management, we present an approach for identifying adaptation needs of IT-infrastructures and for preparing change processes. The main contributions of this paper are (1) to show that a competence model is pertinent in a process of aligning an IT-infrastructure to changed business demands, (2) to present a way of extracting alignment needs from competence information in enterprise models, and (3) to extend our approach for competence profile management with alignment operations.

The paper is structured as follows: Section 2 introduces the background for the paper from the areas of competence modelling and enterprise modelling. Section 3 describes the application context for this work, i.e. competence modelling for network organizations. Section 4 presents the approach for competence profile lifecycle management. Section 5 shows how to use competence models for alignment purposes and extends the lifecycle model. Section 6 summarizes the work and gives an outlook on future work.

2 Background

Work from competence management (Sect. 2.1) and enterprise modelling (Sect. 2.2) form the background for this paper and will briefly be presented in this section. This includes an example of enterprise modelling-based competence modelling in Sect. 2.2.

2.1 Competence Management

Competence management and supply is a research area with tradition in different areas of scientific fields, like human resource management in economics, lifelong learning in educational sciences or organizational learning in organizational sciences. In the literature, different definitions of competence can be found with focus either on individual of enterprise competence. Examples are [2] who define a competence as a set of all knowledge forms and personal abilities that are required for performing tasks and defining competence as the ability to combine in an efficient manner a number of non-material resources and material resources in order to respond to the need of an activity to be performed in an enterprise. We will use the definition of Pepiot et al., [3] with a clarification regarding the non-material resources and an adjustment regarding the application context addressed (see also [4]):

> Competence is the ability of an organization to combine in an efficient manner a number of non-material resources and material resources in order to respond to the need of an assigned task within a joint process with other partners. Non-material resources include organizational resources (best practices, know-how, organizational culture) and individual abilities (personal abilities and knowledge forms).

There have been efforts to standardise competence models, like HR-XML developed by the HR-XML Consortium [5]. HR-XML is a library of XML schemata with focus on modelling of a wide range of information related to human resource tasks. Using such schemas it is possible to define profiles in order to use competency definitions. It specifies data sets like job requirement profiles or personal competency profiles. The former describes competences that a person is required to have. The latter describes competencies that a person has.

Several approaches have been developed to provide a systematic evaluation of competences in enterprises. The approach proposed by Berio & Harzallah [6] suggests the CRAI model (competence, resource, aspect, individual) associated with axioms based on set theory. The approach aims at describing formal competence in order to provide a mapping between required and acquired competence in an enterprise reengineering context. Competences are characterized by sets of knowledge, know-how and behaviour associated to a context and linked to individual actors. Based on a classical evaluation of these characteristics, a mathematical aggregation is suggested to provide a quantitative evaluation of competences.

Recent work in competence management includes best practices for competence modelling [7], application of competence management for performance optimization in industry [8] or for networked defence [9].

2.2 Enterprise Modelling and Competence Modelling

Enterprise modelling is an important tool for strategic planning of any enterprise today. It consists of the process of building models of the whole or a part of an enterprise with process models, data models, resource models, etc. It is based on the knowledge about the enterprise, previous models, reference models and/or domain ontologies. The term "enterprise model" is used in industry to denote differing enterprise representations, with no real standardized definition. Due to the complexity of enterprises, a vast number of differing enterprise modelling approaches have been pursued across industry and academia. Enterprise modelling constructs can focus upon manufacturing operations and/or business operations; however, a common thread in enterprise modelling is an inclusion of assessment of information technology.

Although this is a well-known research field and there are several approaches and modelling languages available, competence modelling has not really been included in the enterprise models. To address this issue the Unified Enterprise Competence Modelling Language (UECML) has been developed [10]. UECML has integrated the concept of competence in the language at three levels; the competence itself, individual competence, and aggregated competence of a group of individuals. UECML is quite unique in explicitly integrating human competence in enterprise models.

Enterprise modelling extended with a competence modelling perspective is a suitable means to capture both organizational and individual competences. Fig. 1 shows an example for such an enterprise-model based competence model, which is based on a case from Automotive supplier industries. The model shows the top level of a model for "country-specific testing", i.e. tests of products with respect to their conformance to legislation and standards in a foreign country. Such a

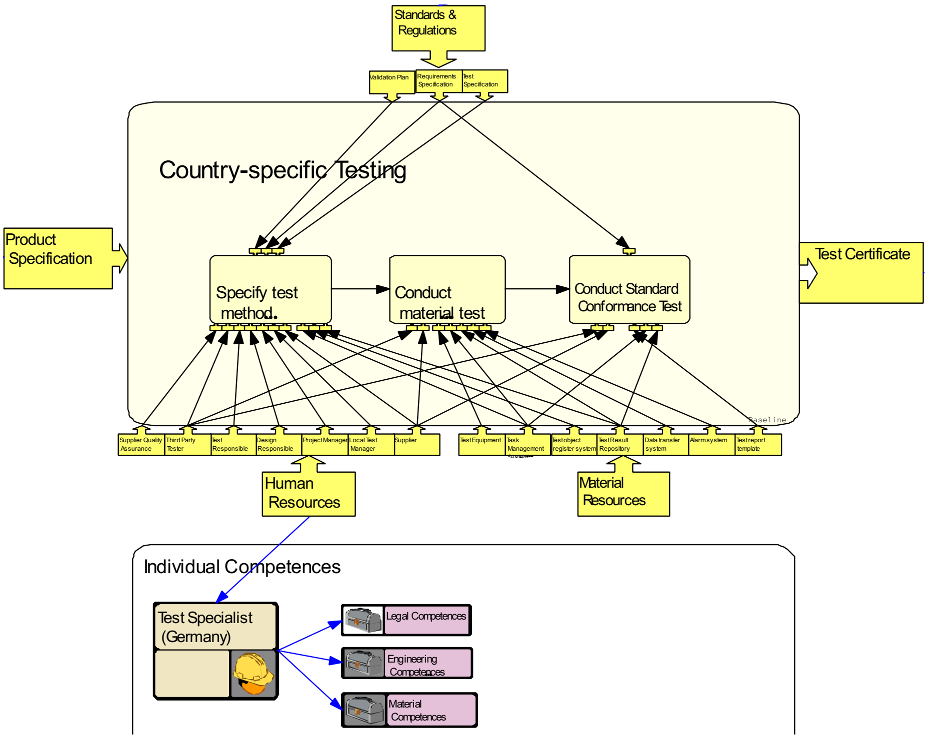


Fig. 1. Example competence demand model (adopted from [4])

testing typically requires specification of test methods and test execution, which in the example is divided in material test and standard conformance test (see task included in the box "Country-specific testing). These three tasks require human resources and material resources, represented by the yellow boxes below the process box. The human resources can be connected to individual competences, which is shown with the example of the "Test Specialist (Germany)" and the required competences. All boxes included in the models should be refined to an adequate level.

In comparison to conventional enterprise models, which often focus on processes, organization structures, products and systems in the enterprise under consideration, competence models have to pay specific attention to organizational and individual abilities:

– A competence perspective has to be added, in order to establish an explicit way of structuring and relating individual competences in enterprise models. In contemporary approaches, competence are either considered as attributes of roles or not included,

– The systems perspective has to be extended in order to provide an explicit way of structuring the material resources needed.

Furthermore, information about the needed capacity for both material and non-material resources when modelling the enterprise has to be captured. In practice, this information can be included by adding adequate attributes to the modelling elements.

3 Application Context

The application context selected for illustrating our approach of using competence modelling for business and IT-alignment are network organizations. This section will introduce the general application context (3.1) and discusses specific requirements of network organizations to business and IT alignment (3.2).

3.1 Network Organizations

In many industrial domains, globalization and the adaptation of supply strategies to global markets resulted during the last 10 years in collaborative or network organization forms. The needs for shorter innovation cycles, lead time reduction or better customization possibilities have stimulated the creation of numerous collaborative partnerships, like networks of suppliers and sub-suppliers, value networks [11] or co-operations in product development or construction projects. In such constellations, a clear understanding what the different suppliers can contribute to a joint project or manufacturing contract is crucial. This understanding has to encompass production capability, services offered, available resources, as well as the organizational competences of the members in the network. In this context, concepts and approaches from competence management can contribute to capture, systematically develop and integrate organizational competence of the member companies in such networks and customer organizations using these competences.

A supply network aggregates independent companies based on the principle of cooperation within a defined application domain and capable of coordinating their activities for production and delivery of the desired product/service. Organizations of this form use information and communication technologies to extend their boundaries and physical location [12] and form multiple links across the boundaries to work together for a common purpose [13]. Such networks experience different phases and situations which form the "life-cycle" of a network organization and can be considered as organizational frame for competence supply. The most important phases are [14]:

– **Community building:** enterprises with joint objectives or interests gather in a community of loosely coupled members. Main purpose is information exchange and communication within the network.

- **Formation:** based on specific requirements for a collaboration project (e.g. a joint engineering activity, like product development), the formation of a project team is started based on the capabilities of the members. As a result of this phase, potential partners with respect to the specific requirements have been identified.
- **Integration:** potential team members have been selected and negotiate the legal and financial conditions for joint project work. Furthermore, a collaboration infrastructure is being implemented for all relevant levels of collaboration. The result of this phase is a project network.
- **Operation:** the collaboration project is carried out within the project network. This operation is supported by the collaboration infrastructure.
- **Discontinuation:** the project network discontinues to exist. Dis-integration on all levels of the collaboration infrastructure and with respect to legal and financial issues is carried out.
- **Community dissolvement:** the joint objectives or interests within the community no longer exist. The network is dissolved.

Usually it is not fully obvious to the network members, which competence and resources are available from which partner in which quantity to which expenses and how to access them. In particular in the formation and operation phase, efficient support for configuration of collaborations and efficient reuse of existing knowledge is a critical success factor. This includes selection of suitable partners based on the competence demand for a planned collaborative project.

Competence demand modelling, as illustrated in Sect. 2.2, forms a basis for these activities. Modelling organizational competences of different members of a collaborative network creates different member competence models, which can be used in the formation and operation phases for identifying the "right" partners and preparing operative solution (cf. [14]).

3.2 Specific Requirements to Business and IT Alignment

This section discusses the specific requirements of network organizations with regards to Business and IT alignment.

Business and IT alignment in general addresses the systematic and efficient adaptation of IT-applications and infrastructures to changing organizational needs. In enterprises, such "changing needs" can usually be expressed in terms of new or modified business goals. In network organizations, the situation is slightly different, since such networks often serve larger customers and changes of the customer's objectives lead to changes in the projects performed by the network for the customer. In the following, we will focus on such customer-motivated changes in networks, which primarily affect the formation and integration phases of the lifecycle model of a network organization. Of course, there also could be changes in the overall objectives of the networked community, which might lead to adaptation needs of the network-internal infrastructure. Such change processes are outside the scope of this paper.

An example of customer-motivated changes could be an additional service completing the production of a physical product. A network of SMEs producing a component for a vehicle, to take one example, could be asked to also provide a new service related to manufacturing. Instead of only producing and delivering the part to the assembly line, ad-hoc configuration changes at the assembly line shall be possible. For this service, an additional competence has to be added to the existing supply network, either by one of the partners already included or by a new partner, which needs to be integrated into the supply network, both process-wise and regarding the IT-infrastructure.

The formation phase requires an efficient way to describe the needed competences completely and on a sufficient level of details. The integration phase adds requirements regarding selection of members. Table 1 summarizes the requirements from both phases, i.e. what specific requirements these two phases pose to support of IT-alignment in case of changed customer demand.

Table 1. Requirements from lifecycle phases to competence demand models

Formation phase	Integration phase
Possibility to specify additional task or task modifications to be delivered to the customer due to changed requirements	Identify competence needed to deliver the additional/modified task
Possibility to define the results to be delivered by the task	Determine competence gap between existing network competence and required competence
Access competence profiles of companies currently serving the customer	Identify additional network member able to deliver the missing competences
Possibility to define resources including capacities required for the additional task	Identify work processes to be integrated between additional member and existing network
	Identify interfaces to be established between IT-applications of the additional member and the existing network

4 Lifecycle Model for Competence Profile Management

Sect. 2.2 presented a method to model competences of enterprises. The result of such modelling can be a competence profile for an enterprise—the set of all competences possessed by the enterprise. Competence profiles are important because they allow for solving different tasks like competence supply or business community creation. However, just having competence profiles is not enough to support aligning network organizations to changing business demands. To do this we need to identify different operations on competence profiles. These operations underpin construction of a system for competence profile management to support business and IT alignment.

To operationalize competence profile management, we first need to define the semantics of competence, competence profile and operations. Formal definitions of the semantics are necessary to facilitate implementation of a competence profile management system. The work presented in [15] introduces definitions of competence, competence level, and competence profile on the level of an individual worker. The definitions of these concepts are based on the context of an enterprise. For example, a competence profile of a worker is defined as a collection of all competences needed to act in the roles that are assigned to this staff member in the organizational structure of an enterprise.

The operations to manage competence profiles specified in [15] include the following ones:

- Create a competence profile for a worker,
- Add competence to the profile of a worker,
- Remove competence from the profile of a worker,
- Limit the competence profile of a worker to a certain part (subprofile),
- Match the competence profile of a worker against a document or another resource,
- Search for competence profiles that can support carrying out certain tasks,
- Rank the found competence profiles according to a given criterion,
- Aggregate a group of competence profiles,
- Track down changes in a group of competence profiles that happen at certain time points.

For instance, "search for competence profiles" is defined as finding all competence profiles that included competences needed to support carrying out the given tasks. The definitions and operations given in [15] are specified at the level of an individual worker. Nevertheless, they can be extended to the level of an enterprise by introducing a definition of a competence profile of an organization and adding new operations needed to support tasks like business and IT alignment. The operations on competence profiles form a lifecycle model for competence profile management. The extended version of the lifecycle model with the new operations is depicted in Fig. 2.

5 Competence Models for Business and IT-Alignment

This section will investigate the potential of using competence models for supporting business and IT-alignment in networks of enterprises and present an approach based on the lifecycle model presented in Sect. 4.

Starting point for this investigation are the requirements identified in Sect. 3.2. In order to support alignment:

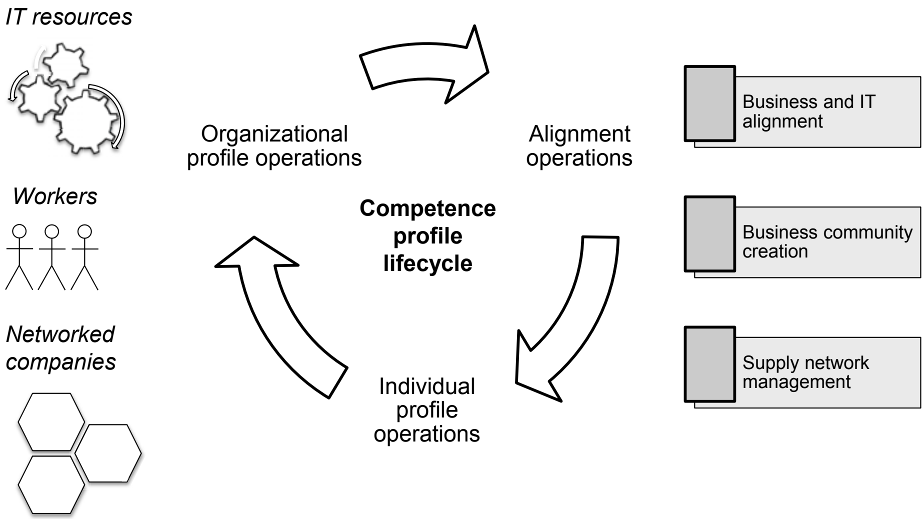


Fig. 2. Lifecycle model for competence profile management extended with alignment operations

- Competence models have to be rich enough from a content perspective, i.e. the required information has to be part of the competence models,
- Functionality for identifying alignment and preparing the alignment as such has to be provided, i.e. the operations provided by the lifecycle model have to be suitable for this purpose and work with competence models.

These two aspects will be discussed in the following sections.

5.1 Alignment Requirements Supported by Competence Models

Regarding the content of the competence models, we will discuss those requirements identified in Sect. 3.2, which include direct requirements to content. These requirements are listed in the left column of Table 2 ("F:" and "I:" are used to indicated the origin from formation or integration phase).

The comparison between requirements and available content in competence models shows that all required information to support the required functionality is included in competence model. Furthermore, the modelling language (expressed by the meta-model) used for enterprise models allows for traversing models from high-level concepts, like business objectives, to low-level elements, like single individual competence parts, and vice versa, by using the relationships included.

This richness and suitability of competence models for IT-alignment is also confirmed by previous work on competence demand models, which can be used for the purpose of specifying required organizational and individual competences (see [4]).

Table 2. Alignment requirements and their support by content of competence models

Requirement from formation or integration phase	Support by competence model
F: Possibility to specify additional task or task modifications to be delivered to the customer due to changed requirements	Competence demand models can be used for the purpose of specifying required organizational and individual competences
F: Possibility to define the results to be delivered by the task	Competence models include work results of tasks (see Sect. 2.2)
F: Access competence profiles of companies currently serving the customer	Not supported by competence models as such; tool support for the actual network configuration is subject of operation phase
F: Possibility to define resources including capacities required for the additional task	Competence models include resources (see Sect. 2.2); capacities need to be added explicitly
I: Identify work processes to be integrated between additional member and existing network	Competence models include work processes (see Sect. 2.2)
I: Identify interfaces to be established between IT-applications of the additional member and the existing network	Competence models include IT-applications related to processes (see Sect. 2.2)

5.2 Alignment Requirements Supported by Lifecycle Operations

Regarding the operations from the lifecycle model, we will proceed in the same way as with the previous section: discuss those requirements identified in Sect. 3.2 which include direct requirements to operations. These requirements are listed in the left column Table 3 ("F:" and "I:" are used to indicated the origin from formation or integration phase).

Sect. 4 has showed that the initial lifecycle presented in [15] provides a suitable framework for supporting business and IT alignment in network organizations with competence profile operations. However, Table 3 shows that the initial set of operations is not enough for business and IT alignment. Hence we need to extend the lifecycle model presented in [15] with operations on organizational profiles and alignment operations. The resulting extended version of the lifecycle model is shown in Fig. 2.

6 Summary

This paper addressed the problem of business and IT alignment in network organizations. The proposed solution is based on the use of enterprise model-based competence modelling and the lifecycle model for competence profile management. First, the requirements to business and IT alignment were analysed based

Table 3. Alignment requirements and their support by operations of the lifecycle model

Requirement from formation or integration phase	Operations from lifecycle model required
F: Possibility to specify additional task or task modifications to be delivered to the customer due to changed requirements	Add a new task to the set of all tasks to be carried out by the network organizations
F: Access competence profiles of companies currently serving the customer	Not supported by competence models as such; tool support for the actual network configuration is subject of operation phase
I: Identify competence needed to deliver the additional/modified task	Find all the competences that are needed to perform the given task
I: Determine competence gap between existing network competence and required competence	Calculate the difference between competences needed for the given set of tasks and the set of all competences possessed by the members of the organization network (collective competence)
I: Identify additional network member able to deliver the missing competences	Search for profiles of companies that include competences needed to perform the specified tasks
I: Identify work processes to be integrated between additional member and existing network	Find the work process that include the given set of tasks
I: Identify interfaces to be established between IT-applications of the additional member and the existing network	Find resources of type "IT-resource" used by the given task. Map the interfaces of the found IT-resource using their descriptions from the underlying enterprise model. The implementation of the mapping of IT interfaces is subject of operation phase.

how a business community is created in the case of network organizations. Then, competence models based on enterprise modelling were shown to meet these requirements. Finally, the lifecycle model for competence profile management was analysed. The initial lifecycle presented in [15] provides a suitable framework for supporting business and IT alignment in network organizations with competence profile operations but lacks explicit definitions and operations on the organizational level. Thus, the lifecycle model was extended with additional operations based on the requirements analysis. Such operationalization allows for support of aligning network organizations to changing business demands.

Future work will focus, first of all, on extending the initial lifecycle model with the operations identified in Table 3. This will also requires defining a competence profile on the organizational level. Besides, we will consider application of our approach to a particular case in order to validate the proposed approach.

References

1. Vernadat, F.B.: Enterprise Modeling and Integration: Principles and Applications, 1st edn. Chapman & Hall, London (1996)
2. Bjurklo, M., Kardemark, G.: Social tests in a model for controlling the enhancement of competence. Journal of Human Resource Costing & Accounting 3(2), 51–64 (1998)
3. Pépiot, G., Cheikhrouhou, N., Fürbringer, J.M., Glardon, R.: A fuzzy approach for the evaluation of competences. International Journal of Production Economics 112(1), 336–353 (2008)
4. Sandkuhl, K., Tarasov, V.: Modelling competence demand for flexible supply networks. In: 13th IFAC Symposium on Information Control Problems in Manufacturing, IFAC, pp. 103–108 (2009)
5. HR-XML Consortium: HR-XML 3.1 release (2010), http://ns.hr-xml.org/schemas/org_hr-xml/3_1/ (accessed May 29, 2011)
6. Berio, G., Harzallah, M.: Towards an integrating architecture for competence management. Computers in Industry 58(2), 199–209 (2007)
7. Albertsen, T., Sandkuhl, K., Seigerroth, U., Tarasov, V.: The Practice of Competence Modelling. In: van Bommel, P., Hoppenbrouwers, S., Overbeek, S., Proper, E., Barjis, J., et al. (eds.) PoEM 2010. LNBIP, vol. 68, pp. 106–120. Springer, Heidelberg (2010)
8. Boucher, X., Bonjour, E., Grabot, B.: Formalisation and use of competencies for industrial performance optimisation: A survey. Computers in Industry 58(2), 98–117 (2007)
9. Albertsen, T., Sandkuhl, K., Tarasov, V.: Towards competence modeling and competence matching for network-based defense. In: Norsell, M. (ed.) Stockholm Contributions in Military-Technology 2007, pp. 9–22. Swedish National Defence College (2008)
10. Pépiot, G., Cheikhrouhou, N., Furbringer, J.M., Glardon, R.: UECML: Unified enterprise competence modelling language. Computers in Industry 58(2), 130–142 (2007)
11. Tapscott, D., Ticoll, D., Lowy, A.: Digital Capital: Harnessing the Power of Business Webs. Harvard Business School, Boston (2000)
12. Skyrme, D.J.: The networked organization (1999), http://www.skyrme.com/insights/1netorg.htm (accessed May 29, 2011)
13. Laudon, K.C., Laudon, J.P.: Management Information Systems: Organization and Technology in the Networked Enterprise, 6th edn. Prentice Hall, Hemel Hempstead (2000)
14. Blomqvist, E., Levashova, T., Öhgren, A., Sandkuhl, K., Smirnov, A.: Formation of enterprise networks for collaborative engineering. In: Post-conference Proceedings of 3rd International Workshop on Collaborative Engineering (2005)
15. Tarasov, V.: Ontology-based approach to competence profile management. International Journal of Knowledge and Information Systems (under review)

Towards Value-Driven Alignment of KMS for SME

Ulrike Borchardt

University of Rostock, Chair of Business Information Systems,
18051 Rostock, Germany
ulrike.borchardt@uni-rostock.de

Abstract. To be successful in business small and medium enterprises have to focus on the optimal utilization of their spare resources. IT - System are supposed to support this process, yet as shown by a survey and systematic literature analysis presented in this paper, by now the concept of KM is not widely established in SME nor are KMS fully integrated or aligned to business strategies. Within the limited resources of SME, the question of how the business value of a well integrated application suiting a SME's objectives can be shown, has to be discussed. In this paper we present our suggestion on how this problem could be approached by using a framework based on the business values and KM dimensions to be able to determine the most effective KMS for a SME.

1 Motivation

To be successful in business small and medium enterprises have to focus on the optimal utilization of their limited resources. This also implies taking advantage of their knowledge and manage it as well as possible. With regard to the demographic development in Europe SME, like all other enterprises, will have to face the retirement of the baby boomer generation within the next 15 years. And those will certainly leave a gap in the enterprises knowledge base if the enterprise does not take precautions. However, this implies that knowledge has to be transferred either to other employees or to systems to store it. Today, the community is well aware of the fact, that KM is supposed to be an holistic approach and should not focus on IT systems only, yet their successful use is a valuable support as well as an indicator for a successful enterprise. Already Sveiby [1] found a relation between the growth of SMEs and their dedication to KM. In addition an holistic approach also makes KM a resource consuming process which might become a problem if resources are already scarce.

Regardless these offered advantages and potentials of KM solutions for SME, not much is known about the recent state of use of KMS in SME and their contribution to the business value and competitiveness of the enterprise. Certainly, do recent studies investigate on the IT or information use, yet they do not cover KMS in general or the business value support in particular. This paper aims to contribute to clarify the current situation and demand of SME with respect to

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 220–231, 2011.

KMS use. Furthermore we will relate this to the field of business and IT alignment, as a KMS is of no use if it does not support the business strategy of the enterprise. Finally, we will propose an approach for value-driven alignment for the use of KMS in SME. To achieve this we structured our paper as follows: Section 2 shortly presents the fundamentals of SME, knowledge management (KM) and knowledge management systems (KMS) as well as the business value of IT (BVIT). Section 3 presents the recent situation of SME and their KM in the federal state Mecklenburg-Vorpommern, Germany, describing the results gained from a survey done on knowledge intensive SME. Moreover the findings of literature research focusing on the question how IT systems can support KM according to the business strategy will be presented there. Finally, section 4 sums up the found results and section 5 presents our ideas on how the business value and KMS can be aligned to match the business strategy, by building up a framework which allows the determination of the suiting KMS or KM application by the business' objectives.

2 Background

This chapter will shortly present the theoretical backgrounds used for our research work and accordingly in this paper, as there are Small and Medium Enterprises (SME), as well as some KM and KMS. In addition some general facts on the BVIT will be provided. as this is the basis of our approach presented in section 5.

2.1 SME

In literature there is no standard definition of the term SME. Generally, there are different declarations on the number of employees in SMEs, to determine their size. Typically, enterprises with less than 10 employees are considered micro enterprises, a small enterprise has 10-49 employees, and enterprises with 50-250 employees are considered medium enterprises within EU guidelines [2]. Besides these restrictions on the number of employees the EU guideline also provides numbers on the turnover or annual balance sheet for categorization. A medium enterprise accordingly should not have a turnover larger than 50 million Euro per year (balance sheet 43 millions), a small enterprise does not have numbers larger than 10 million Euro and a micro enterprise has no larger turnover than 2 million Euro per year.

When looking at the information processing in SME following observation was made: SME tend to have more difficulties related to information overload than larger enterprises. This is probably caused by the limited amount of employees within the enterprise. Consequently, the individual employees have to process a wider range of information than specialists which can be found in larger enterprises. In addition, the qualification of the single employee is more important as there is no possibility for personal backups due to the sparse resources. Accordingly, knowledge should be stored or transferred electronically before an employee retires or falls ill. This is an urgent issue since SME often rely on a competency head start, which is no longer available if knowledge walks out the door. Consequently a need for KM in SME can be identified. [3]

2.2 Knowledge Management and Knowledge Management Systems

As for SME no official definition on the term of KM can be given, since the individual definitions vary widely. However, it should be noted here that two main approaches on KM exist, one human-oriented focusing on the transfer of knowledge between the employees of an organization, where as the other technical-oriented approach, aims at supporting the process of KM by the use of information and communication technology (ICT) [4]. Though both approaches have their limitations this paper will focus on the latter, since the first can go without any technical support at all and is not of interest here. Following this idea a definition of KMS can be given as follows ([5], p.86):

"A knowledge management system is an ICT system in the sense of an application system or an ICT platform that combines and integrates functions for the contextualized handling of both explicit and tacit knowledge, throughout the organization or that part of the organization, that is targeted by a KM initiative. A KMS offers integrated services to deploy KM instruments for networks of participants, i.e. active knowledge workers, in knowledge-intensive business processes along the entire knowledge life cycle. Ultimate aim of KMS is to support the dynamics of organizational learning and organizational effectiveness."

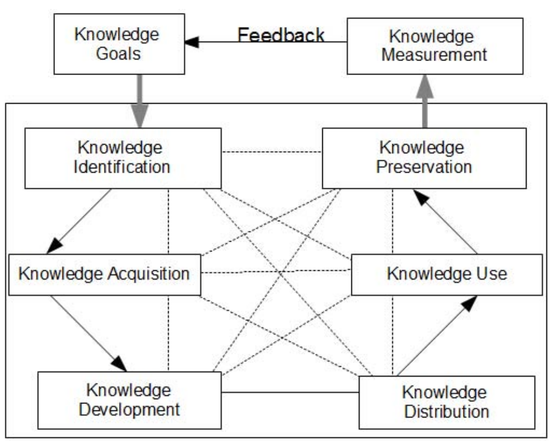


Fig. 1. Building Blocks of Knowledge Management [6]

With regard to the definition of Maier it can be seen that a KMS is supposed to be a part of a bigger KM initiative, meaning that there is more than the pure implementation of a technological system. However, this process should fit into the overall process of KM in an organization. This process itself should orientate towards the business goals as indicated by Probst et. al. [7], "The building blocks of knowledge management", were one block of the outer management cycle is called knowledge goals (see Figure 1). These knowledge goals should be derived from the business goals and strategy of the the concerned organization. Opposing this, one main point of critique in the model of Probst et.al. is that they do not offer a specific recommendation on how a KMS can support their KM cycle, their elaboration on this part offers no specific details concerning a concrete implementation. Furthermore, it is difficult to comprehend which building block can be supported in which manner by IT, a correlation between systems or applications and building blocks cannot be done one-to-one.

2.3 Business Value of IT

The business value of information technology (BVIT) includes all measures which focus on evaluating the effects of IT-changes on the business, especially performance, competitiveness, innovation and economic growth over longer time periods.

During the years numerous research has been conducted on this field of research and consequently different approaches for BVIT can be identified [8]:

- Process-oriented approaches describe the BVIT as means of improvement in business processes.
- The IS Success model of [9] is used as a perceived value approach, which explains the value as the change which can be perceived within several categories within the enterprise, whose improvement has a positive impact on the whole enterprise and therewith the business value.
- The project-focused approaches, try to justify the use of IT within projects. Accordingly, these approaches mostly focus on the impact of IT within a certain project.
- The last approach uses Scorecard structures, and accordingly tries to correlate different dimensions. However, this usually does not cover the holistic interrelations but only the ones with one central dimension as e.g. the cost for IT.

Nevertheless, KMS can be regarded as ICT according to Maier, and therefore can be somehow covered by BVIT. How this can be integrated with the business strategy of a SME will be explained in section 5.

3 KM in SME

This section presents the work done in order to comprehend the current situation of SME with respect to KM and KMS use. Accordingly, two main result were obtained, the first one by initiating a survey among knowledge-intensive SME with regard to the use they make of KMS and the second one by a systematic literature review on the topic in the renowned journals/ conference series. These results are complemented by the short introduction of two approaches explaining how KM can be implemented in SME, however even these do not provide concrete statements on what business value can be expected.

3.1 Knowledge Management in SME in Mecklenburg-Vorpommern

Looking at these suggestions on how to implement KM in practice and the reappearing question whether KM and KMS are really needed in SME a survey was conducted to document the recent state of the art in SME in our region.

This survey on KM and its application, especially in the form of KM systems was done at the Chair of Business Information Systems in Rostock. The survey was conducted in form of an online questionnaire, with focus on SME only. For this survey only knowledge intensive SME in Mecklenburg-Vorpommern (a federal state of Germany) were asked to fill out a questionnaire, which should provide us with a current state of the art in KM and especially focusing on IT support within these enterprises. The survey took place in November and December 2010. The according questionnaire was provided via an online platform and the link was sent to 596 enterprises, which fulfilled the criteria of being knowledge intensive according to [10]. The actual distinction on whether the enterprise was a SME or not was made based on the amount of employees, since an inquiry on the annual turnover was unlikely to be answered.

Out of these 596 enterprises 48 filled out the questionnaire, resulting in a response rate of 7,89% since one of the answers was given by an enterprise which could no longer counted a SME. Before elaborating on the results of the survey one fact on the distrubtion over the enterprise sizes: 6 answers were gained from medium, 14 from small and 27 from micro enterprises. Since these numbers do not fully resemble the official numbers for the distribution (micro: 88.84%, small: 9.02%, medium: 1.9%) for the area [11], as well as the fact that the basic population is not very big, an evaluation according to the different groups will not be presented here. Regarding the obtained general results several points were interesting.

- Firstly, only 9 out of the remaining 47 answering enterprises employ a systematic KM approach.
- Secondly, only 5 out of these 9 enterprises use a KMS to support their activities. When asking for the system itself the answers were: Wiki, Sharepoint2010, One Note, SVN and MediaWiki.
- Thirdly, one of the enterprises answered that it uses several applications to support KM and these do not have the possibility to exchange data. The other enterprises at least have this opportunity enabled.
- Fourthly, though KMS need a certain amount of administration, not all enterprises have a dedicated person for this task (4 out of 5).

With regard to these results it can be concluded that most enterprises still are not familiar with the actual meaning of the term KMS as introduced by Maier, and assume that any application that supports their handling of organizational knowledge is a KMS. Moreover a total of 80% of the enterprises in the survey do not practise KM at all. In the questionnaire we also asked for the goals which were pursued by the introduction of a KMS.

- The most common answers were transparency of knowledge, improving documentation, distribution of knowledge.
- When provided with a list of objectives, the participants considered almost every objective as very important, though the list contained 12 objectives.
- Yet when asked for the individual accomplishment of these goals, half the SME answered they achieved them to a high till very degree, whereas the other half indicated this degree was low.

Drawing conclusions it can be stated that SME are not fully aware which goals belong to KM and therewith focus on basic functionalities. And even more important, it does not seem to be clear which goal can be addressed with which IT solution. So SME want to improve, want to add to their business value and know that IT can somehow deliver to it. Yet, which solution supports precisely which goal needs further clarification.

In addition, the survey also questioned how KM if practiced was integrated into the enterprise. The obtained results here were.

- Firstly, none of the enterprises could name a specific strategy as suggested in literature for their application of KM.
- Secondly, only one of the enterprises actually writes down its goal for KM, and surprisingly these were not even controlled. All other enterprises rely on general oral statements on the goals for their enterprises, and control them by a regular personal estimation. Only one enterprise tries to control success by means of indicators.
- Thirdly, on the terms of how KM and KMS were introduced into the enterprise 6 answers said as a project, the rest mentioned a top-down approach.
- Fourthly, given the question of how much time their employees use for the fulfillment of KM tasks, 8 out of 9 enterprises answered less than 10%, two stated that it were even said less than 5%. Given 40 office hours a week this means less than 2 hours.

Interpreting these results it can be seen, that KM has not yet reached the SME in the area of Mecklenburg-Vorpommern. Some enterprises certainly got in touch with the concept and made up their own idea about it, yet a consequent alignment of KM and KMS to the business strategy was not found. However, the results on the objectives in KM indicate that a strategy is needed, as insecurity leads to a lack of priority setting.

3.2 Systematic Literature Review

Though KM is not widely spread in the enterprises of the region we looked for literature on KM in SME. However, by obtaining a systematic literature research we could see that only few papers on the KM in SME are available. This research was asking for approaches on how to introduce KM in SME published during the last 7 years in whether "Wirtschaftsinformatik" (WI, 2004-2010), "Practical Aspects on Knowledge Management" (PAKM, 2002-2010) and "Conference on Information and Knowledge Management" (CIKM, 2002-2010).

	WI	PAKM	CIKM
Years	2004-2010	2002-2010	2002-2010
Total amount	7	4	9
Amount papers	270	157	1671
Relevant papers	1	8	1

The overall result was rather poor, because even though the two conferences PAKM and CIKM were asking for contributions on the development of KM in SME only 10 out of 2098 contributions actually considered the topic. The search was performed with the keywords: knowledge (Wissen); learning (Lernen); experience (Erfahrung); expertise (Expertise), know-how (Know-how); wisdom (Weisheit); company (Firma); organization (Organisation); corporation (Gesellschaft), as well as KM and enterprise. The result was that out of 2098 papers only 10 were relevant for the topic, however 277 matched the searchwords in some way. Nevertheless, most of these final papers were addressing specific tools or applications in use in SME and not general ideas on how to implement KM in SME, or the question which application supports a SME best in its KM initiative. By this systematic approach no satisfying results were obtained, which showed recent experience with the problem of value-driven alignment.

3.3 Implementing KM in SME

Even though no specific results were found by systematic literature review, by using a more general literature research some approaches for KM in SME were found. Two of those will be presented here, since these were the ones including a technical part, whereas most other suggestions leave out technical details and focus on the introduction of the concept of KM only.

European guideline on good practice in Knowledge Management

After a workshop organized by the European Committee for Standardization in 2004 the results were published under [12], which can serve as a helpful tool through the implementation of KM in a SME. This guideline is divided into five parts, all concerning KM, yet only part 3 covers the implementation of a KMS in the enterprise. The first two parts motivate using KM in the organization at all, and provide a framework for all actions being part of the KM initiative, whereas the two remaining parts provide approaches to measure the success of the concerned KM initiative. The whole approach is based on the general phases of a software project as known from the waterfall model, yet here the initial phases cover the KM aspect, e.g. the question which knowledge is needed and for which goals. The goal is to specify suiting tools for the individual needs of the SME and afterwards also include a phase to estimate the success of the overall project, which is done by determining indicators early in the process. The focus lies on the project in KM including suiting tools, yet the guide mostly provides question catalogues, hardly any direct recommendation on the tools are given.

Knowledge management with available resources

This approach, described in [13], mainly focuses on introducing KM into the enterprise as a project by using prototypes. Accordingly this approach is not supposed to work without a technical component and mainly resembles the waterfall model for software projects with an extensive requirement analysis combined with prototyping. The main concern of this approach is to use already available systems or applications of the SME and create a portal to concentrate these. This creation of portal is accompanied by general workshops on the topic of KM to create e.g. awareness for the initiative and the demand of KM within the concerned enterprise. To support the project contracts between the project team and the enterprise are concluded. With regard to these contracts this method guarantees a successful implementation of the portal technology yet it includes no measurements to control whether the usage of the developed technology satisfies the initial need.

4 Lessons Learned

Summing up the results from the survey as well as from literature research, it became obvious that though KM was one of the buzzwords during the last years thorough common understanding of the concepts of KM and KMS is not yet established. The understanding and implementation of these vary widely. Even though several suggestions on the implementation of KMS exist (see section 3.3), none of these could be found explicitly in the practice of SME. In addition, especially with regard to the limited financial resources of SME an holistic approach as suggested by Maier [5] appears to be too excessive at the beginning. Moreover, no direct connection between the individual enterprise's strategy and the KMS architecture is made, there it is no plausible documented way on how to build up a KMS architecture to suit the business strategy, as it is suggested by Probst et. al. with the building block "knowledge goals". Following that description this building block is supposed to be the reference point where enterprise goals and knowledge goals are to be matched. Yet Probst et.al. provide no further recommendations on how this alignment could be realized.

When looking at the fact that SME prefer smaller applications, since they are easier to install and administrate, one way to approach a thorough alignment of IT to business strategy in the field of KM might be reduced approaches for initial implementations [14]. They should be smaller and easier to apply, in their overall design more like a sandbox: easy to use and to build up continually something that supports the overall success of the enterprise. Possibly this could in the end resemble the architecture as introduced by Maier, but only if all components and knowledge services are of value for the individual enterprise's strategy. One step necessary towards this is to show the individual value of certain groups of applications as further described in section 5.

5 Value-Driven Alignment of KMS

With regard to the conclusion further research will have to focus on the fact how certain applications which can be found in the area of KM can provide business value for certain parts or processes within SME and the question on which price does this value can be achieved. Recently the business values of single applications are analyzed. This analysis is done considering three main questions: Firstly, are the benefits a the implementation of certain applications restricted to special types of SME? Secondly, which specific value is added for instance by a Wiki to the enterprise? And finally, which strategy can be used for successful alignment to business strategy and embedding into business processes.

Accordingly the first step to proceed with these questions, is the determination of the business value of KMS in general. Though there are concepts available on this topic, see e.g. [15] a general reduction to a number indicating the success of a system appears to narrow an approach. Regarding the definition of a KMS as given by Maier it can be understood as an IT system and consequently the approaches of BVIT can be taken into consideration.

Accordingly, the idea is to focus on individual applications supporting the knowledge services as suggested in Maier's architecture and find out how they contribute, first to the individual business strategy of the enterprise and second to the knowledge service. Alongside the question which knowledge service is especially helpful for which kind of enterprise has to be clarified, since based on that knowledge the application to start with can be chosen. Of course, KM is supposed to support the individual aspects without neglecting the overall holistic approach of KM, however, the business value in comparison with the costs of an application is not clarified by now, as well as the question whether or not a real KMS is needed. Yet, we assume that KMS can be of value for SME since some studies have already shown the impact of a KM application an the enterprises' success [1]. Using the example of a Wiki, this means, assuming that Wikis mostly support the knowledge service "publication" ignoring at this place that they have a strong discussion functionality, that it has to be found out what kind of SME takes most advantage out of this KM application.

Our Framework

Looking at the gap identified in section 3.1, further research focuses on investigating how certain applications in the field of KM can deliver business value. For addressing this problem we propose a three dimensional framework for the evaluation of the value creation of KMS in SME.

The overall goal of this framework is to identify the KM technologies offering the best value combination according to the business objective of a SME. The first dimension of the framework should address the aspects and manifestations of value, since KMS can be seen as ICT certainly approaches of BVIT can be regarded.

Here working the concept of BVIT (business value of information technology) is possible, as described in section 2.3. Considering the approaches of BVIT named, not all can be considered appropriate here. We are looking at these approaches as a source for attributes in this first dimension, which have to suit the KM environment. With regard to the immaterial nature of knowledge the precise determination of numbers related to costs cannot be the objective here, accordingly the perceived value approach seems an appropriate candidate to fill this dimension. These leaves us with the perceived value as suggested by [9], which is also promising for its high user orientation. And since KM is considered to have to be highly user-oriented this appears to be suitable. Regarding the approach of DeLone and McLean, the suggested information use might be splitted into knowledge and information use and especially combined with user satisfaction the resulting impact on the organizational impact might be a good way to show how the individual KM solution can align to the enterprises strategy.

Moreover the project-focused approaches can be excluded as KM is supposed to be a long term integral part of the enterprise, which cannot be covered project-focused only. And when having a look at the Scorecard approaches, these can be excluded as we do not want to spend a major effort in correlating the attributes within this dimension but to the other dimensions. And consequently even the process-focused approach is not suitable here as many SME do not have their processes modeled and it cannot be our objective to make that process of modeling a prerequisite here.

The second dimension deals with the tasks and possible architectures for KMS. One possible approach here could be the architecture of Maier. This architecture of an holistic KMS distinguishes four central knowledge services (see Figure 2, layer III) and suggests to integrate the individual solutions via portals to allow one single point of access for the user. The services are documentation, search, collaboration and learning, which each cover a certain functionality, but Maier still leaves room for further functionalities not covered by one of these services. However, building up such an architecture is rather resource consuming (in means of time as well as money), and appears to be a highly complex task to accomplish for a SME. This leads to the question whether individual services might be sufficient, and by which means an individual service can support the enterprises business strategy.

Finally the third dimension should show the cause effects between the two afore described dimensions. These effects traditionally were only shown between the different aspects of business value in the first dimension and are to be extended here to interrelations in dimension 2 as well as to the relations between 1 and 2.

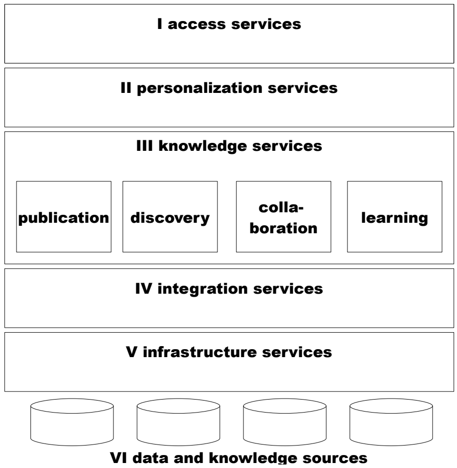


Fig. 2. Knowledge Management System Architecture according to [5]

Prospects

The goal of the suggested framework is to identify the most effective KMS technology, at the best cost ratio for individual business objectives of a certain SME. Consequently, our next step towards a working framework is the population of the framework with attributes within the dimensions. therefore a mapping of the individual goals of the SME is necessary, as well as a mapping between the dimensions of the framework. Moreover, the resulting fields have to be filled through further surveys or projects with SME to find specific results on specific applications within a SME according to a certain objective. Only after this population was done, it is possible to derive solutions for supporting applications by consulting the framework.

References

1. Salojärvi, S., Furu, P., Sveiby, K.-E.: Knowledge management and growth in Finnish SMEs. Journal of Knowledge Management 9(2), 103–122 (2005)
2. The new SME definition: User guide and model declaration (2003), http://ec.europa.eu/enterprise/policies/sme/facts-figures-analysis/sme-definition (accessed April 06, 2011)
3. Lamieri, L., North, K.: Wissensmanagement in Klein- und Mittelbetrieben. Praxis Wissensmanagement (2001)
4. Maier, R., Hädrich, T., Peinl, R.: Enterprise knowledge infrastructures: With 50 tables. Springer, Berlin (2005)
5. Maier, R.: Knowledge Management Systems: Information and Communication Technologies for Knowledge Management, 3rd edn. Springer, Heidelberg (2007)
6. Probst, G., Raub, S., Romhardt, K.: Managing knowledge: Building blocks for success, reprint. edn. Wiley, Chichester (2002)

7. Probst, G., Raub, S., Romhardt, K.: Wissen managen: Wie Unternehmen ihre wertvollste Ressource optimal nutzen, 6th edn. Gabler Verlag / GWV Fachverlage GmbH, Wiesbaden (2010)
8. Sandkuhl, K., Tellioglu, H., Johnsen, S.: Orchestrating Economic, Socio-Technical and Technical Validation using visual modelling. In: 16th European Conference on Information Systems: ECIS 2008, National University of Ireland (2008)
9. Delone, W.H., McLean, E.R.: The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. J. Manage. Inf. Syst. 19, 9–30 (2003)
10. Legler, H., Frietsch, R.: Neuabgrenzung der Wissenswirtschaft - forschungsintensive Industrien und wissensintensive Dienstleistungen (NIW/ISI-Listen 2006). Fraunhofer ISI, Karlsruhe (2007)
11. Statistics agency Germany. Enterpriseregistry (2007), http://www.regionalstatistik.de (accessed April 27, 2011)
12. European Guide to good Practice in Knowledge Management: CWA 14924 (2004)
13. Feggeler, A., Lensing, W., Mühlbradt, T., Schieferdecker, R., Strothotte, M.: Wissensmanagement mit Bordmitteln, Bachem, Köln (2007)
14. Borchardt, U., Grap, F.: E-learning Application Support for SME. In: Forbrig, P., Günther, H. (eds.) Perspectives in Business Informatics Research. LNBIP, vol. 64, pp. 62–72. Springer, Heidelberg (2010)
15. Tseng, S.-M.: Knowledge management system performance measure index. Expert System Application 34, 734–745 (2008)
16. Bullinger, H.J., Bullinger, J.: Managing innovation networks in the knowledge-driven economy. International Journal of Production Research 42(17), 3337 (2004)
17. Wong, K.Y., Wong, Y.: Development of a knowledge management initiative and system: A case study. Expert Systems with Applications 30(4), 633 (2006)
18. Wong, K.Y., Aspinwall, E.: An empirical study of the important factors for knowledge-management adoption in the SME sector. Journal of Knowledge Management 9(3), 64 (2005)

SCOR Based ISS Requirements Identification

Liene Liepina and Marite Kirikova

Faculty of Computer Science and Information Technology,
Riga Technical University, Latvia
{liene.liepina,marite.kirikova}@rtu.lv

Abstract. Information systems requirements should be rooted into the business needs. Business process models are one of the most transparent ways how to represent these needs. Knowledge about business processes in different types of industries has already been accumulated in several process frameworks, such as Supply Chain Operations Reference framework, enhanced Telecom Operations Map, Value Chain Operation Reference framework,, etc. The frameworks are mainly used in process redesign, and improvement. However, they are also the source of information systems and services requirements. While the role of the frameworks as part of enterprise architecture has been widely discussed, there is still a lack of research on the use of frameworks in information systems and services development. This paper attempts to theoretically prove the usability of business process modeling frameworks in information systems and services development by demonstrating two approaches of the use of Supply Chain Operations Reference framework in information systems and services requirements identification.

Keywords: Framework, SCOR, requirements, information systems, services.

1 Introduction

Nowadays different business process frameworks are a common tool in business process analysis, modeling and design. However, the use of frameworks in information systems and services (ISS) requirements identification is less researched. While frameworks differ in how the processes are structured; in their conceptual granularity at different levels of abstraction, in conceptual meaning of the elements of the frameworks, and in the way how they represent ISS related issues, all of them include information that is useful for ISS requirements definition. The amount of this information depends on the features of the frameworks: e.g., concerning inputs and outputs of processes, there are frameworks that show inputs and outputs of all or a part of processes (e.g. VRM (Value reference Model) [1]), there are frameworks that do not show them (e.g., APQC (American Productivity and Quality Center) process classification framework [2]), and there are frameworks that provide some information with respect to inputs and outputs of the elements (e.g. SCOR - Supply Chain Operations Reference framework [3]).

In this paper we use SCOR framework to present the first results of the on-going research on the utilization of business process frameworks for ISS requirements identification. To prove usability of frameworks for ISS requirements identification

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 232–243, 2011.

we propose two approaches of requirements definition using SCOR framework. Both approaches demonstrated in the paper root the requirements into the business process defined on the basis of SCOR. Thus both approaches seek alignment between the business process and the ISS requirements. While the resulting requirements identified by each approach may differ considerably, both approaches show usability of business process frameworks in ISS requirements definition.

The paper presents results on use of SCOR in the context of small and medium enterprises (SMEs), however they may also be applicable for larger enterprises.

The paper is structured as follows. Section 2 is devoted to related work and surveys different uses of SCOR framework in research and practice. In Section 3 the experimental business case expressed in SCOR terms is presented. This business case is used in Section 4 and Section 5 respectively to propose and present two approaches of requirements definition. In Section 6 the results obtained by the requirements definition approaches are compared and discussed. Section 7 consists of brief conclusions.

2 Use of SCOR Framework

The SCOR framework (model) builds on the concept of process reengineering, performance, measurement, and logistics management [4]-[12], [13]-[15], [3]. It is based on 5 processes, namely Plan, Source, Make, Deliver, and Return. It considers all processes at four levels of abstraction:

- The first is Top Level – it looks to major 5 processes and describes process types.
- The second is Configuration Level – it puts all process types in the process categories.
- The third is Process Element Level – it looks deeper in process categories and describes the steps to be performed to execute level 2 processes. It is important that the sequence in which these processes are executed influences the performance of level 2 processes.
- The forth is Implementation Level – it shows industry specific activities required to perform level 3 processes. Level 4 processes describe a detailed implementation of the process. SCOR does not detail level 4 processes. Organizations and industries develop their own level 4 processes.

The SCOR model identifies financing activities and associated risks [10], it links performance metrics, processes, best practices, and people into a unified structure [15]. The framework includes also Enable processes which are not discussed in this section.

Many researchers point to the *benefits of the use* of SCOR framework. In [4] the following advantages of its use are listed:

- Improved speed of deployments
- Faster and greater return on investments
- Common metrics across organization directly linked to processes
- ‘Benchmarkable’ metrics with other reference model users in industry
- Reference model serves as a neutral ‘common language’
- SCOR facilitates easier gap assessment
- SCOR gives an opportunity to use documented best practices derived from thousands of users.

Source [5] comments that SCOR Model provides a step-by-step process that corresponds to the organizational goals, the required training and the support needed to successfully achieve these goals.

The benefits of SCOR framework usage are tangible and intangible. Intangible benefits of it s use include mature improvement, collaboration and team work as well as internal and external processes with fully shared issues and expertise. It also provides a communication platform, standardization and ERP migration to the new information systems [9].

In line with benefits several *challenges* are also reported, since the use of the framework is usually related to changes in organizations and corresponding change management challenges. It is not an easy to introduce SCOR because it requires having specific knowledge, time, and investments and in the beginning it is not possible to see the result unless specific simulation tools are used [13]. There can also be difficulties in problem analysis, in respecting particular organizational constraints, and in taking into consideration experts' opinion [12]. Nevertheless many enterprises had overcome these difficulties and use the framework without big problems (e.g. Proctor & Gamble, IBM, Nokia, Toyota, J&J, Samsung, Wall-Mart, Tesco, Johnson Controls [6]). This suggests that there are possibilities also for other enterprises to use SCOR framework that might be relevant especially because of emerging globalization of enterprise businesses.

In several cases the successful use of SCOR framework is related to use of specific *tools* such as iSCOR [8], ARIS EasySCOR, IBM SmartSCOR [13], (KOS) OWL model [16].

Currently only large companies are reported to build their information systems based on SCOR framework. For instance, ADVA Optical Networking reports reduction in IT costs through minimizing system customization and making better use of standard functionality, Douglas Pharmaceuticals Limited reports that investment in Tooling improves 50 %. SAAB shows that, with use of the framework, information systems are key enablers in utilization of benefits of SCOR [15]. However, there is a lack of research on specific approaches that could be used for framework based ISS requirements definition on regular basis. For achieving ISS requirements definition benefits for different enterprises including SMEs, the first step is to define methods that help to identify requirements for ISS that are necessary for the support of the processes that are defined using SCOR framework. The approaches proposed in the remainder of the paper are the first step towards such methods.

3 The Experimental Business Case

The requirements definition approaches proposed and discussed in this paper are demonstrated by the example of a small supply chain consisting of a state institution (Institution of Government) which every year is making a calendar for the next year; Typography which prints the calendars; Computer Graphics Enterprise that designs the calendar; and external suppliers and customers. In this section we illustrate the supply chain using artifacts that are prescribed by SCOR [4] guidelines.

The Business Scope Diagram (Fig. 1) shows a small supply chain in which:

- Computer Graphic Enterprise is getting photos for making design (1)
- It offers different version of the calendar to State Institution (2)
- After accepting the version (3) the Enterprise gives it to Typography (4)
- Typography is getting paper and materials (5)
- Calendars are delivered to the Institution for the realization (6).

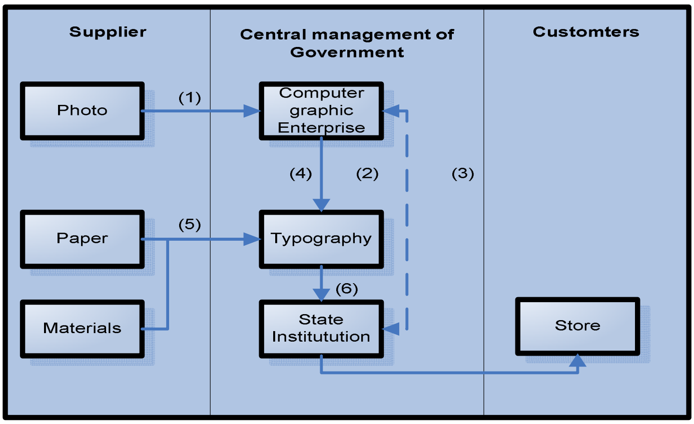


Fig. 1. Business Scope Diagram

In Fig. 1 information only flow is shown by a dotted line, while material and information flow is shown by a solid line.

SCOR Configuration Toolkit [14] was applied to develop all further artifacts reflected in Fig. 2-4 and Table 1.

Figurative illustration of a geographic map is given in Fig. 2.

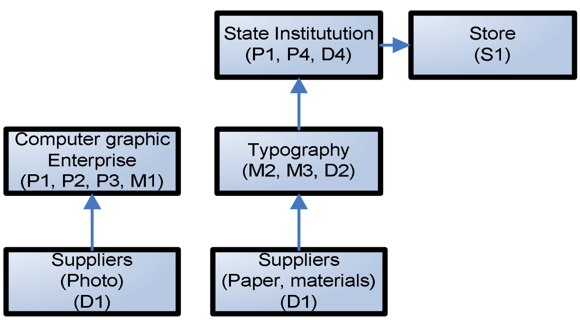


Fig. 2. Figurative ilustrtation of Geogrpahic map

A fragment of Process element thread diagram is shown in Fig 3. The fragment is constructed by hiding several elements of original thread diagram. Only a small subset of elements that are most relevant for describing ISS requirements identification approaches is shown in Fig 3.

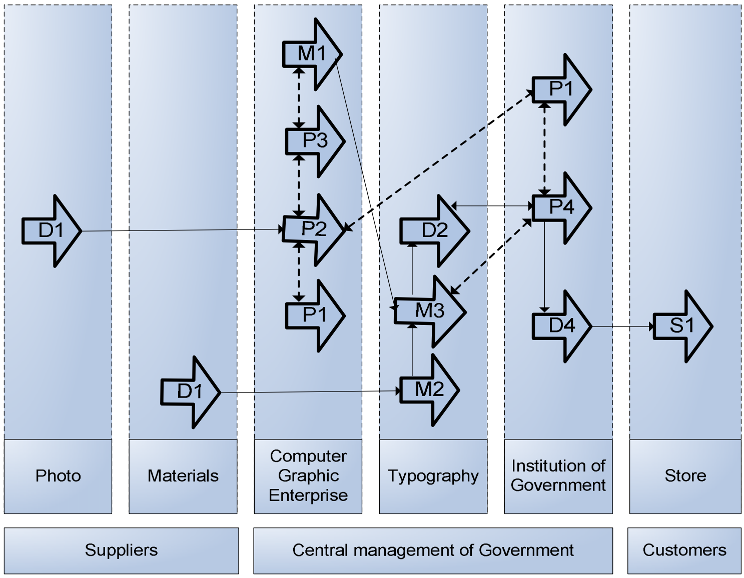


Fig. 3. Thread Diagram

Processes (workflows) are shown in Fig 4. The names of the processes are described in Table 1.

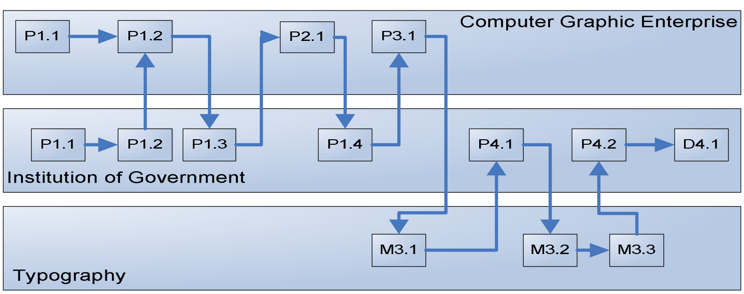


Fig. 4. The process model

Table 1. Sub-processes coresponding to the process model shown in Fig. 4

Computer Graphic Enterprise	Institution of Government	Typography
P1.1. Create Templates P1.2. Create Offer for Market P2.1. Create Examples P3.1. Develop Final Version	P1.1. Make Instruction P1.2. Make Purchase P1.3. Sign Contract with Enterprise P1.4. Accept Finished Version P4.1. Sign Contract with Typography P4.2. Test the Final version of Calendar D4.1.Deliver Calendar	M3.1. Get Accepted Version of Calendar M3.2. Sign Contract with Institution M3.3. Make Calendar

Table 1 and Fig. 4 represent the business process diagram for the supply chain in which the Institution of Government is involved. Organizations in the supply chain depend on one another not only because of the exchange of goods, materials and money. Information exchange between institutions is to be taken into consideration, too. The requirements for information systems may be partly derived from the processes defined on the basis of SCOR in two ways: (1) by analyzing the processes at the third level of SCOR (Fig. 4) and by analyzing them in the top down manner via the whole hierarchy of the framework. The first option is discussed in more detail in Section 4; the second - in Section 5.

4 SCOR 3rd Level Based ISS Requirements Identification

In this section we propose an approach for ISS requirements definition, based on the third level of SCOR processes. At this level it is possible to view the supply chain processes in an integrated manner and it is possible to consider processes of each enterprise separately, too. Regardless of the viewpoint the approach is the same. It focuses on possible inputs and outputs of the processes as well as on key performance indicators or metrics obtainable from the SCOR framework [17]. Defining ISS requirements on the basis of business processes is a quite common approach in Model Driven Development [17] and Service Engineering [18]. Methods from these areas can be adopted for SCOR based requirements identification. Therefore in this section we will focus only on the issues that are not commonly considered in business process based requirements identification, namely (1) frameworks based input and output analysis and (2) metrics analysis.

SCOR 3rd level based ISS requirements identification approach consists of the following phases:

1. For each process identify connecting processes and corresponding outputs and inputs
2. For each identified output and input decide whether it requires maintenance of particular information about it. If yes – identify details of data requirement.
3. For each process decide which metrics are to be used for.
4. For each chosen metrics analyze the following:
   a. Is the data needed for calculation of value of the attribute available among the identified ISS requirements? If yes – mark it as a metrics and proceed further; if – not add new information requirement and mark it as a metrics.
   b. Define metrics target/alert value request requirements for each data entity marked as metrics.
5. Define functional and non-functional ISS requirements using use case analysis or other requirements definition techniques [e.g., 18, 17].
6. Check whether all data identified in phases 1-4 are considered by functional requirements. If not – return to step 5, if yes, proceed to requirements validation.

The results of phases 1 to 4 are illustrated in Table 2 where inputs, outputs, metrics and data requirements are shown for four processes.

Table 2. Output from phases 1-4 of SCOR 3rd level processes based ISS identification

Proc	*Inputs*	*Outputs*	*Metrics*	*Data requirement*
Create Template	Customer requirements; Order backlog plan; Data revised (optional); Aggregate forecast and projections revised (optional); Business assumptions; Shipments; Supply Chain requirements	Description of errors; Report of last actions; Report of lost data; Information of state holydays days – add/edit/delete; Delivery data; Information of pictures; Order information; Delivered date	Forecast Accuracy; Identify, Prioritize, and Aggregate Supply Chain Requirements Cycle Time; Cost to Identify, Prioritize, and	All items from Inputs and Outputs of *Create Template*; Forecast accuracy; Supply chain requirements cycle time; Cost to identify, prioritize and aggregate supply chain requirements.
Create Example	Bill of materials; Delivery plans; Order signal; Planning data; Production plans	Data of pictures; Data of state holydays; Descriptions of Usable Templates, Terms; Limits; Alarms; Last data for completed template	Identify, Prioritize, and Aggregate Product Requirements Cycle Time; Cost to Identify, Prioritize, and Aggregate Product Requirements Forecast Accuracy	All items from Inputs and Outputs of *Create Example*; Product requirements cycle time; Cost of identifying, prioritizing, and aggregating products requirements forecast; Measure of accuracy
Accept Finished Version	Supply Chain Plans; Information about Finished version	Supply Chain account time; Acceptance results	Establish Supply Chain Plans Cycle Time; Cost to Establish and Communicate Supply Chain Plans	All items from Inputs and Outputs of *Accept Finished Version*; Supply Chain Plans Cycle Time; Cost to Establish and Communicate Supply Chain Plans
Get Accepted Version of Calendar	Engineering design methods; Procedures, Processes order information	Length of service; Account information	Finalize Production Engineering Cycle Time; Cost to Finalize Production Engineering	All items from Inputs and Outputs of *Get Accepted Version of Calendar*; Production engineering cycle time; Cost to finalize production engineering
Deliver Calendar	Daily replenishment requirements; On-demand replenishment requirements; Resource availability; Shipping schedules; Stocking requirements; Stocking schedule vendor/DC inventory availability	Time to deliver; Quality service; Car duty; Security service; Forecast time; Real time; Cost of delivery; Planning time; List of inventory; Expire date	Generate Stocking Schedule Cycle Time; Cost to Generate Stocking Schedule;	All items from Inputs and Outputs *of Deliver Calendar*; Stocking schedule cycle time; Cost to generate stocking schedule;

The name of the process (see Table 1) is represented in the first column. In the second column the inputs of the process are represented and in the third column the outputs of the process are shown. the metrics of SCOR relevant to the process are listed in the column “Metrics”.

The last column of Table 2 amalgamates date requirements from Inputs and Outputs of processes as well as from analysis of metrics where additional data requirements can be identified. The table represents two types of Input and Output requirements: (1) directly transferred from SCOR framework, and (2) tailored semantically to the business case. Analysis of metrics give an opportunity to derive other requirements (not presented in the table), for instance values for some of metrics that can be incorporated in ISS or used as non-functional requirements, e.g., interface requirement for providing space for inputs for desired Supply Chain Plans Cycle Time or displaying calculated Cost per calendar. The identified requirements may serve as the background for further requirements elicitation.

5 Top-Down SCOR Based ISS Requirements Identification

In the top-down approach the ISS requirements identification starts at the first or second level of SCOR processes (Fig. 1-3). It is performed using the following guidelines (Fig. 5):

1. Start with the top level SCOR artifacts and look for the basic rules that are needed to perform the processes in the enterprise. Ask the question – can computerized ISS do any tasks in this process? Look at this issue starting at the first level of SCOR and continue up to the third level.
2. If there are no basic rules which can be translated to ISS, look deeper in the process – try to merge it with other processes. Ask the question – are there more benefits than investments in introducing ISS?
3. At the moment when the rule or processes that can be converted to ISS are found - take all information about them to the lower levels of SCOR framework.
4. Collect all processes which are related to potential ISS with input and output data (see SCOR 10.0 processes [15]).
5. Analyze their inputs and outputs and identify ISS requirements.
6. Use Metrics to get values for Requirements.

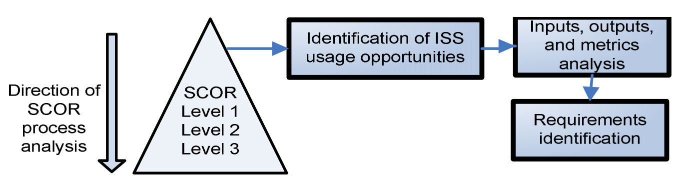


Fig. 5. Illustration of the top-down approach

Last two guidelines may be implemented using the approach proposed in the previous section. If the level of analysis is deeper than the third level then the approach similar to the SCOR 3rd level based approach can be applied.

Suppose top-down analysis of SCOR artifacts in Thread diagram has shown (Fig. 3) that there are processes which can be transformed into ISS (Fig. 6). Actually it appears that all processes of Computer Graphics Enterprise can be transferred into the ISS. From this at a high level of abstraction the following ISS requirements can be stated:

1. ISS must maintain the templates.
2. ISS must check correspondence to market requirements.
3. ISS must store and show examples of the calendar.
4. ISS must support the whole calendar development cycle from the first to the last version.

Fig. 7 illustrates how the decision to introduce ISS is taken to the next level of SCOR processes.

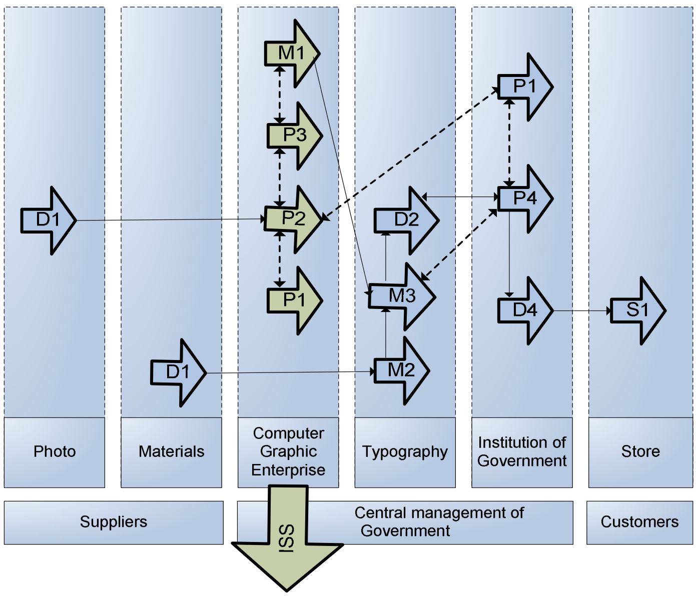


Fig. 6. ISS suportable processes identified in the thread diagram

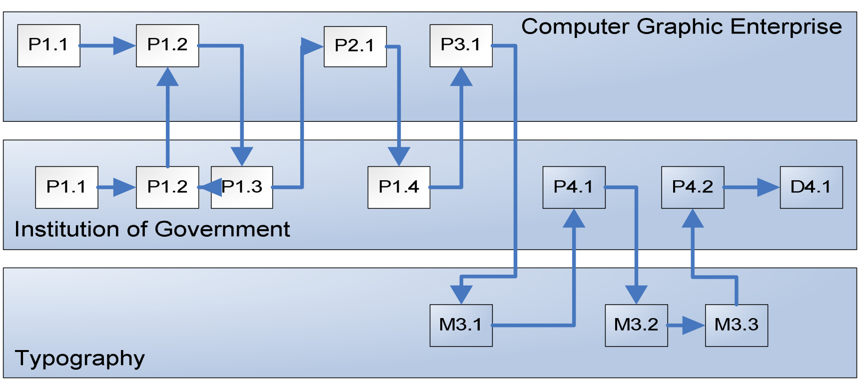


Fig. 7. Propagating the decision to lower levels (ISS processes highlighted)

Requirements 1 to 4 is the starting point for defining more detailed requirements. All processes transferable to ISS (highlighted processes in Fig. 7). These processes can be analyzed using the SCOR 3d level based requirements identification approach (Section 4). Thus for Computer Graphics Enterprise's processes Create Template and Create Example we would get requirements similar to ones reflected in first two rows of Table 2. However, in this approach requirements for Institution's of Government process Deliver Calendar are not be considered.

6 Comparison and Theoretical Evaluation of Approaches

In Sections 4 and 5 we presented two approaches of the use of SCOR framework for ISS requirements identification. Both approaches have the following common features: (1) analysis of inputs and outputs of the 3rd level SCOR processes, and (2) analysis of metrics. The approaches have the following differences:

- **Scope of 3rd level processes under input, output, and metrics analysis:** the SCOR 3rd level based approach considers processes of the whole supply chain and provides the background for decision making with respect to which information is to be included in ISS; on the other hand, the top-down SCOR based approach consider only those processes that have been identified at upper levels of analysis. In case there are no processes identified at upper levels, the top-down approach becomes similar to the 3rd level based approach.
- **Applicability:** the top-down approach is more applicable in situations where there is an intention to introduce new ISS, the third level based approach can be used in all cases. However, it might be less efficient than top-down approach in situations when particular decisions about ISS introduction have already been made.

Both approaches help to identify ISS requirements since they give an opportunity to define a subset of requirements directly from the information available in SCOR framework. Thus good practices amalgamated in SCOR can be indirectly introduced to any ISS project for enterprises involved in supply chains. Especially important it is for SMEs which in most cases do not define their business processes formally and do not apply SCOR in their business process management.

At this stage of research only above the mentioned theoretical evaluation of the use of frameworks in ISS development is available. Practical evaluation has just started and final results are not yet obtained.

7 Conclusions

Global economy forces companies to engage in local and international value and supply chains. Therefore they need flexible ISS that supports their frequently changing processes. A great amount of business process knowledge is amalgamated in business process frameworks. These frameworks can be used as the source of ISS requirements. In this paper we propose two approaches of use of SCOR framework for ISS requirements definition. The approaches give an opportunity to identify requirements in line with decision making about changes in business processes. These first results of the research show the following benefits of the use of frameworks in ISS requirements definition:

- Knowledge amalgamated in business process frameworks (in this case SCOR) can be used not only in business process improvement and design, but also in information systems and services development
- In both proposed ISS requirements identification approaches use of the framework facilitates to transparently consider business processes and ISS as related systems
- The use of the frameworks in ISS requirements definition is beneficial even if the business processes of the enterprise are not defined in advance (see section 5).

At present the approaches mainly consider data requirements of ISS. We plan to utilize more information from SCOR for requirements definition. In their current stage of development the proposed approaches have been applied only in the context of SMEs as they prescribe manual analysis of the artifacts produced during their application process. Specific visualization tools would be needed for their use in large enterprises. The approaches are demonstrated on SCOR framework only; it is a matter of investigation whether similar methods can be applied with VCOR [1], APQC process classification framework [2], and other process frameworks.

References

[1] VCOR, `http://www.value-chain.org/en/cms/4/`

[2] APQC Process framework, `http://www.apqc.org/process-classification-framework`

[3] SCC & SCOR Overview Engen. Joseph Francis, Executive Director SCC (2010)

[4] SCOR and Benefits of Using Process Reference Models. Thomas Phelps. Supply Chain Strategy and Development Manager, HP 2006 Chair, Supply Chain Council Board of Directors, January 12 (2006)

[5] Workflow Improvements and Cost Benefits Using the SCOR Model. William Frank Quiett, C.P.M., A.P.P. Wm. Frank Quiett Associates, 91st Annual International Supply Management Conference (May 2006)

[6] Measuring SCM benefits and SCOR Model. Suhas Range, Director (Planning&Admin) NMIMS University, Mumbai, `http://www.scribd.com/doc/7136865/Measuring-SCM-Benifits-SCOR`

[7] The SCOR Model. An Implementation and Analysis by Merck Serono. Ferhat Oegretmen, Global Data Manager, Merck Serono MS-TSE, Munich, October 26 (2010)

[8] Aligning the Supply Chain Operations Refernce (SCOR) with Enterprise Applications: Real-time Value Chain Intelligence. Thomas Gulledge and Tamer Cavusoglu, Plicy Analysis Center – Enterprise Engineering and Policy Program George Mason University and Thomas Kessler, PROMATISm Corp, San Ramon CA

[9] Irfan, D., Xiaofei, X., Chun, D.S.: A SCOR Reference Model of the Supply Chain Management System in an Enterpise. STH eInterantional Arab Journal of Information Technology 5(3) (July 2008)

[10] Lup, Z., Wong, E., Tong, F., Tan, C.J.: Warehouse Financing Risk Analysis for SMEs in China. In: 2010 International Conference on Service Sciences E-Business Technology Institute, The University of Hong Kong (2010)

[11] Ling, H.: Effect of IT on Supply Chain Outcome: From an Inter-organizational Relationship-specifc View. In: International Conference on E-Business and E-Government, School of Management, Fudan University, Shanghai, China (2010)

[12] Wang, W.Y.C., Chan, H.K., Pauleen, D.J.: Aligning business process reengineering in implementing global supply chain systems by the SCOR model. International Journal of Production Research 48(19) (October 1, 2010)
[13] Dong, J., Ding, H., Ren, C., Wang, W.: IBM SmartScor – A SCOR based supply chain transformation platform through simulation and optimization techniques. IBM China Research Laboratory (2006)
[14] Supply Chain Council, SCOR version 9.0 overview. Online-Archive (2008), http://supply-chain.org/
[15] Supply Chain Council, SCOR version 10.0 overview (2010), http://supply-chain.org/
[16] Zdravkovic, M., Panetto, H., Trahanovic, M.: Towards an Approach for Formalizing the Supply Chain operations. In: I-Semantics 2010 (2010)
[17] Harmon, P.: Business Process Change: A Guide for Business managers and BPM and Six Sigma Professionals, 2nd edn. Elsevier, Morgan Kaufmann Pubishers (2007)
[18] Brown, A.W.: MDARedux: Practical Realization of Model Driven Architecture. In: Seventh International Conference on Composition-Based Software Systems, 0-7695-3091-5/08 © 2008 IEEE, pp 174-183 (2008)

A Review of Information Logistics Research Publications

Darek M. Haftor[1], Miranda Kajtazi[2], and Anita Mirijamdotter[2]

[1] Stockholm University, School of Business,
SE-106 91, Stockholm, Sweden
dh@fek.su.se
[2] Linnaeus University, School of Computer Science,
Physics and Mathematics, SE-351 95 Växjö, Sweden
{miranda.kajtazi,anita.mirijamdotter}@lnu.se

Abstract. 'Information Logistics' has presented itself as an intellectual and professional domain addressing the question of timely providence of the right information. A question that emerges then is: What is Information Logistics? To answer this question, a comprehensive review of research publications was conducted, where 'Information Logistics' was featured in the publication title. A detailed analysis of the content of these publications identified eleven different research directions, where five are currently active, all in Europe. Among various findings, the results show that these research directions have been pursued independently of each other, addressing different kinds of research questions and contexts, utilising different research approaches, and therefore generating a variety of unrelated research results. All the reviewed research here shows that there are numerous unmet empirical needs in our human and social affairs, as well as a need for intra-disciplinary developments, which calls for a joint mobilisation of the research efforts.

Keywords: Information Need, Lack of Information, Literature Review, Content Analysis.

1 Introduction

The Tsunami Catastrophe in South East Asia in 2004, the recent earthquake in 2009 in Italy, the Challenger Space Shuttle Disaster in 1986, a patient being amputated the wrong leg, the Enron Bankruptcy in 2001 are only some dramatic examples of the numerous situations in our societies where the lack of accurate information, at the right time, in the right format, provided to the right actor, gives rise to unwanted consequences. 'Information Logistics' has presented itself as an intellectual and professional domain to address the question of timely providence of accurate information. A question that emerges then is: *What is Information Logistics?* To answer this question, our team conducted a comprehensive review of scholarly publications that included 'Information Logistics' in its title. This paper presents the outcome from that review and its analysis. The result presents eleven identified

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 244–255, 2011.

Information Logistics (IL) research directions, where five are currently active, all in Europe. The volume of research publications is steadily increasing and new domains of application emerging manifest its relevance for our human, industrial and societal needs. However, the Information Logistics research is highly diversified and fragmented, sometimes having the only common feature: the label of Information Logistics. This makes us call for a joint effort to collaborate and establish an accumulation of a theoretical body.

Section two presents briefly the Research Approach employed in this study, while section three presents a summary of the results obtained from this review. The fourth section discusses these findings and derives some implications, while the final section provides a short summary.

2 Research Approach

In order to answer the here posed research question of what is information logistics research about, a comprehensive literature search and review was conducted. The search was conducted through the conventional search engines provided at a university library, followed by a reference list reviews of the retrieved texts.

As mentioned above, the review was driven by the identification of research publications that feature ‘Information Logistics’ in its title, which implies a concept-driven research approach. The review was limited to publications in three languages: English, German, and Swedish. The analysis of the retrieved publications was guided both by J.P van Gigch’s meta-modeling research approach [1], and by Kuhn’s notion of a scientific paradigm [2]. The specific questions of inquiry included: what is the research object, what is the research methodology employed, what are the received theoretical bodies, what are the research results, and what are the basic presuppositions of the research approach; all this with a critical review of strengths and limitations.

The research design presented here includes several limitations. One is its explicit focus only on research publications that have ‘Information Logistics’ in its title; clearly there are research publications that address aspects of Information Logistics yet do not mention information logistics in their title. However, this limitation was made on purpose to reduce the research efforts needed that would otherwise require multiple resources compared to available. Secondly, the research conducted here was concept-driven, meaning that it started from a given concept – Information Logistics – and then it explored its various meanings. This is a common approach utilized in literature reviews. However, the limitation is that very different and unrelated phenomena may be covered with the same concept. This requires careful attention in the research process, which was assumed here. A final limitation here is that although more than one hundred references were identified; only three quarters were successfully retrieved; particularly some of the German-language references were not possible to retrieve through the international university library system, mainly older publications however. It is reasonable to assume though, that the retrieved and analyzed publications are representative to the whole population.

3 Results

The result of the conduced review of 'Information Logistics' publications is presented in this section. This is organized into two main sub-parts: firstly is the brief account of the volume and kind of publications analyzed; secondly is the content of each research frontier identified from the review of these publications.

3.1 Identified Kinds of Publications

The present review succeeded to identify 108 publications featuring 'Information Logistics' in its title. Out of these 71 (≈ 65 %) were retrieved and consequently analyzed. The earliest publication identified originates from 1978, from Harvard University, USA. From the beginning of 1990's the volume of publications increased and more than doubled between 2000 and 2009. Various kinds of publication formats and channels are utilized, where the conferences are the dominating channel. Thirdly, two languages of publications dominate: English and German, which is reflected in that Germany is the single country where majority of the authors' institutions are present.

The following conclusions may be derived from this account. Information Logistics research has been published for more than three decades; the last decade shows a significantly growing number of publications; Information Logistics research utilizes all the typical kinds of publication channels of a scholarly discipline, and Germany is the dominating country of Information Logistics research.

3.2 Research Directions Identified

The content of the reviewed and analyzed Information Logistics publications were organized into *eleven research directions* with regard to their research topic addressed. Each of these identified research directions is given a short account in the following.

Firstly is the '*Local Distribution of Information*' (1978), originating from the Harvard University, USA. In this, Information Logistics is understood *"…as a function of the business enterprise, devotes primary attention to the production, storage, packaging, and movement of information (products). With this in mind, information logistics will, for our discussion, refer to the management of all activities which facilitate information (as a product) movement in order to supply customers with the place and time utility in information goods and services they demand."* [3]. The empirical concern here addresses the so-called information industries, such as broadcast and cable television, broadcast radio, telephone, mobile radio systems, motion pictures, organised sports, theatres, computer systems, postal services, private information delivery services, newspaper, periodicals, business consulting services, advertising and marketing, brokerage industries, book publishing and printing, libraries, banking and credit, insurance, legal services. The key empirical concerns in this research proposal include the emergence of new type of information services, the increase of costs of information provision, change of information demand due to demographic transformation, and the development of information and communication technologies. While no new information logistics knowledge is provided here, the key merit of this research direction stems from its formulated research questions that seem to still be valid; these include: what are the crucial problems associated with local distribution of information?; what are the costs associated with local distribution of information?; what are the trade-offs among various modes of information delivery?; and: what are the regulative and legal issues to be considered? For example, the latter question shows its relevancy particularly this 2010 year, as the European Union decided to de-regulate the postal services within its member countries. Only one publication was found within this research direction, where the only received theoretical foundation is the so-called General Systems Theory [4]. The presented study elaborates the newspaper industry by means of statistical analysis of secondary data, hence a positivist approach. The results presented are somewhat speculative and suggest the following trends: decline of the volume of newspaper readers; increased costs of newspaper distribution; potential of joint delivery of newspapers and magazines; new technology for the delivery of news to homes; purposeful handling of information flows will increase physical goods distribution; and barriers between current industries will transform. In summary, the key merits of this research direction seem to be its early focus on Information Logistics phenomena and the formulation of a set of generic research questions that still seem to be valid for research purpose.

The second research direction is called the '*Information-Production Flow-Time*' [5], [6]. The empirical concern here is the limited productivity increase within the information-producing organisations – e.g. banks, insurances, movies, legal, marketing, and sales. In the beginning of the 1990's, the difference in performance improvements between blue and white collar work was significant; the gained productivity in offices was only 3% in the 1980's compared with an improvement of 75% in factories. The research strategy assumed here is somewhat unique, namely to investigate how theoretical bodies from goods logistics, which thus became the received theoretical body, may be transferred to the domain of information logistics. Therefore, the research questions are how to shorten Information-Production Flow-Time and how to establish relevant control of information-production-processes. This research resulted in three distinct proposals: a model for the information-production time-elements, a modelling language of flow-time elements in organisations, and a model for the control of information-production processes. Again, the merits of this research direction seem to be the articulation of the research challenge: the productivity limitations in organisations where the product is information-based not physical goods. At least three limitations may be mentioned here,: the danger of theoretical transduction, not critically reflected, may lead to reductionism; the selected root metaphor reduces an organisation to a manufacturing-flow, and information to a physical entity; and the lack of benefit studies and data. The presented models are based on a minor sample of case studies within the context of action research.

The third research direction is the '*User-demand information-supply*' [7], [8], [9], [10], and [11] among others. This is definitively the most productive research direction within Information Logistics, as identified here, measured both in terms of the number of publications and also in terms of the research outcomes transferred into industry and organisations. Its home base is the *Fraunhofer Institute* for Software and Systems Engineering, in Germany, particularly at the Technical University in Berlin, however with affiliation elsewhere in Germany and outside, such as Sweden, China, USA, Poland, and Russia. Two research concerns are addressed here; one is the coordination of information transfer, from a source to its destination, second is the provision of information that is needed by an actor, whether human or machine. This research direction as a whole focuses on the design and development of software artefacts to improve the two empirical challenges for organisations and its individuals. In the first case, it is multi-agent software that retrieves and conducts semantic matching of information needs vs. information available [12]. In the second case, software and various modelling tools are conceived to identify automatically the needed information and to filter and provide the right information [13], [14], [15]. Three key advances of the latter include technology for the so-called role-based information provision, then situation-based information provision, and finally the context-based information provision. The received theoretical bodies come mainly from computer science and particularly software engineering. The research results may be summarised as conceptual frameworks for information demand identification, information matching, information transfer, and database representations. Some of these outcomes have found implementation in various industrial software applications, such as traffic information systems, weather warning information for the insurance industry, the tourist industry, applications for journalists, automotive and construction industries. The strength of this successful research direction of Information Logistics includes its addressing of empirical and instrumental needs in organisations, its well defined and focused research program and as a consequence the productive research outcome and its accumulative research progress; also the experimental approach employed secures relevancy of research outcomes. The limitations, on the other hand include, that the research does not investigate and provide any benefit studies, that although the empirical needs of the human actors are addressed, only software solutions are provided disregarding other than mechanistic aspects, such as the social, political, economical, psychological, legal, ethical. Further, only information-needs and information-transfer is addressed, not information generation. This research approach clearly builds on a mechanistic root metaphor, where a human being is assumed to rationally know what information is needed prior its reception, that it is possible to identify information-demand prior information-reception, that a machine can detect the information-demand, and that the identified information-demand can be satisfied by available information in various computerised databases – all these assumptions are clearly difficult to justify.

The fourth research direction of Information Logistics is the '*Efficiency of Information-Flow*', which also has its home in Germany, in this case at the Frankfurt University [16], [17]. The concern here is an optimisation of information flows in information networks, as applied to the information supply chains in the automotive industry. The basic problem addressed thus is the lack of information-flow efficiency, which leads to production and supply inefficiencies due to unwanted interruptions caused by the lack of the needed information at the right moment. In this, the Information Logistic system is understood as a network of nodes linked with some channels for information transfer between these, including their standards or protocols, and then the information flow pattern that emerges in such a network depending on how the information flow is directed. Two models are advanced for an optimal information flow in such networks. The first assumes that all network nodes, i.e. various companies in the automotive supply chain network, are homogenous, and the goal is to find the optimal cost of information flow there. Key operational questions include: how many nodes and intermediaries, what channels and what information-flow patterns, and what economic effects are derived? The results generated is an economic model of information-flow that covers all related tasks; e.g. selection of communication paths, channels, and nodes. The algorithms developed are then applied to the German automotive supply chain, where the information flow cost was estimated annually to app. € 80 million and the application of the proposed optimisation algorithms would imply an 80 % reduction [17]. The key limitation of this optimisation model is its assumption that all actors in an information exchange network are homogenous – this is seldom an empirical case, and is not valid for the automotive supply chain, where information asymmetry and opportunistic behaviours is an inherent property. Therefore, the secondly advanced model assumes that the network nodes are heterogeneous, i.e. autonomous, self-interested, and utility maximizing. The goal here is then to find the optimal coordination of information-flow allocation. In this, it is recognised that it is not possible to find an optimal coordination of information-flow allocation that is centralised for the whole information network, i.e. supply chain. Therefore the research strategy selected assumes the utilisation of decentralised algorithms. The result generated is an information allocation algorithm based on an update mechanism that maintains a weak consistency of replicated information in the network. This allows increased efficiency with maintained quality of information-flow for decentralised and distributed computing, based on local information. The received theoretical bodies here are mainly the Transaction Cost Theory; the Network Effect Theory; and several Operations Research Optimisation Algorithms, as well as some basic Information Logistics Frameworks. The research approach utilised includes empirical data collection and its statistical analysis, which means the positivist epistemology. The assumed root metaphor seems to be the organism, i.e. a Darwinist Adoption of the global behaviour with local actors. The strengths of this research direction include the impressive economic implications of the models developed, that conceptually regarded Information Logistics systems are not reduced, nor bound, to information and communication technology, the methodological rigour, and the unique macro-system perspective. The limitations include the few publications generated from this research program, the assumption of the biological root metaphor directing toward a global optimum as desirable.

The fifth research direction of Information Logistics, here identified, is the '*Cross-Functional supply of Analytical-Information*' [18], [19], [20] with its home at the University of St. Gallen, in Switzerland. The key empirical challenge here seems to be: how to realise synergies in an organisation, when different actors and functions collaborate, by means of information exchange and provision between these actors and functions. The approach assumed to reach that end is the search for process-oriented Information Logistics in organisations, which have embedded analytic information and/or analysis capabilities into the context of organisational processes. In this, two levels of investigations emerge: one is the enterprise-wide Information Logistics infrastructure, enabled by information and communication technologies, and the second is the business processes, particularly those not fully automated. The research methodology assumed, builds on hypotheses formulation, survey-based data collection from organisations, followed by statistical analysis, all aimed to find so-called best practices in organisational behaviour – this constitutes a positivistic approach. One key research result generated, is a survey that classified organisations with regard to their IL practices, whether IL is centralised and integrated into processes, or focusing on data system quality, or standardisation of IL applications, or sophistication of Information Logistics Strategy. The balanced approach includes all these to a certain degree, which is said to be optimal for the organisation [19]. Another research contribution is the elaboration of an Information Logistics Strategy, which is understood in three terms: the sourcing approach, the delivery model, and the portfolio strategy [20]. The received theoretical bodies in this research stream include various conceptual frameworks for the governance of information and communication technologies, and also some supply chain frameworks. The strength of this research includes its empirical orientation and research methodology employed for the search of best practices as well as cost and benefits of investments into information and communication technologies. It also highlights the importance of the right information for human decision-making. A key limitation here, however, is bound to the specific notion of Information Logistics that this school assumes, namely: "*Information logistics (IL) comprises the planning, control, and implementation of the entirety of cross-unit data flows as well as the storage and provisioning of such data. In order to differentiate IL and operational data integration, only those data flows are considered to be IL components which support decision making. If data is used for decision making in the same organizational unit where it originates, such flows do not fall under our IL definition because in this case, most of the managerial challenges do not occur.*" [20] The proposal that Information Logistics may only address information flows, that are cross-functional and for decision-making only, is clearly arbitrary and unjustified, which is implicitly challenged by all the other research directions of IL, here presented. We see no reason to omit the information flows that do not happen to be cross-functional and that are not aimed at decision-making, as the latter's challenges and benefits may be just as important. A second limitation here seems to be the basic research assumption that an organisation's global optimum may be reached by means of cross-functional information provision, as this is supposed to facilitate cross-functional co-ordination. Clearly, this may be the case in some instances, however organisational sociology as well as strategic management has shown that this does not need to be the case.

The sixth research direction of Information Logistics is the '*Process-Improvement through Information-Flows*', with its home at the Nyenrode Research and Innovation Institute, in the Netherlands [21], [22]. As is the case with the previous direction, the key concern here is that the lack of needed information creates organisational inefficiencies. Various proposals are advanced to this end, including a procedure for business process re-design, an approach for the information-flows improvement within the health care, and also a model for the evaluation of information access technologies, such as the Google search engine on the Internet. Research wise, this approach may not be called a rigorous scholarly knowledge production, it leans more toward highly applied research, almost firm consulting. The research methods are not declared, and in practice employ a more eclectic and common-sense making approach. No empirical research is presented and no explicit received theories are visible. The strengths of the outcome of this Information Logistics research group is its highlighting of the importance of information-flow metrics and also showing the value of Information Logistics solutions for the health-care organisations. This Information Logistics group manifests a more fragmented approach, addressing various domains that are not clearly related. The results and its recommendations are unjustified and unmotivated, sometimes simplistic and not evidence-based.

The seventh research direction of Information Logistics is the '*Information Logistics for e-Maintenance*', with its home at the Luleå University of Technology, in the northern Sweden [23], [24]. Its key concern is the provision of the right and timely information about a complex entity – e.g. aircraft, transportation vehicle – to the service module, in the context of the maintenance operations of such an entity. The overall challenge is the concern for the life cycle of such an entity, its costs, quality, or safety matters. e-Maintenance, understood as the application of information and communication technologies to various maintenance activities, is assumed to constitute a solution for these concerns. The received theoretical bodies are mainly various conceptual frameworks from computer sciences and particularly software engineering, and also some generic maintenance operation frameworks. The research approach chosen seems to be engineering and experimental, particularly in the context of aircraft maintenance operations. The results seem to be the focus on the data content management, in terms of its format and structure, to enable efficient data exchange between heterogeneous information technology systems and their different contexts. The strengths of this research direction include its attention to the opportunity that Information Logistics may provide to future maintenance operations, the need to focus on the data contents and its interoperability, and also the recognition of the various stakeholders for such data, including the human decision maker. The limitations here stem from the implicit assumption that complex operations management and its information logistics may be understood in software engineering terms, and an appropriate configuration of various existing software engineering standards and artefacts will generate an appropriate solution.

Finally, the present investigation has identified four additional Information Logistics research initiatives. Compared to the above presented, however, these do not represent an outcome of a research program, rather minor research attempts or ad-hoc wise elaborations and comments. The following presents these four directions very briefly only.

The eighth research initiative is called here the '*Outsourced Information Handling*', presented in research paper from the Washington University, USA [25]. The empirical challenge addressed is the costs of data management in organisations, where high volume information exists. The solution proposed is to outsource all the electronic data handling capability to the so-called Application Services Providers. The approach assumed here is a kind of common-sense elaboration, without any explicitly received theoretical bodies, or empirical investigations, only some literature review. The main merit from this initiative seems to be its attention of the high data processing costs in organisations.

The ninth research initiative is the '*Information-Flows in Supply-Chains*' conduced at the Halmstad University College, in Sweden [26], [27]. This initiative addresses the industrial supply chains and starts with the observation that the efficiency gain realized through increased efficiency of physical goods flows is not enough; it is now time to increase the efficiency flows of information as well. In consequence, the overall research question assumed is: what factors can improve and rationalize information flows in supply-chain-oriented organisations? The research approach assumed includes explorative and empirical-based approach, mainly with interviews and also several case studies. The received theoretical bodies typically come from a comprehensive literature review on information flows in supply chains. The results provided, include observations that are specifics of supply chains malfunctioning, with regard to their information flows, such as that there is a lack of information-flows between sales and marketing, the needed information is provided too late to the needing agent, and the wrong information is provided to the needing agent. The strengths of this research initiative seem to be the defined research question and also its explorative and empirical research approach. The limitations are, however, that the research results are not very informative or original and also that the notion of IL is tightly bound to Logistics and Supply Chain Management, which is a limited notion compared to the other presented research directions of Information Logistics.

The tenth initiative identified here is the '*Work-flow Modelling*' from Växjö University in Sweden [28]. The key empirical concern here seems to be that in order to provide the right information to the user of some information and communication device, there is a need to represent this information-need in a way that a machine can read and understand it. The assumed approach is the modelling of the context of the information demanding agent – man or machine. The received theoretical bodies are several methods for conceptual modelling, also known as ontological modelling. The results are generated through a conceptual elaboration of the received theories versus the question of investigation, hence no empirical research is provided. The results constitute some general guidelines for how to conduct work-flow modelling, including work-activities, story telling, and role specification. Further, the very modelling process is recommended to be participative, i.e. to include the stakeholders that typically are or will conduct the workflow at hand. The key strength of this contribution seems to be its attention to the content of the human and machine actors in a work-flow situation, as well as the participative modelling approach highlighted. On the other hand, the limitation is that there is no real novel contribution; the guidelines provided are today common knowledge of the Information Systems Development literature, nor that there is any empirically based justification of the elaboration exercised.

The eleventh initiative for Information Logistics research comes from one publication originating from Norway, and is labelled here as '*Global Cross-Reference Database*' [29]. Its empirical concern is that currently, most databases have different formatting and nomenclature standards. This, in turn, makes information exchange difficult, causing slower operations and costs, hence operational inefficiencies and quality problems. The contribution of this research is the proposal to create a global cross-reference database for all major nomenclatures, aiming to facilitate information cross-referencing and therefore information exchange. The key merit of this proposal is its attention to an important issue that makes information logistics inefficient in and between organisations and their actors. However, no scholarly research is presented to either elaborate the empirical issues or to advance its resolution.

4 Discussion and Implications

It is now time for a collective analysis of the above presented research initiatives. Starting with the research objects and the empirical dilemmas addressed, a diversity may be noted, where some initiatives have their locus of research on a whole industry and its information flows, several initiatives focuses on the lower level of an organization and its various processes, while still other initiatives focuses yet at a lower level of data and software. Also, within each such a category, different questions are addressed. This shows the research review here is highly diversified and uncoordinated. The same may be noted with regard to the kind of received theories that are assumed, that range from transaction-cost theory, through optimization algorithms, to various software engineering conceptual frameworks. As a consequence of this, different research methodologies are assumed, including the typical positivistic approach with hypotheses formulation, survey-based data collections and the statistical analysis, through more hermeneutic case studies, to typical engineering studies based on experiments and conceptual elaborations.

Yet another partial-diversity is that each of the research initiatives presented has its own notion of what is information logistics about. However, underneath, there seems to be an implicit agreement on some commonalities, such as that it is about the timely and accurate provision of the needed information to the needing actor, whether human or machine. A second common characteristic is that all the reviewed research initiatives of Information Logistics include information and communication technologies as a key component. Some regard this technology only as an enabler, which is conceptually disregarded in theoretical elaborations, while other explicitly focus only on some aspect of it, typically the software.

As we understand it here, there are at least two major shortcomings with the Information Logistics research as a whole. The first is that the given contributions focus extensively on the information needs, and information provision, however very little, if any, elaboration addresses the source of information, its generation. The limitation of this is that a successful information logistics operation may only be established if a compete information value chain is conceptualized and addressed, from the information generation, through its transfer, transformation, storage and also consumption, aimed at some value generating aspect.

The second shortcoming identified is that Information Logistics research addresses the very ‘information’ in its syntactical manner, disregarding its semantic quality: the meaning. At least since the seminal work of Winograd and Flores [30], we know that a syntactical-approach to information and its context as well as technology, is limiting and hinders us from exploring the full potential of an information society.

Finally, only five of the eleven Information Logistics research initiatives are currently active, (as far as it is known), all in Europe, which is not plausible. However, research that addresses the challenges of Information Logistics is conducted under other labels within various disciplines, such as computer sciences, industrial manufacturing and engineering, business administration and economics, psychology, social-psychology and sociology, as well as semantics and information studies.

5 Summary

Information Logistics, understood as the timely provision of the right information to the needing actors – human or machine – shows its importance frequently in media, when incidents caused by the lack of needed information lead to unwanted consequences: the 2004 Tsunami catastrophe in South East Asia, or the financial collapse of the 2001 Enron bankruptcy. This text presents a review of published research that explicitly addresses 'Information Logistics'. The findings here show that Information Logistics as an intellectual domain is highly diversified, fragmented, and uncoordinated, which is a typical characteristics of a young and underdeveloped research domain. On the other hand, the volume of publications is steadily growing and new domains of applications are addressed in human, social and industrial affairs. This dichotomy of empirical relevancy and intellectual fragmentation calls for joining the separated efforts through a closer collaboration and exchange, aiming to establish the accumulation of an intellectual body that address the pertinent needs of information society.

References

1. van Gigch, J.P.: System Design Modeling and Metamodeling. Plenum Press, New York (1991)
2. Kuhn, T.S.: The Structure of Scientific Revolutions, 2nd edn. University of Chicago Press, Chicago (1975)
3. Wormley, P.W.: Information logistics: Local distribution (delivery) of information. J. Quart. 55(1/9) (1978)
4. von Bertalanffy, L.: General System Theory. George Braziller, New York (1969)
5. Gerrits, J.W.M., Sijbrands, M.J.C.: Information Logistics towards Logistical Concepts for Information Production. Research Memorandum, Vrije University (1992)
6. Gerrits, J.W.M.: Towards Information Logistics – an Exploratory study of Logistics in Information Production. Ph.D. Thesis, Vrije University (1995)
7. Deiters, W., Heuwinkel, K.: Smart-Wear: A personalized information broker. In: Conference Proceedings of ICEWES, Cottbus (2002)
8. Heuwinkel, K., Deiters, W., Königsmann, T., Löffeler, T.: Information Logistics and Wearable Computing. In: Proceedings of the 23rd International Conference on Distributed Computing Systems, Washington (2003)
9. Sandkuhl, K., Smirnov, A., Weber, H. (eds.): The Knowledge Gap in Enterprise Information Flow: Information Logistics concepts and technologies for improving information flow in networked organisations, Jönköping School of Engineering, Center for Information Logistics, Ljungby (2004)
10. Haseloff, S.: Context Gathering - an Enabler for Information Logistics. In: Deiters, W., Kutsche, R.D., Sandkuhl, K. (eds.) Knowledge Supply and Information Logistics in Enterprises and Networked Organizations Proceedings of the International Workshop Associated with the Multi-Conference on Wirtschaftsinformatik of the Gesellschaft für Informatik e.V. University of Duisburg-Essen, Duisburg (2004)
11. Lundqvist, M., Sandkuhl, K., Seigerroth, U., Stirna, J.: Method Requirements for Information Demand Analysis. In: Proceedings of the 2nd International Conference on Adaptive Business Systems, Glasgow, pp. 15–21 (2008)

12. Smirnov, A., Pashkin, M., Chilov, N., Levashova, T., Haritatos, F.: Knowledge Source Network Configuration Approach to Knowledge Logistics. I. J. of Gen. Systems 32, 251–269 (2003)
13. Sandkuhl, K.: Information logistics in networked organizations - selected concepts and applications. In: Filipe, Cordeiro, J., Cardoso, J. (eds.) ICEIS 2007. LNBIP, vol. 12, pp. 43–54. Springer, Heidelberg (2008)
14. Meissen, U., Pfennigschmidt, S., Voisard, A., Sandkuhl, K.: Situation-based Message Rating in Information Logistics and its Applicability in Collaboration Scenarios. In: Proceedings of the 30th Conference on EUROMICRO, Rennes, pp. 484–491 (2004)
15. Lundqvist, M., Sandkuhl, K., Levashova, T., Smirnov, A.: Context-Driven Information Demand Analysis in Information Logistics and Decision Support Practices. In: American Association for Artificial Intelligence. AAAI, Pennsylvania (2005)
16. Grolik, S.: Information Logisitcs – Decentralization Approaches of Information Allocation in Information Exchange Networks. Ibidem-Verlag, Hannover (2007)
17. Fricke, M.: Information Logistics in Supply Chain Networks. Concept, Empirical Analysis, and Design (Paperback). Ibidem-Verlag, Hannover (2007)
18. Lahrmann, G., Stroh, F.: Towards a Classification of Information Logistics Scenarios - An Exploratory Analysis. In: Proceedings of the 42nd Hawaii International Conference on System Sciences, pp. 1–10. IEEE Press, Hawaii (2009)
19. Bucher, T., Dinter, B.: Process Orientation of Information Logistics – An Empirical Analysis to Assess Benefits, Design Factors, and Realization Approaches. In: Proceeding of the 41st Hawaii International Conference on System Sciences, p. 392. IEEE Press, Hawaii (2008)
20. Dinter, B., Winter, R.: Information Logistics Strategy - Analysis of Current Practices and Proposal of a Framework. In: Proceedings of the 42nd Hawaii International Conference on System Sciences, pp. 15–17. IEEE Press, Hawaii (2009)
21. Olthof, W., Willems, J.: IL Quadrants for Information Access Technology. Working Paper, Nyenrode Research and Innovation Institute, Nyenrode University (2008)
22. Willems, A., Willems, J., Hajdasinski, A.: Information Logistics Research Report: Frameworks in the Healthcare Industry. Working Paper, Nyenrode Research and Innovation Institute, Nyenrode University (2009)
23. Karim, R., Candell, O., Söderholm, P.: E-maintenance and information logistics: aspects of content format. J. of Quality in Maint. Eng. 15, 308–324 (2009)
24. Candell, O., Karim, R., Söderholm, P.: eMaintenance - Information Logistics for maintenance support. J. of Robot. and Computer-Integrated Manuf. 25, 937–944 (2009)
25. Chaffe, B.C.: Information Logistics: Outsourcing Solutions Can Improve Productivity. White Paper, Graduate School of Business, University of Washington (2001)
26. Chibba, A., Hörte, S.Å.: Information and physical flows in supply chains. Research Report, Centrum for Information Logistics, Ljungby (2003)
27. Chibba, A., Rundquist, J.: Mapping flows - an analysis of the information flows within the integrated supply chain. In: Proceedings of the 16th Annual Conference for Nordic Researchers in Logistics, NOFOMA 2004, Linköping, pp. 137–153 (2003)
28. Flensburg, P.: Information Logistics and Content Management - buzzwords or important concept? In: Karlsson, C., Flensburg, P., och Hörte, S.-Å. (eds.) Knowledge Spillovers and Knowledge Management, vol. (2004), pp. 476–491. Edward Elgar, Cheltenham (2004)
29. Simonsen, G.H.: The Global Cross-Reference Data Base - A Tool for Information Logistics in e-Business. In: Fifth eGov-conference, Antwerp, pp. 635–645 (2005)
30. Winograd, T., Flores, F.: Understanding Computers and Cognition: A new Foundation for Design. Addison-Wesley Company Inc., Massachusetts (1987)

Challenges for Electronic Identity Documents

Mirosław Kutyłowski and Anna Lauks-Dutka

Wrocław University of Technology
{miroslaw.kutylowski,anna.lauks}@pwr.wroc.pl

1 Motivation for Electronic ID Cards

A personal ID card is a certified copy of some personal data which proves the authenticity of the data in an offline setting. Majority of countries either plan or already are running projects aiming to deploy personal identity cards equipped with an electronic chip. Apart from a more reliable protection against forgery, this should give an opportunity to enable electronic services implementing the idea of e-government. In this scenario, personal identity cards have to serve as secure devices for running cryptographic protocols such as authentication, proof of presence, signing electronic documents, etc. In practice, there is a no universally accepted and clear concept of how to use electronic identity cards. Moreover, decisions are often based on myths (e.g. concerning security of smart cards) and tradition (e.g. the concept of qualified signatures) rather than on technical facts and feasibility. This lack of concept resulted so far in numerous strategic mistakes, concerning both the technical layer and business model.

2 Remote Identification

The validity of personal data form an eID card can be confirmed in two different ways. A naïve solution assumes that each data item is digitally signed by the card issuer. This solution provides a very high data confirmation level, that enables data collection and selling and facilitates creation of copies of the state registers. The second approach is from the area of zero knowledge proofs (ZKP). It assumes that data is authentic since it comes from an authenticated chip. Moreover, the data transferred between the chip and the terminal is encrypted with a session key, which prevents its modification and eavesdropping. General authentication procedure in this model consists of three phases: authentication of the owner of the card, terminal authentication and chip authentication.

PACE (Password Authenticated Connection Establishment) is a good example of the first step of authentication. It is a modified version of Diffie-Hellman key agreement protocol (proposed by BSI) that provides secure communication and password-based authentication. It should be stressed that the password is not used as a PIN to unlock the device (terminal). Instead, it is used to derive a session key that is later applied to encrypt communication. A huge advantage of this scheme is that a different challenge is used in each authentication attempt, which prevents replay attacks. On the other hand, the weakness of the password is an issue. It cannot be too long as it would be difficult for the user to memorize, and there is no other way to improve its entropy. Consequently, a password guessing attack is possible.

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 256–257, 2011.

Terminal authentication (TA) is necessary before the terminal gets any non-trivial data from the card (such as a digital image of the owner's face or fingerprints) and before the card starts any important protocol (like installing a qualified signature by the terminal, protocols confirming presence for medical transactions etc.). Chip authentication (ChA) is necessary to check the authenticity of the eID card and its presence and to confirm data transmitted later by the chip. It is technically possible to implement TA and ChA in such a way that no replay attack is possible and the transcript cannot serve as a proof for a third party that the interaction took place. The order in which TA and ChA are executed may be fixed (we can observe some kind of a German-French war over this matter).

There are different challenges for remote identification, depending on the area of application. According to the travel document inspection directive (ICAO) no terminal authentication is executed. In other words, a biometric passport shows the data just to anybody. On the other hand, an eID (because of the personal data protection) must not reveal the personal data to unauthorized terminals.

3 Restricted Identification

Another privacy issue of an eID card is the owner's anonymity while using some e-government services. The main idea of restricted identification (RI) is to hide the identity of the user while binding the rights with the physical person at the same time. To achieve that, RI splits activity areas into independent sectors (like health care system, citizen-police interactions, local referenda etc.) and provides a different identifier for authentication for each sector. Basic property of this method is unlinkability of different sector's identifiers that belong to the same person. It should be stressed that the typical pseudonym solutions are not enough in this case due to sybil attacks and identity transfer. For RI implementation different techniques and architectures are possible (sometimes specific for particular application areas). So far, there is no universal solution. Austrian Buergerkarte for example assumes that the card computes a password for each sector from a personal number and the sector identifier. Then a central password verification similar to PIN number verification of bank cards is used. The disadvantages of this solution are static passwords and the possibility of impersonating of the password owner by the recipient. German version of RI assumes that when logging into a sector, the eID card computes a unique password for the sector. Then the terminal of this sector service provider checks if it is "talking" with an eID card, receives a password and checks the password against the sector's blacklist (the list of excluded users). Cryptographic techniques guarantee that an eID card can not generate two different passwords for the same sector, so the blacklist approach is effective. A very strong data protection and strong unlinkability of passwords from different sectors is also provided. In Wroclaw University of Technology we develop an approach somewhat similar to the German scheme. We consider however the management of users with a "whitelist" (list of legitimate users) and/or "blacklist" and assume that each time a different password is generated so the terminals do not have to be trusted.

Diagnosis of Multi-Agent Systems and Its Application to Public Administration

Alexander Boer and Tom van Engers

University of Amsterdam/Leibniz Center for Law
{A.W.F.Boer,T.M.vanEngers}@uva.nl

Abstract. In this paper we present a model-based diagnosis view on the complex social systems in which large public administration organizations operate. The purpose of diagnosis as presented in this paper is to identify agent role instances that are not conforming to expectations in a multi-agent system (MAS). To this end, we introduce model-based diagnosis of an imperfectly observable multi-agent system. We propose the model-based diagnosis problem as an explanation of major driving forces behind policy making, and requests for change to IT and business process design departments, in public administration. This makes model-based diagnosis a useful legal knowledge acquisition model for public administration.

1 Introduction

In previous papers [3,2,4] we have characterized the problem of adapting business processes and decision support software in public administration, in response to changing society and government policy, as a design & diagnose problem solving cycle, based on the observations on problem solving made by [6]. Although few would argue that large organizations are designed, with *agents* as the components, little has been done to address complex social systems as a model-based diagnosis problem. Localization of fault is however an important part of a sterotypical legal problem: we look for *someone* responsible for norm violation. In enforcement scenarios, we search for a reasonable *responsibility assignment* to an agent, and subsequent punishment of, or remedial action by, that agent. Problems that threaten public administration performance also involve localization of agents causing performance problems, leading to design-based remedies that can be characterized as *component repairs*; resource reallocation, redesign of software, revision of guidelines, etc.

In this paper we explain model-based diagnosis in section 3, and apply the conceptual model to the complex social systems in which large public administration organizations operate. The purpose of diagnosis as presented in this paper is to identify problematic agent role instances in a multi-agent system (MAS). To this end, we introduce the model-based diagnosis problem of a diagnostic agent having a model of an observable multi-agent system to be diagnosed in its environment in section 4. We introduce the *agent role* as component of the multi-agent system, and agent role descriptions as models of normal, functional components and abnormal, faulty components. The predicament in law is that the diagnostic agent is *part* of the multi-agent system, and

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 258–269, 2011.

has only limited access to the information exchanged in the multi-agent system. The diagnostic agent's ability to control and change the multi-agent system is limited, and manipulations to obtain information have side-effects in the system.

We pesent a subjective version of the model-based diagnosis problem, based on diagnostic agent role descriptions, as a model for the acquisition of diagnostic compliance knowledge in public administration. In the paper we use the domain of real estate property transactions, from the point of view of a tax administration, as an example in section 4.2.

This work is part of a project with the Dutch tax administration and the immigration service. These organizations have a lot of legal knowledge in the form of *critical incidents* and, for want of a better word, *noncompliance storylines*, that has no natural place in traditional forms of legal knowledge representation, like normative rules or legal argument schemes, but is important for monitoring and enforcement. A central place in these stories is taken by stereotypical agent motives like evading income taxes, and these stereotypical agent motives tend to remain relevant, even though policy is made to put a stop into specific legal gaps.

2 Related Research

Fig. 1 presents the functional classification of generic tasks in public administration that we proposed in [3,2]. The organization principle for this classification is functional dependency between knowledge roles, inspired by the typology of problems and views on problem solving first presented two decades ago in [6]. The *suite of problem types* in [6] was based on an analysis of the problem and task decompositions found in then-state-of-the-art knowledge-based system literature [7,8,21]. The suite of problem types presents us with a generic problem solving cycle, and two different vocabularies for describing it, depending on the type of model of the domain that is available:

1. *Model* → *Design* → *Implement* → *Monitor* → *Diagnose*
2. *Classify case* → *Plan* → *Execute* → *Monitor* → *Assessment*

Design in public administration can be approached from the first perspective. We therefore argue for multi-agent model-based diagnosis. The principles behind model-based diagnosis have been explained well by [18]. In this paper we follow the problem formulation given there, i.a. because Reiter's approach to the problem can be straightforwardly mapped to more recent logics used for representation of law.

The problem of multi-agent diagnosis has been addressed in different contexts. There are several interesting ways to combine multi-agent systems with model-based diagnosis that have been explored by others. One approach focuses on distributed diagnosis of (generally non-agent) systems [17,19], for instance the diagnosis of systemic failure of distributed sensors in a complex system. The systems being diagnosed are of the traditional kind, and external to the agents, but the diagnostic problem solver is a multi-agent system. This approach addresses the information and coordination problems that arise in diagnosis by a complex multi-agent system. The diagnostic agents only have local access to a greater system, and need to coordinate the diagnostic hypothesis generation

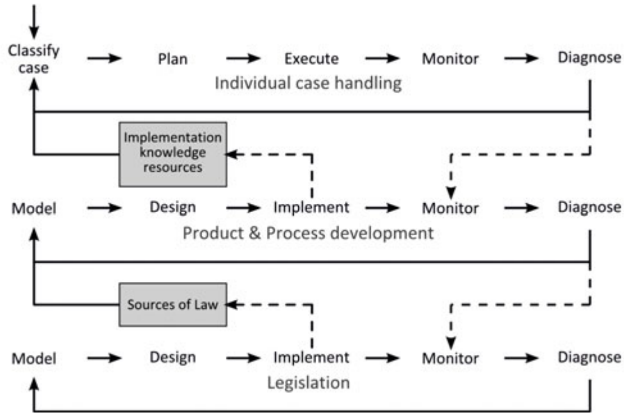


Fig. 1. The case handling, development, and legislation problem solving cycles

and testing process. In principle the centralization of diagnostic information in a single agent that discriminates between hypotheses and orders tests is the most efficient solution wherever practical [17].

Another approach is to diagnose *multi-agent plans*. A diagnostic hypothesis from this point of view is an identification of failing parts of a plan [22]. Although this approach treats a multi-agent system as the subject of model-based diagnosis, the agents are not the components of the system of interest. This problem formulation has some similarity with the one addressed in this paper, but the setting to which it applies is a fully cooperative setting. It shares with [17,19] the assumption that problems are at the root caused by different or false beliefs about the system, and not by competing objectives. We do not believe this model captures the diagnosis problems of the social organization we are interested in, but it addresses an important aspect of the problem.

Lastly, one could consider the *normative assessment* approach found in normative multi-agent system problem formulation [1] to be a kind of diagnosis approach to multi-agent systems, but the classical normative assessment problem formulation does not make a central issue out of localization of faults, which is an essential characteristic of model-based diagnosis. Although related in function, it is a specialized case of the simpler assessment problem formulation in the second perspective on problem solving in [3,2].

3 The Nature of Diagnosis Problems

The model-based diagnosis setting can be viewed as a problem of a single diagnostic agent having a model of a system to be diagnosed as its environment. Diagnosis presumes that a system can be decomposed into small components with well-understood behaviour models. Effectiveness of model-based diagnosis approaches in the literature is largely determined by the extent to which it is possible to obtain observations on the states implied by the model of the system, and the possibility to test components independently from the rest of the system. How to do this is of course not directly obvious for non-trivial social systems.

Following [18], we characterize a description of a system as a pair $(SD, COMP)$, where SD, the system description, is a set of first order sentences, and $COMP$ a finite set of constants identifying components. Typically, a system description describes how a system normally behaves, and it often distinguishes a description of structure from a description of function of the components. The functional model causally relates input and output terminals of components, and terminals of components are connected.

An observation of a system is a set of first order sentences OBS on the events happening at component terminals. If OBS and SD are inconsistent with the assumption that all components are normal, certain of the components behave abnormally. A diagnosis is a hypothesis that components $AB \subset COMP$ in $(SD, COMP, OBS)$ are abnormal and the rest normal. The diagnostic process usually involves making additional observations as evidence for ruling out hypotheses (measurement). Diagnostic reasoners may use the functional model for both causal and evidential reasoning.

Diagnoses can be generated on the basis of component fault modes. The diagnosis in this case conjectures alternative behaviour descriptions for the components that are behaving abnormally. The set of fault modes of a component may be complete: in this case the component must be behaving according to the health mode or one of the fault modes. Alternatively there may be generic *unknown fault* models.

Default reasoning about normality of component behaviour can be modeled with some predication of (ab)normality, here n, and a normal default theory of the form $DT = (\{\, Mn(c)/n(c) \mid c \in COMP\}, SD \cup OBS)$ [18]. For this default theory, Reiter's default logic extensions are exactly those of the generic diagnosis problem for $(SD, COMP, OBS)$ directed towards a *minimal* set of abnormal components [18]. In SD a complete set of fault modes $f_1, f_2, \ldots, f_n$ can be expressed by the first order axiom $t(x) \wedge \neg n(x) \supset f_1(x) \vee f_1(x) \vee f_2(x) \vee \ldots \vee f_n(x)$ [18], and alternative health modes $n_1, n_2, \ldots, n_n$ by $t(x) \wedge n(x) \supset n_1(x) \vee n_1(x) \vee n_2(x) \vee \ldots \vee n_n(x)$, where t is the component type to which the partition into behaviour modes applies. Additional sources of knowledge may guide the exploration of the diagnosis search space as it is characterized here. Diagnoses may for instance be ruled out on grounds of impossibility, faults in components may imply faults in other components, and a probability distribution of fault modes may be known that guides selection of hypotheses.

Default logic extensions are stable extensions in terms of Dung's argumentation frameworks [10]. This makes the notion of a diagnosis concrete enough to implement it with logic programming [10]. We use the multi-agent programming environment Jason, which has such an extension [5]. The conceptual problem is to present the multi-agent system as a set of components to be diagnosed.

4 Agent Roles as Components

In [4] we argued for the use of agent role instances rather than agents as the knowledge components in a simulation of noncompliance storylines. The agent role construct is associated to *reflective function*. By attributing beliefs, desires, and intentions to others, we make the behaviour of others meaningful. Agents are able to flexibly activate, from a collection of self-other representations organized by prior experience, the one(s) best suited to the circumstances [11]. We have similar collections of representations of the self in relation to the affordances of an environment, for instance for the use of tools. Agent role knowledge distinguishes itself from other self representations by the potentiality of agent role inversion: we can imagine being in the other agent's shoes, and reflect on the beliefs, desires, and intentions we would have. For agent roles such as buyer–seller we literally have the option of experiencing the transaction from the other point of view, alhough this is the exception rather than the rule in social interaction. Fonagy and Target in [11] for instance focus on parent–child attachment; Although children evidently simulate the parent, and in the process of doing so learn to reflect on themselves, agent role inversion cannot play a central role in the ability to acquire useful self–other representations. It does presumably play a role in the acquisition of the same descriptive vocabulary for the self and the other.

In MAS literature we find many proposals for adding agent roles to MAS in order to model social systems [9]. The most important operations in MAS role dynamics are enact and deact, which means that an agent starts and finishes to occupy an agent role, and activate and deactivate, which means that an agent starts executing actions belonging to the role and suspends the execution of the actions [9].

We propose to drop the notion of a deliberative agent-level process that exercises control and enacts, deacts, activates, and deactives roles, and to make the coordination between agent roles part of agent role specifications. The agent role description itself specifies when it activates another agent role or deactivates itself. This means that the deliberative agent self of a natural person becomes an empty shell: nothing more than an identity marker to impute legal and moral responsibility to. In [4] we explained this choice in the context of simulation of legal storylines in a MAS. In the MAS as a model-based diagnosis problem setting, the possibility of localization of faults is another reason to keep the behavioral repertoire and planning ability in agent components simple.

Complex combinations of agent roles, with some coordination mechanisms between them, may be considered stereotypes of persons and organizations. This notion is somewhat reminiscent of the notion of the *society of mind* [20], with the addition that the agents of which such a society is composed are coherent agent role instances. If the agent role description is adequate, its behaviour in response to the events to which it is susceptible is as predictable as that of a machine. If the agent knows the agent role it is dealing with, and has an appropriate matching self representation, it does not need to model the intentions and beliefs of the other agent explicitly because it has plans for all expected outcomes of interaction.

We assume that all relevant events are messages between agents. A payment from a to b is for instance reinterpreted as an order from a to its bank, and an acknowledgement to b from its bank. This works adequately in bureacratic environments, but we recognize the limitations of this stance.

Finally, note that not all events that are *unexpected* lead to diagnosis. Competence at task performance often allows for a certain degree of flexibility of behaviour within a role. What matters is that the event indicates a *problem* calling for reinterpretation of the environment. For this reason we distinguish variation in health modes and fault modes. An uncommon turn of events of a healthy environment is not a diagnosis problem per se.

4.1 Social Structure and Agent Role Component Descriptions

To start, we briefly sketch social structure diagnosis in Reiter's terms. Objectively, a description of a social environment is a pair $(ENV, ROLES)$, where ENV is a set of first order sentences, and $ROLES$ a set of agent role instances in the environment. All $a \in ROLES$, are assigned an agent role class $A(a)$ in ENV, which determines possible health modes $Ah_1, Ah_2, \ldots, Ah_n$ and fault modes $Af_1, Af_2, \ldots, Af_n$ for the agent role, as specified in section 3. The structural topology of the system is determined by messages $M(a_1, a_2, m)$, where a_1, a_2 are agent roles and m is a message, and agents have goals $G(a, g)$, where a is an agent role and g a goal. Behaviour descriptions in ENV appeal to agent role class and behaviour mode, e.g:

$$A(x) \wedge Ah_1(x) \wedge A(y) \wedge G(x,z) \supset M(x,y,z)$$

Objectively, that is, if the diagnostic agents are not participating in the system, the following rules constrain behaviour descriptions:

1. The message sent may be conditional on the agent class and health or fault mode of the sender;
2. The message sent may be conditional on the agent class of the receiver, but not of the health or fault mode of the receiver, since the sender does not model it;
3. The specific health mode of the receiver should not normally affect completion of goals, only the path towards it.

It is clear that, if participating agents reason about the behaviour mode of other agents, and use this reasoning in their reactions, model-based diagnosis is more complicated. It is also clear that, in real social systems, the participants do such reasoning. This observation is relevant to business process design in public administration: if personnel on the work floor implements its own workarounds for perceived problems, this may lead to impoverished diagnostic information on higher levels, and therefore a problem in the design of the system will not be addressed.

A problem with *objective* description of a social system is that system breakdown may result from participants holding different subjective social system descriptions. The existence of social structures, and the conventional or healthy behaviours associated with them, must be voluntarily accepted by the participants in the structure. Diagnostic agents participate in the social system being diagnosed. Subjectively the model-based diagnosis problem is more complicated:

1. A diagnostic agent does not have access to all messages exchanged between relevant agents;
2. A diagnostic agent may behave in a fault mode;
3. There is a distinction between topological connections to other agents that are real, in the sense that the agent really sends and receives messages, and topological connections that are hypothesized, of which the agent may receive information from other trusted agents, that may however behave in a fault mode; and
4. A diagnostic agent generally has to assume that other agents may reason about behaviour modes and adjust their behaviour accordingly, and agents operating in a fault mode may diagnose for signs of the fault mode being caught, and reactively hide their tracks.

In Fig. 2 the predicament of collecting information for diagnosis is shown in graphical form. The only ways in which agent c can know of the messages from a to b are 1) notification by a, 2) notification by b, and 3) carbon copy of a message sent through a physically limited communication channel from a to b. Often the diagnostic agent must trust either a or b. A diagnostic agent tests the trustworthiness of others by comparing reports from different sources.

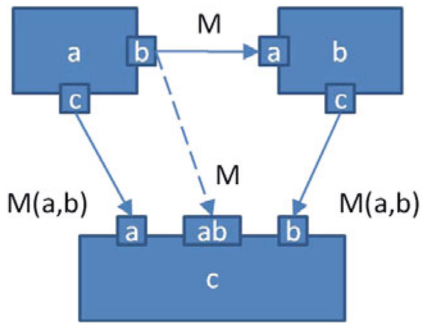


Fig. 2. There are three basic topological solutions M[a], M(a,b)[a], M(a,b)[b] for letting diagnostic agent role c directly listen in on the messages from agent role a to b. Different topologies allow for different misrepresentations of actual message events.

The pragmatic approach to the subjective model-based diagnosis problem we have chosen, depends on the possibility of translating between a representation of behaviour in first order sentences, and agent descriptions in a BDI-style agent representation language, AgentSpeak. For this purpose others in the project are working on a semi-automatic mapping from business rules specified in RIF to AgentSpeak [13].

The example behaviour rule may translate to: an x_{A,Ah_1} has the following plan operator $+!z : a(Y) \leftarrow m(Y, z)$. The plan operator means: attempt $m(Y, z)$ if goal z is added, and $a(Y)$ is the case. The partial translation reflects the fact that the same rule, depending on perspective and the reasoning abilities of the agent to which it is added, may be implemented differently.

The collection of legal diagnostic knowledge depends on the validity of the conceptual model of the diagnosis problem, with its health modes and fault modes, rather than its implementation in the form of model-based diagnosis algorithms as in [18]. Each mode represents a possible agent role in a MAS simulation, while the first order sentences describing agent roles, in combination with default logic reasoning, allow for diagnostic reflection on the social environment by agents with diagnostic ability.

4.2 An Example Knowledge Acquisition Environment: Real Estate

In this section we discuss an example setting for diagnosis, and present agent role descriptions in a form suitable for knowledge acquisition from domain experts. The tax administration is the primary responsible party for collection of taxes, including those on income presumably generated from illegal activity. It also potentially plays an important supporting role in crime fighting as an informer and witness, and its customs and fiscal intelligence departments routinely participate in organized crime fighting. Finally, it provides feedback to the legislator on new tax avoidance strategies that erode the tax base.

A representative knowledge domain subject to monitoring by the tax administration is real estate transactions. The tax administration has no direct interest in buying or selling real estate, but real estate transactions do involve a lot of money and are a good vehicle for tax evasion. In the field of real estate we find a variety of types of crime:

tax evasion: selling below or buying beyond real market value to avoid income taxes, at the expense of paying a lower property transfer tax;

bid rigging: typical of foreclosure auctions, the seller unknowingly selling his property below value to a buyer cartel, which distributes the profit among the cartel participants;

kickbacks: use of real estate transfers below or beyond real market value to hide kickbacks; and finally

extortion: a buyer or seller deviating from real market value because because of threat with some form of violence or blackmail.

These types of crime usually involve large deviations from apparent market value, or untypical quick depreciation or appreciation of real estate property value. It is not, however, trivial for an outsider to determine what the real market value of a property should be, and the market value may be misrepresented even to outsiders by appraisers that participate in a scam. The deviation itself is for a tax administration not a reason to act: to enforce, a plausible story is needed, involving collusion and a motive. This story usually involves an offense besides tax evasion.

Knowledge acquisition is about eliciting these stories, with a special focus on intentions and plans of the participants, and the messages that must have been exchanged.

The health modes of the real estate transaction depend on the intention of the buyer and seller. A healthy buyer wants to obtain at least market value mv, and expects little more. A healthy seller wants to pay no more than market value, and expects little less. Both may refer to alternative bids to judge market value, or may trust an appraiser to estimate market value mv'. A *healthy* factor that will affect value negatively is for instance haste.

Fault modes for the buyer or seller are based on an intentional deviation $addv$ from market value. This is either the intention of buyer or seller alone, in collusion with fake bidders, a fraudulous agent of the other party, or a fraudulent appraiser ($mv' = mv + addv$), or it is a coordinated intention between both, because the amount $addv$ is owed for some service, or is a gift, or is being extorted. When the buyer or seller is the victim, this will become apparent if the participants are interviewed about the transaction directly. Doing this does however tip off the participants in a coordinated transaction, and allows them to hide their tracks.

The following is a generic AgentSpeak plan for a fault mode seller who intends to transfer an amount $addv$ to the buyer (with predicates in natural language, and variables of central importance emphasized, to enhance readability):

1. ! transfer $addv$ to recipient
2. $+$! transfer $addv$ to recipient $\leftarrow$ propose to sell a property worth mv for $mv - addv$.
3. $+$ Plan to sell a property worth mv for $mv - addv$ accepted $\leftarrow$ propose a property worth mv.
4. $+$ Property worth mv is accepted $\leftarrow$! secure property for mv; ! sell property for $mv - addv$ to recipient.

In this form, agent mode descriptions can be directly written down during an interview with experts.

For a tax administration, various sources of information are available to test the hypothesis that it is dealing with this type of buyer. The buyer leaves a conspicuous evidence trail if he has to acquire the property and then immediately sells it for a loss. Also an appraisal mv' before the property sells for $mv' - addv$ is conspicuous, just like the buyer immediately selling it again for mv. These transactions are of course suspicious and attract attention to the relationship between buyer and seller. One way to hide a relationship between participants, and therefore the motive of the transaction, is to make transactions part of larger packages of transactions: a for instance sells property p_1 to foreign party b for $mv - addv$, who then sells property p_2 to c for $mv - addv$, the right holder for a payment of an amount $addv$ towards a, who then cashes $addv$ by selling p_2. Much more complicated arrangements exist, making $addv$ for instance smaller as a percentage of the whole, and therefore less conspicuous, by increasing the mv of the property package.

Since participants in a complicated scam will have trouble trusting eachother, they are likely to sign all deeds at the same time in the presence of a notary lawyer. Notary lawyers that are suspiciously often involved in setting up such complicated transactions will therefore also draw attention, etc.

As we consider these scenarios in the context of model-based diagnosis, the number of relevant agent roles (appraisers, agents, notary lawyers, auction houses) involved grows, because we rely on interception of messages provide us with a motive for a behaviour.

5 Conclusions and Discussion

We propose the model-based diagnosis problem as an explanation of the driving forces behind requests for change and policy making in public administration. Identification of fault models with specific agent role descriptions explaining the fault will – over time – increase the accuracy of policy forecasts and the robustness of design in public administration.

This approach has advantages over ad hoc approaches to embedding tacit knowledge about noncompliance in the business rules and business processes of the organization. Although there are good reasons to believe that social systems are too complex for application of model-based diagnosis algorithms, a diagnosis-based methodology for compliance knowledge representation:

– distinguishes knowledge of conventional behaviour, and intended, designed behaviour, from common incidents and noncompliance in a systematic manner;

- permits the modeling of misunderstandings between network partners arising from different views of the social environment;
- encourages early simulation of abuse of new policy by stereotypical policy abusers; and
- encourages systematic collection of diagnostic knowledge for enforcement and redesign purposes.

An aspect of our approach that clearly needs more development is the use of generic agent coordination design patterns, to deal with the variety of approaches of coordinating essentially the same plan among multiple agents. This is especially relevant because coordination also takes the place of a higher order deliberation process within an agent, and because normative positions play an important role in this domain. The approach of [16], and later work based on it, specifically attracts our attention, and combines well with our views on the use of normative rules in law.

5.1 Discussion: Applicability of Model-Based Diagnosis to Social Systems

Based on knowledge availability requirements, one could argue that administrative organizations exist in an environment in which one should plan instead of design. As pointed out by [6], when we feel able to control a domain by encapsulating processes into fixed structures we decide on a *design* for dealing with a type of problem and implement it. When we feel this is not feasible, given the characteristics of the domain and problem, we make situation-specific *plans* to address a problem. Ideally, a fully articulated problem description would allow for both terminologies, but in practice we generally address some aspects of problems as planning problems and others as design problems. Proposals for increased robustness in large organizations are likely to argue for less rigid design solutions, i.e. for more flexible planning solutions as in [12].

Based on this argument, complex social systems should not be analyzed as a model-based diagnosis problem, because lack of control over the environment plays too large a role. Model-based diagnosis presumes not only that 1) we have a correct and complete functional specification of a system, but also that we have enough control over the environment of the system to 2) isolate it from undesirable inputs, 2) create diagnostic inputs, and 3) reliably observe the outputs. These conditions are obviously only met in controlled social experiments. Hence the assessment alternative mentioned in section 2.

Still, no one hesitates when speaking about design of social systems or diagnosis of social problems, even if we accept that the models exhibit less desirable properties than those in many engineering disciplines. Most importantly, design of social systems in public administration is indeed driven by a logic of component replacement and repair, and this is a clear characteristic of the design & diagnose perspective. A MAS configuration represents a design, and the flexibility in the behaviour repertoire of participating MAS agents represents a potential for (single agent and multi agent) plans. It requires no great leap of the imagination to think of diagnostic hypotheses about a MAS configuration.

As indicated in section 4, the utility of the model-based diagnosis does depend on the extent to which we separate diagnosis problems from routine day-to-day processes. The model assumes that the agent components themselves refrain from doing diagnoses and repairs. The distinction between routine primary business processes and background diagnostic processes is indeed a common problem decomposition in large organizations. Simple problem monitors are built into the primary business process logic, and diagnosis problems that require deep understanding of the domain and may lead to enforcement, resource reallocation, redesign, or new policy, are assigned to dedicated experts.

This type of diagnostic task is typically less well-defined and not supported by knowledge representation. There are obvious reasons for this. Judging by the successes and failures of expert systems in the real world [14], complex models of social environments are less suitable for decision support systems. *Centralizing* the hypothesis generation and discrimination process in a diagnosis problem is a key factor in keeping communications within the organization to a manageable level [17]. A deeper understanding of rare and abnormal scenarios typically requires a much more complex model of the environment, and eventually this model will become too complex to be also of value in primary business processes. When the hypothesis generation and discrimination process is centralized, there is no economic case for a model of the environment useful for diagnosis.

Even if the diagnostic process is in practice rarely supported by decision support software, and in an ad hoc manner if it is, the tacit knowledge used in this process has considerable influence on policy making and redesign of business processes and software. The alternative for modeling agent role fault modes is knowledge management in the form of stories revolving around the motives, collusions, and plans of agents. The model-based diagnosis model gives a natural place to these stories in knowledge representation, and at least conceptually embeds the activities of design and policy making department into the business processes.

References

1. Boella, G., Pigozzi, G., van der Torre, L.: Five guidelines for normative multiagent systems. In: Proceeding of the 2009 Conference on Legal Knowledge and Information Systems, pp. 21–30. IOS Press, Amsterdam (2009)
2. Boer, A., van Engers, T.: Generic problem solving tasks and agent roles in public administration. In: Winkels, R. (ed.) Legal Knowledge and Information Systems. Jurix 2010: The Twenty-Third Annual Conference. IOS Press (December 2010)
3. Boer, A., van Engers, T.: Knowledge acquisition in public administration. In: Cimiano, P., Pinto, H.S. (eds.) EKAW 2010. LNCS, vol. 6317, Springer, Heidelberg (2010)
4. Boer, A., van Engers, T.: An agent-based legal knowledge acquisition methodology for agile public administration. In: ICAIL 2011: Proceedings of the 13th International Conference on Artificial Intelligence and Law. ACM, New York (to be published, 2011)
5. Bordini, R., Hübner, J., Vieira, R.: Jason and the golden fleece of agent-oriented programming. In: Weiss, G., Bordini, R., Dastani, M., Dix, J., Fallah Seghrouchni, A. (eds.) Multi-Agent Programming, Multiagent Systems, Artificial Societies, And Simulated Organizations, vol. 15, pp. 3–37. Springer, Heidelberg (2005), `http://dx.doi.org/10.1007/0-387-26350-0-1`, doi:10.1007/0-387-26350-0-1
6. Breuker, J.: Components of problem solving and types of problems. In: Steels, L., Van de Velde, W., Schreiber, G. (eds.) EKAW 1994. LNCS, vol. 867, pp. 118–136. Springer, Heidelberg (1994)

7. Chandrasekaran, B., Johnson, T.R.: Generic tasks and task structures: History, critique and new directions. In: David, J.M., Krivine, J.P., Simmons, R. (eds.) Second Generation Expert Systems. Springer, Heidelberg (1993)
8. Clancey, W.J.: Heuristic classification. Artificial Intelligence 27, 289–350 (1985)
9. Dastani, M.M., van Riemsdijk, M.B., Hulstijn, J., Dignum, F.P.M., Meyer, J.-J.C.: Enacting and deacting roles in agent programming. In: Odell, J.J., Giorgini, P., Müller, J.P. (eds.) AOSE 2004. LNCS, vol. 3382, pp. 189–204. Springer, Heidelberg (2005)
10. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence 77(2), 321–357 (1995), http://www.sciencedirect.com/science/article/B6TYF-4031CDM-P/2/b6a314405d7de8cafd03b1f4f457235e
11. Fonagy, P., Target, M.: Attachment and reflective function: Their role in self-organization. Development and Psychopathology 9(04), 679–700 (1997), http://dx.doi.org/10.1017/S0954579497001399
12. Gong, Y., Janssen, M.: Agent-based simulation for evaluating flexible and agile business processes: Separating knowledge rules, process rules and information resources. In: Aalst, W., Mylopoulos, J., Sadeh, N.M., Shaw, M.J., Szyperski, C., Barjis, J. (eds.) EOMAS 2010. LNBIP, vol. 63, pp. 41–58. Springer, Heidelberg (2010), http://dx.doi.org/10.1007/978-3-642-15723-3-4, 10.1007
13. Gong, Y., Overbeek, S., Janssen, M.: Business rules for creating process flexibility: Mapping RIF rules and BDI rules. In: 1st International Workshop on Web Intelligent Systems and Services, WISS 2010 (in print, 2011)
14. Grandon Gill, T.: Early expert systems: Where are they now? MIS Quarterly 19(1), 51–81 (1995)
15. Greiner, R., Smith, B.A., Wilkerson, R.W.: A correction to the algorithm in reiter's theory of diagnosis. Artificial Intelligence 41(1), 79–88 (1989), http://www.sciencedirect.com/science/article/B6TYF-47WTTF0-69/2/f6804b2a121735af3b2a03c9b440e121
16. Jennings, N.R.: Commitments and conventions: The foundation of coordination in multiagent systems. Knowledge Engineering Review (1993)
17. Kalech, M., Kaminka, G.A.: On the design of coordination diagnosis algorithms for teams of situated agents. Artificial Intelligence 171(8-9), 491–513 (2007), http://www.sciencedirect.com/science/article/B6TYF-4NB99GY-2/2/1011c067c72d15338a29a6c0657ae8da
18. Reiter, R.: A theory of diagnosis from first principles. Artificial Intelligence 32(1), 57–95 (1987), http://www.sciencedirect.com/science/article/B6TYF-47WTTBS-4T/2/b2d7b5acbca6116be93d292e56ee81aa
19. Roos, N., ten Teije, A., Witteveen, C.: A protocol for multi-agent diagnosis with spatially distributed knowledge. In: Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS 2003, pp. 655–661. ACM, New York (2003), http://doi.acm.org/10.1145/860575.860681
20. Singh, P.: Examining the society of mind. Computing and Informatics 22, 521–543 (2004)
21. Steels, L.: Components of expertise. AI Mag. 11(2), 30–49 (1990)
22. Witteveen, C., Roos, N., van der Krogt, R., de Weerdt, M.: Diagnosis of single and multi-agent plans. In: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS 2005, pp. 805–812. ACM, New York (2005), http://doi.acm.org/10.1145/1082473.1082596

Technical and Legal Meaning of "Sole Control" – Towards Verifiability in Signing Systems*

Mirosław Kutyłowski[1], Przemysław Błaśkiewicz[1], Łukasz Krzywiecki[1], Przemysław Kubiak[1], Wiesław Paluszyński[2], and Michał Tabor[2]

[1] Wrocław University of Technology
[2] Trusted Information Consulting, Warsaw
{miroslaw.kutylowski,przemyslaw.blaskiewicz, lukasz.krzywiecki,przemyslaw.kubiak}@pwr.wroc.pl, {wieslaw.paluszynski,michal.tabor}@ticons.pl

Abstract. One of the fundamental ideas of the framework of electronic signatures defined in EU Directive 1999/93/WE is "sole control" over signature creation data. For a long time "sole control" has been understood as using black-box devices for which a certain third party has issued a certificate, whereas the signer was supposed to trust blindly the authorities and certification bodies. This has been claimed as the only feasible solution.

Recent advances in technology and development of verifiable systems show that it is possible to provide systems such that the signer has much more control over the signing process and can really maintain control over the signature creation data. The main idea is that breaches in the system cannot be excluded but if they occur, then the signer can provide evidence of a fraud of a third party.

Keywords: electronic signature, secure signature creation device, qualified signature, mediated signatures, end-to-end verifiability.

1 EU Directive and Advanced Electronic Signatures

Inevitably, in the near future, paper documents will be replaced by electronic ones. This process is quite advanced in some areas. For instance, banks and other financial institutions use paper documents in most cases only for direct contacts with clients, but even in these areas electronic channels are preferred due to processing costs. B2B and B2C contacts are increasingly dependant on electronic flow of documents.

The only area that is lagging behind is flow of documents to and from public bodies. The main obstacle is legal requirement for a *written form* omnipresent in legal proceedings. Consequently, the act of signing a document is a crucial phase in finalization and confirmation of a legal step. In order to create a legal institution of *electronic written form* it is necessary to mimic handwritten signatures in the electronic world.

* Supported by Polish Ministry of Science and Higher Education fund for research & development in years 2009/2011, grant No. O R00 0015 07. The first and second authors have been supported by Foundation for Polish Science, "Mistrz" Programme.

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 270–281, 2011.

Technology of digital signatures, started with emergence of RSA in 1978, provided firm cryptographic background for electronic signatures. Even if security of digital signatures rely on some algebraic assumptions (computational hardness of factorization and of discrete logarithm in some groups), forging signatures without private signing keys is considered infeasible with achievements of contemporary mathematics. Based on this we can be convinced that a signature verifiable with a certain public key has been created by a person holding the matching secret key.

EU Directive 1999/93/EC. In order to pave way for legal application of asymmetric cryptography, European authorities decided to initiate introduction of electronic signatures into national legal systems. The aim of EU Directive 1999/93/EC [1] was to provide a common framework for signing electronic documents in an electronic way. The requirements for *advanced electronic signatures* (Art. 2, point 2 of [1]) are stated as follows:

> *"advanced electronic signature" means an electronic signature which meets the following requirements:*
> *(a) it is uniquely linked to the signatory;*
> *(b) it is capable of identifying the signatory;*
> *(c) it is created using means that the signatory can maintain under his sole control; and*
> *(d) it is linked to the data to which it relates in such a manner that any subsequent change of the data is detectable;*

The basic cryptographic properties related to asymmetric cryptography make it possible to fulfill requirements (b) and (d) in the sense that if a signature s for a given document D is verified positively, then one can guarantee that it was the holder of a certain private keys who created the signature s under D. However, asymmetric cryptography itself does not link the holder of the secret keys and the declared signatory. This is also the main difficulty and Achilles' heel of the concept.

Usually, the signatory is given a technical device, called *secure signature creation device*, SSCD for short, that implements *signature creation data*. "Implementing" means here providing a technical solution which enables the signatory to keep the signature creation data under his *sole control*. Then we have only to link the signatory with the matching public keys (or more generally: *signature verification data*) in order to enable fulfillment of (a) and (b) as well. However, the concept of advanced electronic signatures does not point directly to the concept of SSCD; other means can be used as well as long as they fulfill the stated requirements.

Rationale for the definition of advanced electronic signatures. There are some important features of the EU definition of advanced electronic signatures. First, the focus is set not only on a single cryptographic key. Protection of a given cryptographic signing key is not enough for multiple reasons. Above all, for some schemes there are multiple signing keys corresponding to the same public key and a service provider could easily circumvent the legal measures and use a fully functional key to forge signatures without violating protection of a *given* private signing key. Therefore the definition must be broader and encompass all cryptographic material that is necessary and sufficient to create a signature.

The second important point is that the notion "means" concerns also the actions that are needed to activate SSCD, such as the use of PIN numbers, biometric authentication, one-time passwords and similar mechanisms. An important point is that even the weakest option (PIN) should guarantee an appropriate level of security of the signatory.

One of the key elements of the definition is that the signatory *can maintain* the means discussed *under his sole control*. Only the ability is mentioned here, as in many cases it is impossible to actually guarantee that the signatory behaves in a correct way. Consequently, signatory's sloppy behaviour cannot be considered a valid reason for denying his signatures. On the other hand, it protects the signatory against solutions that contain a built-in malicious trapdoor: when existence of a trap-door is revealed, it becomes also evident that the signatory could not maintain the keys under his sole control.

2 Black-Box Concept for Secure Signature Creation Devices

In the past decade there have been two dominating solutions for SSCD. One is implementing SSCD on cryptographic smart-cards. The other approach is storing cryptographic keys of all signatories on a secure server and providing relatively strong authentication methods for accessing the server (the later approach was used in Danish NemID project). A common goal of these solutions is protection of the signing keys: they are stored and used in an environment that protects against cloning and exporting them outside. For this reason the devices in question must prevent any access to these keys (such as reading, overwriting), except for creating signatures after successful authentication of a signatory. This is realized by strict design rules for device operating systems. In particular, such system should be as simple as possible in order to ease its inspection, all options must be well documented and limited to the ones that are necessary.

Apart from a secure implementation of the operating system, the construction of SSCD itself must prohibit retrieving the secret key in an indirect way, through side channel information, probing its memory or techniques such as fault cryptanalysis. As a minimum, the device should be tamper-evident. Preferably, it should be tamper-proof so that any attempt of invasive access results in destruction of the protected keys.

Smart cards used as SSCD have the advantage that the signatory has the SSCD under direct physical control. On the other hand, smart cards demand particular protection from the owner - they cannot be left unattended as activation mechanism with PIN is not sufficiently reliable to protect against random guessing. This becomes a problem since a person cannot protect a growing number of different security items and remember any number of PINs. Major problems are also related to possible PIN capture by the PC, which connects a smart card reader and a keyboard (of course, there are smart card readers with a keyboard for entering the PIN and even smart cards with touch fields enabling direct input, but these solutions are not widely used).

Certification concept. The dark side of tamper-proof devices is that protection measures not only prevent retrieving keys, but also impede security inspection regarding compliance of the actual device implementation with its description. In particular, an inspection may require:

– invasive access to the actual hardware (thereby it is limited to a random probe of the devices),
– confidential information concerning layout design, executed code, security mechanisms, as well access to confidential documentation,
– expert knowledge.

Due to complex nature and high costs, inspections must be performed by specialized bodies with access to appropriate laboratory equipment.

For this reasons, the trust concept is based on *certification bodies* performing inspections. The certification bodies should satisfy the following conditions:

– they should be trustful in the sense that, for instance, they are independent in their opinions and not connected in any economic way with the companies providing products under inspection,
– they should posses necessary know-how and equipment.

Implementing the first condition is a challenge: it is hardly possible to formulate legal conditions that would guarantee independence; all we can do is to exclude some situations where independence is violated. Supervision over certification bodies (e.g., like in Germany) does not solve the problem since it merely shifts the problem to another area: *who will guard the guard?*

Apart from the question *who performs the inspection?* there is the issue of *what should be checked?* In order to avoid incomplete inspection the targets of certification process should be defined independently of the process itself. Moreover, they should be in some sense comparable – therefore well designed *protection profiles* within the framework of Common Criteria are very helpful.

Limitations of certification framework. Even though certification is very useful for quality enforcement, it is also somewhat perpendicular to the concept of *sole control* formulated in the electronic signature Directive. It trades one problem for another: in the worst case, the manufacturer and the certification body may collude and conceal trapdoors and security flaws. Moreover, they both can be under pressure of state security agencies to tolerate (or even provide) some backdoors for these agencies for the sake of public security. So in reality a certified product is a product that is under control of the manufacturer and the certification body (or bodies, if many certificates are issued). Of course, the companies must be extremely careful - revealing any malicious behavior means loosing clients. However, the case of CryptoAG [2] shows that there are at least serious suspicions concerning installation of backdoors by major vendors. Finally, even if the certification process is performed correctly, this does not contribute to signatory's "sole control".

3 Attacks and Threats

Protection of signing keys stored in a SSCD device is a never ending race between protection technology and sophisticated attack methods. Moreover, this is a race that takes place behind the curtains: it is not only common people who have no access to up-to-date information, but also academic world, which has proven to be very efficient in finding security flaws and attack techniques. In fact, we are afraid that the only state authorities who have real insight in the situation are the ones in the countries with leading enterprises on the smart card and HSM market. One could expect that information about newest attack methods is kept secret. After all, if published, it might facilitate revelation of signing keys and, more importantly, decryption keys of confidential documents. Concluding,

- in most countries it is impossible even to determine if a signatory has sole control over an SSCD,
- the situation is dynamic in the sense that an SSCD may suddenly become exposed to an attack which is purposefully kept secret.

Unreliable randomness. In the realm of practical security of digital signatures some unresolved problems are still present. One of them is wide use of randomization: except for a few schemes (among them RSA), creation of a signature involves selection of a random parameter. Cryptographic schemes specify that these parameters are drawn uniformly at random from a given space. It has been observed that not fulfilling this requirement may lead to complete collapse of security features [3], [4]. On the other hand, there are no reliable methods that would prove that the output of a random number generator fulfills the uniformity requirements. What can be achieved is the elimination of poor generators with outputs evidently violating statistical properties of uniform probability distribution.

The problem is even more acute, since usually a composition of purely random source (but possibly a biased one) and a pseudorandom generator using the output of the random source as a seed [5] is used. There are many arguments why such a combination is reasonable, but on the other hand it is no longer true that the signing scheme is implemented according to the original description, and consequently all security claims might be not relevant.

Tricks with random parameters. The main problem is that even if a device is certified, the signatory cannot be sure that the device in his possession is identical with the devices inspected by a certification body. From the signatory's point of view, the device is a black box and as long as its observable behavior is not suspicious, the signatory cannot claim any problems.

On the other hand, the manufacturer can implement malicious code (which can even be claimed non-malicious). This code (which is a special implementation of standard cryptographic operations to be executed by the device) creates a very efficient channel for deriving the secret keys from the device without any physical contact with the device. All that the adversary has to know is a certain secret (not stored in the device and not derivable from the data it stores) and a few signatures the device generated (typically two consecutive ones) – for details see [6] and the papers following it. These malicious methods work particularly smoothly for DSA, ECDSA and other similar schemes, but even RSA is not completely immune against them (see e.g. [7]).

Tricks with key generation. Since a service provider generating the keys outside the signatory's device may potentially retain and misuse them, it is often (but not always, see e.g. [8]) believed that generating the keys on the card solves all problems. This is not really the case. The signatory cannot be really sure if the command for generating the keys is not a bogus operation and the keys that are supposedly "generated" have already been installed on the card by the manufacturer and the whole operation only activates them. Of course, this attack can have more subtle implementations, but the principle remains the same.

4 Concept of Verifiability

Another idea comes from a recently developed concept of security enforcement in the field of e-voting protocols. They are far more complicated than signature schemes and the risk is not limited to a single person but might have severe consequences to the society as a whole.

Trust based on independent certification is not sufficient for enforcing fairness and completeness in electronic elections. This led to the idea of the so-called *end-to-end systems* which are based on the following principles:

- no expert knowledge is required to evaluate relevant security mechanisms,
- evaluation results should be self-evident and undeniable,
- effectiveness of security mechanisms can be checked while the system is running and not only as a prior inspection,
- no assumption is made about honesty of any party of the protocol, it concerns in particular the manufacturers and supervision authorities.

Security levels according to E2E verifiability may be formulated. In such a hierarchy the solution in which "sole control" is based on the *assumption* that a third party or parties do not gain control over the device or leave some trapdoors must be definitely at the very bottom of the hierarchy.

5 Solution Examples for Electronic Signatures

One may claim that people are forced to blindly believe that the system is fulfilling all the requirements and that a signatory has sole control over signature creation data. In this approach, the legal framework may deny the right to challenge the fact that a signature does not fulfill the conditions of advanced electronic signatures. The following situations are worth mentioning:

- there are lobbying efforts to extend the legal consequences of qualified signatures to signatures verified with qualified certificates but not generated with an SSCD. In this case the signatures are not even advanced electronic signatures and the user has no sole control over the cryptographic signing keys.
- Polish law on electronic signatures denies the right to challenge electronic signatures created during validity period of corresponding qualified certificates, when the device used turns out not to be an SSCD [9, Article 6].

As we shall see below, there are emerging techniques that provide effective control mechanisms preventing attempts of breaking "sole control" of the signatory.

5.1 Fail-Stop Signatures

Fail-stop signatures (see e.g. [10]) were designed to address the question of a powerful adversary which:

- has access only to the public key of the signatory,
- has sufficient algorithmic knowledge and necessary computational power to break the scheme so that he can generate a correct signature on behalf of the signatory.

This scenario is not unrealistic as public knowledge of cryptanalytic capabilities of major institutions is based on speculations only. The history teaches us that some knowledge has been kept secret for decades.

The idea of fail-stop signatures is to protect sole control of the signatory but make the scheme fraud-evident rather than tamper resistant. The main features of the scheme are the following:

1. for each public key there is a large number of matching private keys,
2. each of these keys generates different signatures for each plaintext, however each of them can be verified correctly,
3. given two signatures generated for the same plaintext but with different private keys matching the same public key, it is possible to solve some hard computational problem which is fundamental for the security of the scheme.

The idea is that if a powerful agency succeeds in breaking a given signature scheme then it still has to determine the keys held by the signatory. Failure to guess the private key and an attempt to forge a signature leads to the situation from the point 3 above: When a forged signature is used, the legitimate signatory may sign the same plaintext with his key and solve a hard computational problem (it might be computing discrete logarithm for some challenged elements) and thereby prove that he has lost sole control over the cryptographic material.

In the scenario of fail-stop signatures we do not prevent computing private keys by the third parties, we make any usage of these keys evident. This way cryptanalysis looses any practical sense, as the results cannot be used against the signatory. This is an example of a solution in which the signer has effective control over very powerful adversaries.

5.2 Mediated Signatures

The idea from [11] is to split signature creation data into a number of shares distributed among two or more different devices. The way the splitting is performed guarantees that:

- it is impossible to recover one share from all the remaining shares,
- for creating a signature it is necessary to execute a protocol in which all shares are used in a substantial way.

In this scenario one share is left in the SSCD, while other shares are located for instance at a PC of the signatory and on a remote server.

The main idea is that while it is inevitable that some of the devices are stolen, lost, or captured by an adversary, putting the shares of one signature creation data on different devices forces the adversary to take control over more than one device of the same type. Moreover, since the devices holding shares are different, it is even harder for the adversary to succeed. For instance, she would not only have to steal the signature card of a signatory but also break into his PC.

Ideally, there is no procedure of splitting the keys, but a distributed protocol for key generation, in which no device gets any information save its own share. The other method is to generate the keys at one place, split and distribute them, and finally to erase all material that would enable to recover the full cryptographic material at one place.

5.3 Validation and Authorization Service

The idea of this solution is implicitly contained in the Polish law on electronic identification cards [12] and already implemented in Denmark in NemID framework. In order to create a signature, an SSCD is forced to contact a central server. The reasons for this are the following:

- it is easy to implement strong authentication mechanisms, much stronger than a PIN number for a smart card; the procedure can be partially independent of the SSCD and use alternative communication channels (e.g. SMS with a one-time password sent to mobile phone of the signatory),
- one can implement security monitoring, like for credit cards usage, with security alert in any irregular situation,
- signatory can be informed via an independent channel (SMS, email) about created signatures,
- signatory can restrict the time when the SSCD is admitted to create signatures; any policy of this kind can be easily implemented.

This way, unlike in classical smart cards scenario where keeping PIN secret leaves false impression of control over the card, the signatory gets real possibility to control usage of signature creation data.

Additionally, one can easily achieve the following goals:

- signing time can be determined quite exactly,
- archive version of the signature can be created immediately,
- control data can be stored in a hash chain, prohibiting misuse of the SSCD and backdating electronic documents.

In fact, we can talk about a new service "Online signature status protocol" (OSSP), which provides answers to the question "have you finalized this signature?". For more discussion on such an approach see [13].

5.4 Floating Exponents

Mediated signature described in Sect. 5.2 can be extended by an additional security mechanism [14,15]. The trick is that the keys used by the SSCD and the mediator server evolve slightly in a synchronized way. Namely, if the original key d is split into $d' = d - \delta$ (SSCD's share) and δ (mediator's share), then in the modified scheme the keys are updated at each step as follows $d' := d' + \epsilon$, $\delta := \delta - \epsilon$, where ϵ changes pseudorandomly (i.e., in each step is different).

Some care must be taken in order to prevent de-synchronization in case of session interruption. However, floating exponents may be used to detect existence of a cloned SSCD. Indeed, once a cloned SSCD is used and a signature is created in cooperation with the mediator, the key of the mediator gets updated, while the key on the original SSCD does not. So, once the original SSCD is used, signature creation yields an invalid (i.e., not verifiable) signature, indicating that something undesired has occurred. Therefore, this solution is an attack detection mechanism.

5.5 Two-Head Dragon

The idea from [16] is that the signature scheme has built-in mechanisms that prevent usage of a cloned device in the sense that any attempt to use it leads to key exposure. This is a severe countermeasure against any use of leaked keys: the only thing the adversary can do with the stolen keys is to publish them, as a forgery leads to an effectively same result. This mechanism is particularly useful, if the service provider has full technical control over key generation service and delivery, while at the same time bears the whole financial risk of key exposure.

In some sense this mechanism extends the features of fail-stop signatures. It changes the legal situation inevitably, since in case of key exposure there is no way to hold the signatory responsible for the signatures created with these keys.

5.6 Accountability

Another solution is based on one-time signatures converted to multiple-signature schemes [17], [18]. In this case a signature device can generate only a limited number of signatures, each of them based on a one-time signature scheme. The main feature of this algorithm is that the signatures are assigned to leaves of a Merkle tree, and any attempt to put two signatures at the same leaf provides an undeniable proof of "double signing" (either by a faulty SSCD or by a clone).

6 Proposed Framework

6.1 Requirements

The following key questions have to be answered when determining a level of "sole control":

1. to what extent does the signatory depend on honesty of the third parties? What are the potential consequences of malicious behavior of these parties?
2. In what way can the signatory transfer the sole control to other parties, if he wishes to do so (or is coerced to do so)?
3. Is the sole control evident to the signatory, or is it merely a technical fact that can be checked by third parties?

The answers to these questions must be included in any set of design requirements for a system with functionality of signing digital data.

6.2 Security Levels

In our opinion it is necessary to formulate common criteria for determining levels of "sole control" over signature creation data.

There are different aspects of "sole control" and we propose to consider them separately with different levels of security.

trust :
- **level TRUST-0** :
  the signatory has to trust blindly the system providers
- **level TRUST-1** :
  it suffices to trust certification and auditing bodies supervising the system of digital signatures
- **level TRUST-2** :
  the signatory does not have to trust any third party

transfer :
- **level TRANSFER-0** :
  the signatory may transfer the possibility of creating signatures to a third person without substantial effort
- **level TRANSFER-1** :
  transfer of sole control is detectable with substantial probability
- **level TRANSFER-2** :
  transfer of control is impossible neither by technical nor organizational means

Guarantees :

level GUARANTEE-0 :
"sole control" is based on a declaration of the system provider

level GUARANTEE-1 :
"sole control" is confirmed by certification bodies and runtime audits

level GUARANTEE-2 :
"sole control" is self-evident in the sense that any breaches can be recognized at least by the signatory and are undeniable

Usage control :

level CONTROL-0 :
the signatory has no record about usage of signature creation data, especially those created without his consent

level CONTROL-1 :
the signatory has access to data indicating actual usage of the signature creation data

level CONTROL-2 :
the signatory has access to data proving actual usage of the signature creation data in undeniable way

6.3 Open Evaluation Framework

One of the major issues in providing sole control of a signatory over the SSCD is providing a framework within which any security flaws can be identified and informed to the public. We believe that the following rules should be applied:

1. Security evaluation does not need to be performed by bodies that are independent of the manufactures, system providers, system administrators. In fact, the reliable search for weaknesses is done by competition that has strong interest in detecting any security flaws.
2. The result of security evaluation may not depend on unrevealed mechanisms: for protection of signature creation data we need to apply a "no security by obscurity" principle as it is done elsewhere in cryptography.
3. Evaluation should be performed in public, and all results must be available for potential users. European Community may contribute a lot in providing an international and unbiased forum for exchange of information.

7 Conclusions

The analysis of available cryptographic mechanisms and contrasting it with real requirements reveal that it is possible to design mechanisms supporting "sole control" of the signatory over signature creation data. Moreover, this can be done far more effectively than technical protection offered by smart cards alone. One can achieve security features which are beyond the range of smart cards and thereby really implement "sole control".

Apart from the technical solutions, we propose to reshape evaluation schemes and move from certification procedures to open evaluation that has shown to be extremely effective in case of NIST competitions for cryptographic standards.

References

1. European Parliament and the European Council: Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a community framework for electronic signatures. Official Journal of the European Communities L(13) (January 19, 2000)
2. Wer ist der befugte vierte. Der Spiegel 36 (1996)
3. Klein, A.: Attacks on the RC4 stream cipher. Des. Codes Cryptography 48(3), 269–286 (2008)
4. Biryukov, A., Shamir, A., Wagner, D.: Real Time Cryptanalysis of A5/1 on a PC. In: Schneier, B. (ed.) FSE 2000. LNCS, vol. 1978, pp. 1–18. Springer, Heidelberg (2001)
5. Konstantinou, E., Liagkou, V., Spirakis, P.G., Stamatiou, Y.C., Yung, M.: Electronic national lotteries. In: Juels, A. (ed.) FC 2004. LNCS, vol. 3110, pp. 147–163. Springer, Heidelberg (2004)
6. Young, A., Yung, M.: The dark side of "Black-box" cryptography, or: Should we trust capstone? In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 89–103. Springer, Heidelberg (1996)
7. Young, A., Yung, M.: A Space Efficient Backdoor in RSA and Its Applications. In: Preneel, B., Tavares, S. (eds.) SAC 2005. LNCS, vol. 3897, pp. 128–143. Springer, Heidelberg (2006)
8. Handschuh, H., Naccache, D., Paillier, P., Tymen, C.: Provably secure chipcard personalization, or, how to fool malicious insiders. In: CARDIS, USENIX (2002)
9. Polish Parliament: Electronic Signature Act. Dziennik Ustaw 130(1450) (September 18, 2001)
10. Bleumer, G.: Fail-stop signature. In: van Tilborg, H.C.A. (ed.) Encyclopedia of Cryptography and Security. Springer, Heidelberg (2005)
11. Boneh, D., Ding, X., Tsudik, G., Wong, C.M.: Instantenous revocation of security capabilities. In: USENIX Security Symposium (2001)
12. Polish Parliament: Ustawa od dowodach osobistych (act on personal identity documents). Dziennik Ustaw 167(1131) (2010)
13. Initiative: PKI 2.0 (2011), http://www.pki2.pl
14. Nicolosi, A., Krohn, M.N., Dodis, Y., Mazières, D.: Proactive two-party signatures for user authentication. In: NDSS. The Internet Society (2003)
15. Błaśkiewicz, P., Kubiak, P., Kutyłowski, M.: Digital signatures for e-government – a long-term security architecture. China Communications 7(6) (2010)
16. Błaśkiewicz, P., Kubiak, P., Kutyłowski, M.: Two-head dragon. clone-fail signature creation devices. In: Chen, L. (ed.) INTRUST 2010. LNCS, vol. 6802, pp. 173–188. Springer, Heidelberg (2011)
17. Merkle, R.C.: A Certified Digital Signature. In: Brassard, G. (ed.) CRYPTO 1989. LNCS, vol. 435, pp. 218–238. Springer, Heidelberg (1990)
18. Rohde, S., Eisenbarth, T., Dahmen, E., Buchmann, J., Paar, C.: Fast Hash-Based Signatures on Constrained Devices. In: Grimaud, G., Standaert, F.-X. (eds.) CARDIS 2008. LNCS, vol. 5189, pp. 104–117. Springer, Heidelberg (2008)

Creating Rules Using Abduction for Legal Reasoning by Logic Programming

Tanapon Tantisripreecha and Nuanwan Soonthornphisaj*

Department of Computer Science,
Faculty of Science,
Kasetsart University, Bangkok, Thailand
{g5317400457, fscinws}@ku.ac.th

Abstract. To create rules for a legal knowledge base, a knowledge from a lawyer cannot be represented in a predicate rule immediately. On the other hand, a programmer cannot create the rule to cover all of the legal knowledge base. Therefore, we need the procedure to suggest a legal expert. Indeed, the rules cannot be created only by facts or rules from legislation but also depend on the interpretation of each expert. In this paper, we proposed a procedure to create rules for legal knowledge base. First, a set of rules is created using *based rule*s and *compound rules* procedure. Second, the rules are selected by abduction for the legal reasoning process. Finally, we improved a Switch of Burden of Proof (SBP) [1] which is the legal reasoning system to examine the rules obtained from abduction. Thai Civil and Commercial Code is used as a knowledge base. The result showed that using our approach can find rules for an explanation of each legal case.

Keywords: logic programming, knowledge representation, legal reasoning, abduction rule, rules formalization.

1 Introduction

Research on legal reasoning is an interesting topic in legal informatics. Rules used in legal expert system [2] are generated by legal expert. The rules can be created in different ways depending on not only facts and rules in legislation but also different experiences of each expert. However, these knowledge bases are difficult to extend or modify.

Consider the predicate for cheque law; there are different styles of legal rules used in a knowledge base. Research works done by [3, 4] employed predefined predicate names to represent the concept as follows:

```
pay(plaintiff,defendant).
endorse(plaintiff,defendant).
```

* Corresponding author.

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 282–293, 2011.

The meaning of the predicate, `pay(plaintiff, defendant)`, is that a plaintiff pays a cheque to a defendant, and the predicate `endorse(plaintiff, defendant)` means that the plaintiff endorses the cheque to a defendant.

Satoh, *et al.* [1, 5] proposed the framework related to a switch of burden of proof. In this system, rules are created by the legal expert. They used a predicate naming, p, to represent facts as follows:

```
p(pay,plaintiff,defendant).
p(endorse,plaintiff,defendant).
```

The predicate `p(pay,plaintiff,defendant)` means that there is an evidence that a plaintiff had given a cheque to a defendant. The predicate `p(endorse,plaintiff,defendant)` means that a plaintiff gave a piece of evidence of endorse to the defendant.

As mentioned above, the predicates can represented only one meaning which is not sufficient to explain for legal reasoning. Therefore, we need more powerful knowledge representation to enhance the capability of the legal expert system. Since using such a predicate has some limitations in answering the questions, i.e. what is a type of cheque?, does a plaintiff sign a signature?, or can a plaintiff authorize endorsing to a defendant?

The objective of this work is to solve the problem as state above. We propose a procedure to create a set of rules, and to select the appropriate rule from the knowledge base for further inferences. We apply the switch of burden of proof [1] in our work and add new feature to the system in order to test the correctness of the rules. The dataset and case studies are based on Thai Civil and Commercial Code Law related to cheque law (section 898-1008).

2 Related Works

Knowledge base was realized as a normal abductive logic program and observation are either a literal or a rule. To generate the rule, Lakkaraju, S.K. and Zhang, Y [6] introduced a procedure to perform rule based abduction in knowledge base. A procedure of SLDNF resolution was employed to achieve the rule base abduction. The result of using this algorithm is that one can always find a minimal explanation for the observation. In 2003, Tuzet, G [7] dealt with the role of abductive inference in legal reasoning. Abduction was taken in proper explanatory sense. There were two kinds of abductive reasoning i.e. fact-finding and rule-finding. The fact-finding goes from effect to cause and the rule-finding goes from the characters of the accounted by abduction to the type of fact. In 2008, Ken Satoh *et al.* [5], used abductive reasoning for finding the rule to explain the legal case, and showed that an abductive framework is useful for reasoning about the next suitable measure in order to win case.

Graca, N and Quaresama, P [8] proposed the use of dynamic logic programming (DLP) to model the legal dynamic situation. The update and query commands were sent to a legal server. Therefore, the problem of laws that change over time and the problem of laws produced by different sources with different reliabilities were dealt. In 2007, Ken Satoh *et al* [1], introduced an approach to formalize a switch of burden of proof in legal reasoning. It can be formalized in non-monotonic reasoning by formalizing burden of proof by the Japanese civil procedure law. In 2009, Gordon T.F. [9] presented the Legal Knowledge Interchange Format (LKIF) which was a new interchange format developed specifically for application in the legal domain. The requirements of rule interchange language were summarized and were used to evaluate rule interchange languages in the legal domain.

3 Rules Creating

Sergot, M.J. *et al.* [3], proposed the formalization of legislation and the development of a computer system to assist with legal problem solving. It provided a rich domain to develop and test artificial-intelligence technology. The operators, *and/or*, was used to distinguish rules which required an expert to identify them.

In this work, we introduce a procedure to create rules in the legal expert system. There are two approach to create a set of rules; *base rule* and *compound rule*. Both of rules were designed as a top-down model, called *Horn clauses* [10]. The form of *Horn clauses* is as follows:

$$A \text{ if } B_1 \text{ and } B_2 \text{ and}, \ldots, B_n \tag{1}$$

Where B_1 *and* B_2 *and*, …, B_n are the conditions, and A denotes the rule effect which ought to follow when those conditions hold.

To create rules, *Horn clauses* were improved to both of *base rule* and *compound rule*. A set of rules that obtain from both approach will be selected by abduction for using in legal reasoning.

3.1 Based Rules

Based rules are created from facts. The name of *head_rule* is considered from noun which related to a case law e.g. person, evidence, or thing. There is only one argument in the predicate of the head of rule. This argument can be substitute by any variable that represent different meaning. In the body of a predicate, there is only one *sub_rule* which represents a fact in legislation. We employ ← as the implication operator, and ⇐ as a semantic operator. The structure of base rule can be written as follows:

$$head\_rule(Arg) \leftarrow sub\_rule\ (Arg_1, Arg_2, \ldots, Arg_n) \tag{2}$$

Consider an example of cheque case, a set of facts which involve in the scenario consist of drawer, payee, signature, and type of cheque. To create a based rule, the structure of a drawer's rule can be represented as follows:

```
drawer(A) ← payment(A1)
drawer(A) ← payment(A1,A2)
drawer(A) ← payment(A1,A2,A3)
drawer(A) ← payment(A1,A2,A3,A4)
A1 ⇐ drawer
A2 ⇐ payee
A3 ⇐ signature
A4 ⇐ cheque_type
```

Where `drawer` is a head of rule with one argument `A` (`A` is variable), and `payment` is a sub rule with argument `A`*s* (`A`*s* is a semantic meaning of each argument).

3.2 Compound Rules

A compound rule is used to create a set of rules in case that a *base rule* cannot be used because only one argument in the head of rule cannot represent a meaning of each semantic word. Sometimes a predicate needs several sub rules to explain the head of rule. Based rules and other compound rules are used together to create a set of rules. The structure of compound rule can be written as follows:

$$\begin{aligned}&head\_rule\ (Arg_{11}, Arg_{12}, ..., Arg_{1n})\\&subrule_1\ (Arg_{21}, Arg_{22}, ..., Arg_{2n})\\&subrule_2\ (Arg_{31}, Arg_{32}, ..., Arg_{3n})\\&\ldots\\&subrule_n\ (Arg_{m1}, Arg_{m2}, ..., Arg_{mn})\end{aligned} \tag{3}$$

For example, section *904* stated that *Holder means a person who is in possession of a bill as a payee or indorsee, or the bearer if the bill is payable to bearer* [11]. We found that there are a lot of rules related to holder such as possession, payee, endorser, bearer, transferring. In the formalization of compound rule, the structure of holder's rule can be represented as follows:

```
holder(A) ← possession(A) ∧ not_transfer(A).
possession(A) ← payee(A).
possession(A) ← endorsee(A).
possession(A) ← bearer(A).
bearer(A)← payee(A) ∧ cheque_type(bearer).
```

Considering a predicate `holder(A)`, there is an argument *A* which represents the holder. To prove this rule, the predicate holder needs two conditions, which obtain from the law [11] i.e. possession and transferring. Moreover, there are three conditions to prove a predicate `possession` which are payee, endorsee, and bearer. Predicate `bearer` must prove a type of cheque as a bearer. Therefore, the structure of predicate holder can be written in a form of *and/or* graph as follows:

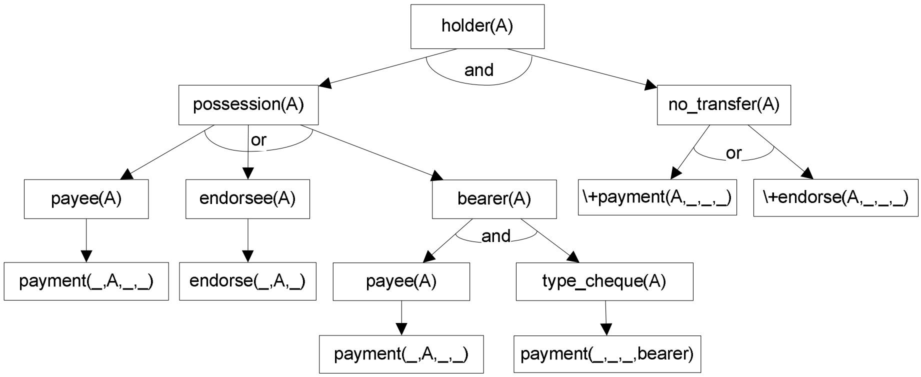


Fig. 1. *And/or* graph of holder(A)

In Figure 1, *base rule* are used to create the *compound rule*. Therefore, the system can find the truth of the predicate holder by using all of leaf nodes. However, some rule can be written together using or-operator (⋁) as follows:

`possession(A) ← payee(A) ⋁ endorsee(A) ⋁ bearer(A).`

Considering a possession rule, there are three kinds of bill i.e. bill of exchange, promissory note, and cheque. If `payee(A)`, `endorsee(A)` and `bearer(A)` are written together in `possession(A)`, this rule is valid for the cheque concept but it is not valid with other kinds of bill. Therefore, this rule should be written separately.

4 Abductive Reasoning

There are three types of reasoning; deduction, induction, and abduction. Abductive reasoning is a method of logic reasoning which is used for the guessing of an explanation to the observation. An abductive framework T can be defined as follows:

Definition 1. Given a set of rules $< T, A >$ (or called theory), a goal O (or called observation), and a set of possible hypotheses A (or called abducibles), abduction is used to find a set of explanation Δ such that $\Delta \subset A$ and Δ satisfies [12]:

1. $T \cup \Delta \models O$
2. $T \cup \Delta$ is consistent.

When the set of rules are created by *base rule*s and *compound rules*, they are selected by abduction to explain the legal case. To find the rule from based rule, a fact from user and a set of rules from rule based will be considered to select rules for legal reasoning. In the example of drawer, if the set of abducibles is {`payment(A)`, `payment(A1,A2)`, `payment(A1,..,An)`}, then all of

Δ1 = {`payment(A1)`}
Δ2 = {`payment(A1,A2)`}

Δ3 = {`payment(A1,A2,A3)`}
Δ4 = {`payment(A1,A2,A3,A4)`}

are the possible explanations for payment. We found that traditional expert systems prefer the predicate with less number of arguments [12]. But this is the drawback because the system cannot provide the complete explanation to user. Therefore we propose to use a set of facts provided by user to match among these arguments. The predicate with the highest number of matching arguments would be selected during the inference process. The advantage of our system is that the explanations provided by the system are more complete and easier to understand by user.

In case that the set of possible explanations are in the form of compound rules, we consider a set of rules which is a part of each compound rule. The way to select the base rule is the same as state above. In the example of holder (Fig. 1), if the set of abducibles is `possession(A)`, `not_transfer(A)`, `payee(A)`, `endorsee(A)` and `bearer(A)` then all of

Δ1 = {`possession(A), not_transfer(A)`}
Δ2 = {`payee(A), not_transfer(A)`}
Δ3 = {`endorsee(A), not_transfer(A)`}
Δ4 = {`bearer(A), not_transfer(A)`}
Δ5 = {`bearer(A), not_transfer(A), cheque_type(bearer)`}

are possible explanations for payment.

5 Experiment and Setup

5.1 Switch of Burden of Proof

We improved the framework of Satoh, *et al.* [1] in order to test the rules obtained from abduction. This system can be formalized in non-monotonic reasoning based on Prolog language. In equation 4, an argument `P` of predicate `proved` was sent to a predicate `p` in sub rule. The predicate `p` was introduced in order to reify plausibility of `P` and meaning of `p(P)` for an ultimate fact in a law is that there is a piece of evidence that a fact `P` is plausible. For example, a drawer predicate can be written as `p(drawer)`.

```
proved(P) :- p(P).                                                    (4)
```

We change the rule design to be the equation 5 that an argument `P` of predicate `proved` can be used to represent fact or rule. And the sub rule, P (the right hand side) will be proved to infer the fact or rule, P (the left hand side). For example, the drawer predicate can be represented as `drawer(A)`.

```
proved(P):-  P.                                                       (5)
```

The modified system can be applied using base rule and compound rule for legal reasoning shown in an appendix. Moreover, the framework of Satoh, *et al* [1] can be represented contexts of inference process as well.

5.2 Case Study

We studied about the liability of people in cheque law (see Fig. 2.). There are different roles among individuals such as drawer, endorser, endorsee, payee, and holder. The drawer is a person who pays the cheque to the payee. The endorser is a person who transfers the cheque to an endorsee, and the holder is a person who possesses the cheque.

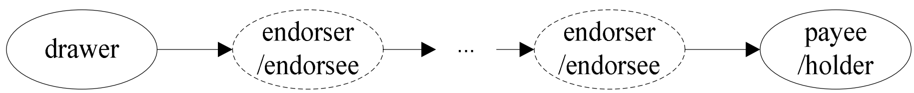


Fig. 2. Transferring of people in cheque law

During the court procedure, the judge considers the status of lawful holder by considering the rule about uninterrupted series of endorsement. The status of holder, drawer and endorser are proved. If each status is true, then the drawer or the endorser is liable to the holder. Unless the court cannot make a decision to the case, the authorizations of a signature are considered. The first person who transfers an unauthorized signature cheque is liable for the debt. Besides, the person who receives and transfers the cheque with unauthorized signature is liable as well. In case that the court considers only the issue of signature, there is no liability for those prior endorsers. The knowledge base can be written in Prolog code shown in an appendix.

Example 1. A drawer *a* pays a cheque to a payee *b*. after that *b* endorses this cheque to *c*. Therefore, `payment(a,b,a,-bearer)` and `endorse(b,c,b)` are input to the system. To ask the system whether *a* is a debtor of *c*. we ask the system using the sentence, `proved(liabiliy(a,c))`. Hence, the inference processes are done and the steps of proof are shown as follows:

```
proved(liabiliy(a,c))
=== Liability between drawer and holder. ===
  proved(lawful_holder(c))
  can_proved(lawful_holder(c))
  proved(drawer(a))
  can_proved(drawer(a))            a is a drawer
  proved(signature(a))
  can_proved(signature(a))
  can_proved(debtor(a,c))          a is a debtor of c
true.
```

The system proves the rule, `proved(liabiliy(a,c))`, by proving the uninterrupted series of endorsements between drawers and holders. In this step, *c* is proved as a lawful holder, *a* is proved as a drawer, and a signature of *a* is proved as true. Therefore, *a* is liable for *c*.

Various questions can be asked by users e.g. *b* is a debtor of *c* or not. If the question cannot be proved by first rule then it will find another rule to prove. Consider the question `proved(liabiliy(b,c))`, the inference processes are done and the steps of proofs are shown below:

```
  proved(liabiliy (b,c))
=== Liability between drawer and lawful holder. ===
  proved(lawful_holder(c))
  can_proved(lawful_holder(c))
  proved(drawer(b))
  cannot_proved(drawer(b))        b is not a drawer
  cannot_proved(lawful_holder(c))
=== Liability between endorser and lawful holder. ===
  proved(lawful_holder(c))
  can_proved(lawful_holder(c))
  proved(endorser(b))
  can_proved(endorser(b))         b is a endorser
  proved(signature(b))
  can_proved(signature(b))
  can_proved(debtor(b,c))         b is a debtor of c
true.
```

The system proved the rule, `proved(liabiliy(b,c))`, by proving the uninterrupted series of endorsement between drawer and holder. In this case, the system cannot prove that *b* is a drawer. Therefore, the next step proves uninterrupted series of endorsement between endorser and holder. In this step *c* is proved as a lawful holder, *b* is proved as an endorser, and a signature of *b* is proved as true. Therefore, *b* is liability of *c*.

Example 2. A drawer *a* pays a cheque to a payee *b*. after that *c* signs without authorization of *b* to *d*. Therefore, the `payment(a,b,a,-bearer)` and `endorse(b,d,c)` are input to the system. To ask the system whether *a* is a debtor of *d*. we ask the system using the sentence, `proved(liabiliy(a,d))`. Therefore, the inference processes are started and the steps of proof are shown below:

```
 proved(liabiliy (a,d))
=== Liability between drawer and lawful holder. ===
  proved(lawful_holder(d))
  cannot_proved(lawful_holder(d))    Δ1: d is not a lawful holder
=== Liability between endorser and lawful holder. ===
  proved(lawful_holder(d))
  cannot_proved(lawful_holder(d))    Δ2: d is not a lawful holder
```

```
=== Signature between drawer and payee. ===
  proved(\+lawful_holder(d))
  can_proved(\+lawful_holder(d))
  proved(payment(_G581,d,a,_G584))
  cannot_proved(payment(_G581,d,a,_G584))
  cannot_proved(\+lawful_holder(d))
```

∆3: *a* do not pay this cheque to *d* holder

```
=== Signature between drawer and endorsee. ===
  proved(\+lawful_holder(d))
  can_proved(\+lawful_holder(d))
  proved(payment(_G581,_G582,a,_G584))
  can_proved(payment(a,b,a,-bearer))
  proved(a\=a)
  cannot_proved(a\=a)
  cannot_proved(payment(_G581,_G582,a,_G584))
  cannot_proved(\+lawful_holder(d))
```

∆4: *a* do not sign this cheque for paying to *d*

```
=== Signature between endorser and endorsee. ===
  proved(\+lawful_holder(d))
  can_proved(\+lawful_holder(d))
  proved(endorse(_G581,_G582,a))
  cannot_proved(endorse(_G581,_G582,a))
  cannot_proved(\+lawful_holder(d))
```

∆5: *a* do not sign this cheque for endorsing to *d*

```
=== Signature between endorser and last endorsee. ===
  proved(\+lawful_holder(d))
  can_proved(\+lawful_holder(d))
  proved(endorse(_G581,d,a))
  cannot_proved(endorse(_G581,d,a))
  cannot_proved(\+lawful_holder(d))
  cannot_proved(debtor(a,d))
false.
```

∆6: *a* do not last sign this cheque for endorsing to *d*

The system proves the rule, `proved(liabiliy(a,d))`, but all of the rules cannot be proved whether *a* is a debtor of *d*. In step 1 and step 2, the inference process found that *d* is not a lawful holder, Therefore, *a* is not liability of *d*. In another step (step 3-6), the system consider for an authorized signature and *a* is not a debtor of *d*. Because, in step 3, the system found that *a* is not a drawer of this cheque. In Step 4, the concept of a lawful holder is determined and the system found that *a* is not a drawer. In step 5, the endorsement concept is determined and the system found that *a* does not sign this cheque for an endorsement to *d*. Finally in step 6, the system found that the signature of *a* is not the prior signature of *d*.

Next question, the system can be asked that who will be responsible for the reason mentioned above. Because we do not know who is a debtor of *d*. Hence, the open-ended questions are able to be used for the question. For example, `Debtor` is a variable of `proved(liabiliy(Debtor,d)`. Therefore, the inference processes are done and the steps of proof are shown as follows:

```
…
=== Signature between endorser and endorsee. ===
  proved(\+lawful_holder(d))
  can_proved(\+lawful_holder(d))
```

```
    proved(endorse(_G950,d,_G874))
    can_proved(endorse(b,d,c))
    can_proved(debtor(c,d))
  Debtor = c
```

Δ5: *c* sign this cheque for endorsing to *d*

The answer of the question: "who is a debtor of *d*", in a step of authorize signature between endorser and endorsee can be successfully proved. Because *d* is not the lawful holder and *c* sign the cheque without the authorization of *b* to *d*. Therefore, *c* is a debtor of *d*.

6 Conclusions

In this paper, we have introduced the procedure that could create a set of rules and select a rule using abduction. The legal expert can create the rules using our procedure to the legal knowledge base hence it can eliminate the gap between the lawyer and programmer. When a set of rules are provided in legal knowledge base, the rule that represents most of the fact will be selected for legal reasoning process.

We studied the structure of the rule in legal expert system and found that using a complicated rule is difficult to reuse for other types of bills (cheque, bill of exchange, promissory note). Nevertheless, the complicated rule needs less computational time. Therefore, we proposed to separate those rules in order to make them possible to reuse in other areas of bill of exchange law e.g. promissory note and bill of exchange. Another advantage of these separated rules is that they are able to provide the answers for the open-ended questions. This is important for the expert system in law domain because it increases user's belief and promotes the use of legal knowledge base in the future.

Acknowledgments. This research was partially supported by Department of Computer Science, Faculty of Science and The Graduate School of Kasetsart University, Thailand.

References

1. Satoh, K., Tojo, S., Suzuki, Y.: Formalizing a Switch of Burden of Proof by Logic Programming. In: Proceedings of The 1st International Workshop on Juris-Informatics (JURISIN 2007), pp. 76–85 (2007)
2. Satoh, K., Kubota, M., Nishigai, Y., Takano, C.: Translating the Japanese Presupposed Ultimate Fact Theory into Logic Programming. In: Proceedings of the 22nd Annual Conference on Legal Knowledge and Information Systems (JURIX 2009), pp. 162–171 (2009)
3. Sergot, M.J., Sadri, F., Kowalski, R.A., Kriwaczek, E., Hammond, P., Cory, H.T.: The British Nationality Act as A Logic Program. Communications of the ACM 29, 370–386 (1986)
4. Merritt, D.: Building Expert Systems in Prolog (1967)

5. Satoh, K., Tojo, S., Suzuki, Y.: Abductive Reasoning for Burden of Proof. In: Proceedings of the 2nd International Workshop on Juris-informatics (JURISIN 2008), Japan, pp. 93–102 (2008)
6. Lakkaraju, S.K., Zhang, Y.: Rule Based Abduction. In: Ohsuga, S., Raś, Z.W. (eds.) ISMIS 2000. LNCS (LNAI), vol. 1932, pp. 525–533. Springer, Heidelberg (2000)
7. Tuzet, G.: Legal Abduction. In: Bourcier, D. (ed.) The Sixteenth Annual Conference Legal Knowledge and Information Systems, Jurix 2003, Amsterdum, pp. 41–50 (2003)
8. Graca, N., Quaresma, P.: How to Model Legal Reasoning using Dynamic Logic Programming: a Preliminary Report. In: Proceedings of Legal Knowledge Information System. Jurix 2003: The Sixteenth Annual Conference, pp. 163–172. IOS Press, Amsterdam (2003)
9. Gordon, T.F., Governatori, G., Rotolo, A.: Rules and Norms: Requirement for Rule Interchange Languages in the Legal Domain. In: Governatori, G., Hall, J., Paschke, A. (eds.) RuleML 2009. LNCS, vol. 5858, pp. 282–296. Springer, Heidelberg (2009)
10. Truszczyński, M.: Logic Programming for Knowledge Representation. In: Dahl, V., Niemelä, I. (eds.) ICLP 2007. LNCS, vol. 4670, pp. 76–88. Springer, Heidelberg (2007)
11. Thai Civil and Commercial code, section 898-1005 (1992)
12. Ma, J.: Distributed Abductive Reasoning System and Abduction In the Small. In: [MEng] Computing Final Year Project Report (2007)

Appendix: Prolog Program

```
:- dynamic endorse/2, endorse/3, payment/4, indent/1.
%%  Fact
payment(a,b,a,-bearer).  endorse(b,d,c).

%%  Legal reasoning process
proved(Rule):-    tab(2),print(proved(Rule)),nl,fail.
proved(Rule):-    Rule,
proved(Rule):-    tab(2),print(cannot_proved(Rule)),nl,fail.


%%  Base rule
drawer(A):-       payment(A,_,_,_).
payee(A):-        payment(_,A,_,_).
signature(A):-    payment(_,_,A,_).
signature(A):-    endorse(_,_,A).
cheque_type(A):-  payment(_,_,_,A).
endorser(A):-     endorse(A,_,_).
endorsee(A):-     endorse(_,A,_).
no_transfer(A):-  \+endorse(A,_,_).
bearer(A):-       payee(A),cheque_type(bearer).

%%  Compound rule
holder(A):-       prossession(A),no_transfer(A).
lawful_holder(A):-drawer(D), holder(A), uninterrupted(D,A,D).
prossession(A):-  payee(A).
prossession(A):-  endorsee(A).
prossession(A):-  bearer(A).
send_to(A,B,A):-  payment(A,B,A,_).
```

```
  send_to(A,B,A):-  endorse(A,B,A).
  uninterrupted(P,H,P):- send_to(P,H,P).
  uninterrupted(P,H,P):- send_to(P,X,P),uninterrupted(X,H,X).

  %%  Steps of proof
  debtor(A,B):- print(' === Liability between drawer and lawful
holder. ==='), nl,
      proved(lawful_holder(B)),
      proved(drawer(A)),
      proved(signature(A)).
  debtor(A,B):-  print('===  Liability  between  endorser  and
lawful holder. ==='), nl,
      proved(lawful_holder(B)),
      proved(endorser(A)),
      proved(signature(A)).
  debtor(A,B):- print('=== Signature between drawer and payee.
==='),nl,
      proved(\+lawful_holder(B)),
      proved(payment(_,B,A,_)),
      proved(holder(B)).
  debtor(A,B):-  print('===  Signature  between  drawer  and
endorsee. ==='), nl,
      proved(\+lawful_holder(B)),
      proved(payment(X,D,A,_)),
      proved(X\=A),proved(holder(B)),
      proved(uninterrupted(D,B,_)).
  debtor(A,B):-  print('===  Signature  between  Endorser  and
endoresee. ==='), nl,
      proved(\+lawful_holder(B)),
      proved(endorse(_,X,A)),
      proved(uninterrupted(X,B,_)).
  debtor(A,B):- print('=== Signature between Endorser and last
endorsee. ==='), nl,
      proved(\+lawful_holder(B)),
      proved(endorse(_,B,A)).
```

Supporting the Harvard Model of Principled Negotiation with Superexpertise

Xenogene Gray[1], Pamela Noel Gray[2], and John Zeleznikow[3]

[1] Department of Computing, Faculty of Science,
Macquarie University, Australia
xen@grayske.com
[2] Centre for Research in Complex Systems,
Charles Sturt University, Bathurst, Australia
pam@grayske.com
[3] Laboratory of Decision Support and Dispute Management,
School of Management and Information Systems,
Victoria University, Melbourne, Australia
John.Zeleznikow@vu.edu.au

Abstract. An expert epistemology is a theory about knowledge in the expertise and includes knowledge representation, semantics, specifications, heuristics, reasoning, etc. Superexpert systems employ computer capabilities to extend ordinary human abilities, and are derived from expert epistemologies. A superexpert system, Negotiation Game or NeGame (NeG), is designed to support the Harvard Principled Negotiation model; it is illustrated by a Civilisation application which reframes the Israel-Palestine conflict as the task of negotiating a mutually acceptable civilisation. The negotiation epistemology of NeG evolved from the adversarial epistemology of eGanges (eG); both handle tasks in a user-friendly and transparent way. NeG manages hierarchical complexity of the conflict issues and differences in their subjective values, advises on cumulative scoring of Wins and Losses, and, through mathematical techniques, maximises Win-Win options. Adversarial epistemology requires four-valued logic, whereas negotiation epistemology requires six-valued logic; they share a common knowledge hierarchy, called a River.

Keywords: epistemology, superexpertise, eGanges, Principled Negotiation, quality control fishbone.

1 System of Legal Epistemology

An expert epistemology can be derived from expertise, such as legal practitioner's expertise [1]; it can include knowledge representation, semantics, specifications, heuristics, and reasoning. Compared to human experts, computers can have larger memories, faster retrieval, and faster processing of the possible combinatorics of complex logic. Superexpertise uses these computer capabilities and is elicited from expert epistemology; it can be deployed to support Principled Negotiation of civilisation.

W. Abramowicz, L. Maciaszek, K. Węcel (Eds.): BIS 2011 Workshops, LNBIP 97, pp. 294–305, 2011.

1.1 Two Types of Legal Problems

In the legal domain, there are two types of problems in resolving and avoiding conflicts for clients: adversarial problems and negotiation problems. These types of problems use different but overlapping epistemologies. Their similar epistemological features are, firstly, the use of finite information, and secondly, their suitability for a knowledge representation based on a River paradigm. Whereas the superexpert adversarial shell, eGanges (**e**lectronic **G**lossed **a**dversarial **n**ested **g**raphical **e**xpert **s**ystem) (eG) [2], uses the paradigm of tributaries joining to more major Rivers which all lead to a single River mouth (the Final result node), NeGame (NeG) is designed as a superexpert negotiation shell which uses a River delta paradigm where a single subjective valuation source node flows and splits along various branches which have ends in several delta-mouth nodes.

Adversarial problems apply enforceable rules to determine a winner between opposing parties. A court case is an adversarial problem example; the rules of law are applied to determine who wins the case. Negotiation problems attempt to resolve issues in dispute between parties in conflict, through cumulative agreement, without a predetermined enforceable rule set being applied; there may be no enforceable rule set, no recognised neutral party to enforce such a rule set, or the parties may prefer not to enforce a rule set as it could cost them more. Differences between adversarial and negotiation epistemologies arise from these distinctions.

Negotiation usually precedes adversarial resolution. However, an agreed settlement of a conflict is *prima facie* enforceable. A suitable alignment of adversarial and negotiation epistemologies may assist transformation of negotiated agreement into an adversarial framework. NeG is aligned with eG and designed to provide an aid for Principled Negotiation throughout a negotiation, and give a precise evaluation of each point in the cumulative agreement; not just of known possible alternatives, but also of newly realised options devised during the negotiation.

1.2 Three Features of Legal Epistemology

There are three features of legal epistemology that are used in an alignment of adversarial and negotiation epistemologies to produce NeG: English common law epistemology, Bologna glosses and the Harvard model of Principled Negotiation.

The practice of common law implemented and expanded its rule system as if it were like a River tributary structure. To some extent this structure was visualised by Gray's Inn lawyer, Fraunce [3, 4], a Ramist, who was a Reformation forerunner of legal knowledge engineering; the Ramist school of logic at the Sorbonne used graphical representations of logic, sometimes with three dimensional logic space [5].

A gloss is related information peripheral to the rules, and originates from the Bologna margin notes of the resumed study of the Roman Civil Code in the eleventh century AD. eG allows for glossing of each node to assist the user to give informed answers, and to resolve issues of law. In NeG, glossing may be used to expand on issue nodes and simplify or clarify further detail, without further complicating the River graphic.

The Harvard Negotiation Project [6, 7] produced a model called 'Principled Negotiation' which identified seven elements: interests, options, alternatives, legitimacy, communication, relationship, and commitment. Spencer [8, p.23-4] suggests that reframing conflict for negotiation is a communication skill. Howsoever a conflict is described or perceived, is the way it is framed. Mayer [9, p.139] captured the resource of reframing well:

> The art of reframing is to maintain the conflict in all its richness but to help people look at it in a more open-minded and hopeful way.

A major element of principled negotiation is alternatives. This refers to the alternatives to the BATNA (Best Alternative To a Negotiated Agreement), a concept established in the Harvard model [8, pp.33-5]. Reframing may prevent any limitations on the formulation of alternatives that the BATNA presupposes.

2 Reframing: Civilisation River

An example of the delta River structure is shown in a NeG application, called Civilisation. The Israel-Palestine conflict is reframed as a task of negotiating a civilisation acceptable to both the parties (cf. [10]). This reframing moves the conflict from a competitive to a cooperative and collaborative orientation. The definitional hierarchy of some issues to be resolved in the Civilisation River is shown in the Rivers window of Figure 1. Further definition may be expanded or adjusted during the negotiation. Figure 2 illustrates the nested nature of the River system: it is the further detail of the negotiation issue node, Environment, in Figure 1. Nesting is required as the sub-branches from the Environment node are too extensive to be seen in one window with the non-Environment branches.

In Figure 1, 'Environment' is seen to be on the same branch as 'System of law' through to 'Foreign policy'; this is the primary branch which has the Civilisation node as its encompassing node. An arrow on the branch, pointing from the Civilisation node along the primary branch indicates that the Civilisation node is the encompassing node of all nodes on the primary branch; the issues covered by the nodes on the primary branch are all sub-issues of the Civilisation issue. Sub-issues on the primary branch may be encompassing nodes on their sub-branches. For instance, the branch consisting of the 'Land tenures', 'Co-ownerships' and 'Disposition' nodes, is a sub-branch of the Environment node, as can be seen in Figure 2. Sub-issues which do not have sub-branches, wherever they occur, are called delta-mouth nodes; they are where negotiation becomes decisive.

In NeG, issues are clustered into branches and flow directions are used to simplify the normalisation of subjective values given to issues by the parties. The subjective values of each party are normalised relatively amongst the group of nodes on the same branch; if a node is also an encompassing node for another River branch, that node's normalised value will be propagated along its sub-branches to aid in the fair

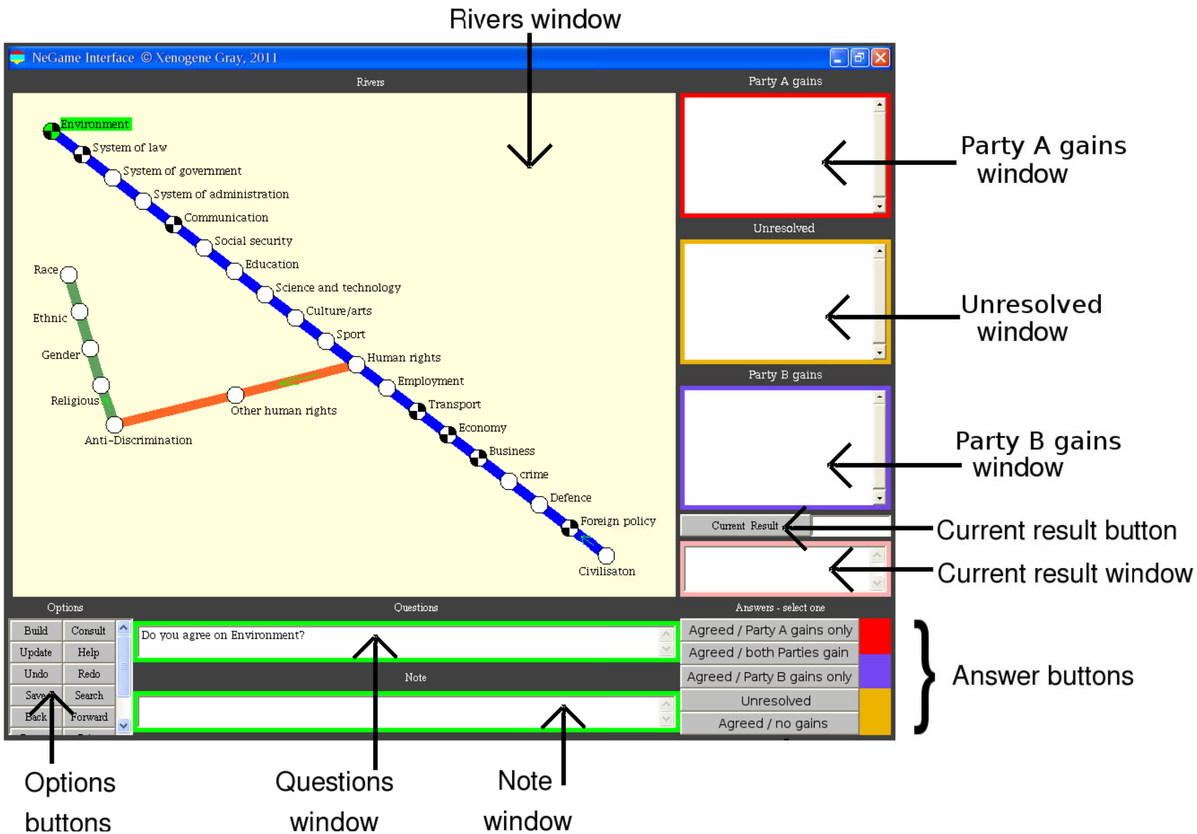


Fig. 1. NeGame interface with Initial map of Civilisation application in Rivers window. © Xenogene Gray and Pamela N. Gray, 2011.

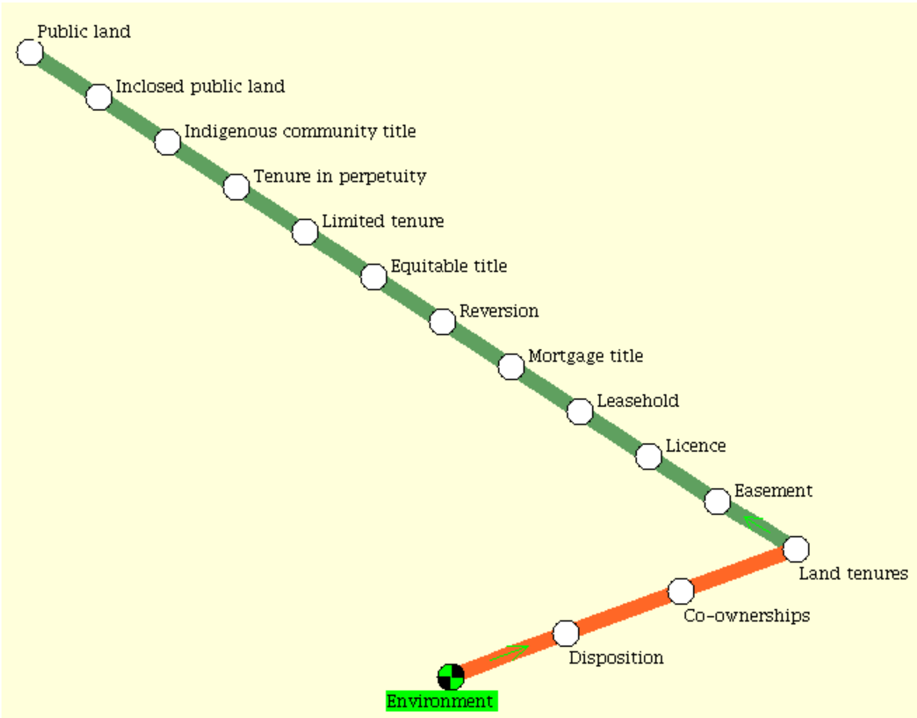


Fig. 2. Submap for Environment © Xenogene Gray and Pamela N. Gray, 2011

assessment of gains. The paradigm of the delta may secure quality control in Principled Negotiation of civilisation, just as the tributary River secures fishbone quality control in an adversarial system of rules [11].

The River graphic in the NeG interface provides transparency of complex hierarchical issues and allows the creation of alternatives for negotiation at any level of the hierarchy. In NeG, issue points may be renegotiated at any time, with or without expansion of options, to achieve a bargain; options can be a continuation of the issues for negotiation. Repeated reframing might assist better understanding of each other's interests and positions, and also reveal solutions that suit or advance both parties; it may create a definition of the conflict, acceptable to both parties, that increases the potential for a common agreed solution.

In formulating a NeG application, River branches can be nested to enable as extensive a River delta as the conflict and potential solutions require; the River allows for consistent option generation as issues are refined. This resolves the difficulty recognised by Menkel-Meadow [12, p.772]:

> … it may be impossible to represent graphically the negotiation of a complex, multi-issue transaction as a two dimensional structure, without imagining a many-planed axis with hundreds of potential coordinates.

Menkel-Meadow also asserted that adversarial negotiation inhibits the creative option generation posed in the Harvard model; a negotiation issue may be modified or added by the parties, whereas an adversarial rule can only be varied by a rule-making authority. The overarching principle of conflict resolution by negotiation is that sufficient issues must be resolved by agreement to reach a Final outcome; it is not an adversarial matter where sufficient rules of law must be applied to determine a Final outcome of the conflict. The River paradigms are expansive and malleable for law-making and for negotiation.

3 Adversarial Superexpertise: eGanges Rivers as Graphical Rules

The eGanges River is a refinement of the quality control fishbone of Ishikawa [11, 13]. The basic knowledge units of eGanges are simple questions, each represented by a node that, with definitive knowledge, is a dichotomy; e.g. “Did you drive the car?”. However, eGanges does not assume users always have a definitive answer to a question, so it allows a third possible answer, uncertain. Uncertain represents a type of unknown, due to evidential uncertainty. Another type of unknown is due to no evidence; this is called unanswered and indicates that further evidence is required. If there is a final result of uncertain or unanswered, the distinction also assists a court to rule on the evidence; reasons for the resolution which may be recorded.

Thus, each node can have one, and only one, of four possible logic values: Positive (true), Negative (false), Uncertain or Unanswered. These are epistemic logic values of the adversarial domain. For processing, these four logic values amount to a complete lattice [14], with a logic ordering whereby: positive > unanswered > uncertain > negative. This means that the logic value of Positive is logically 'greater than' Unanswered, which is logically 'greater than' Uncertain, which is logically 'greater than' Negative. The eG epistemology uses this four-valued logic to resolve who wins the adversarial dispute, as each value in respect of each question may result in a different outcome.

The use of multi-valued logics was posed by Lukasiewicz [15], and extended by Kleene [16], Belnap [17], Ginsberg [18], Fitting [19], Gray and Gray [2], and Majkić [20]. The distinction between the two types of unknowns is epistemically significant and was introduced in [2], followed by a more detailed semantic analysis of an equivalent four-valued logic [20].

An eG River is a graphical representation isomorphic to a four-valued logic Horn clause system and shows the hierarchical nature of the rules; each tributary of the River is a rule and the rules are interlocked where they share the same antecedent or consequent node; the same node can be an antecedent in some rules, and simultaneously the consequent of others. Rules with the same consequent indicate a disjunction; there are alternative ways of establishing a consequent so that the tributaries appear as a fan. A consequent node is at one end of a tributary, and is differentiated from the antecedent nodes of the tributary by an arrow on the tributary graphic that points to it; this represents the inference arrow in the rule.

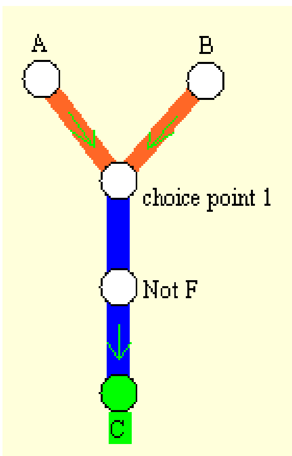
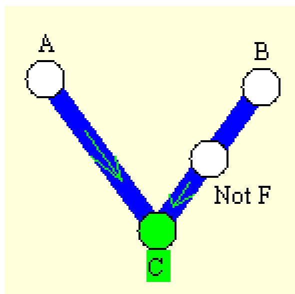
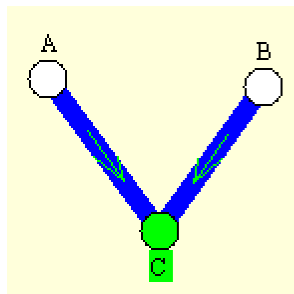
In Table 1, the leftmost eG River graphic shows three River tributaries linked together into a single River system. The three Horn clauses are:

1) 'choice point 1' ← A
2) 'choice point 1' ← B
3) C ← 'choice point 1', 'Not F'

Together, 1) and 2) constitute a fan with the common consequent, 'choice point 1' (some other node label could be used, e.g. D).

An eG River can be as extensive as the rule system requires. The logic value of the rule head (consequent) must be kept greater than or equal to the logic value of the rule body (minimum antecedent value) in order to prevent inconsistencies; e.g. if every antecedent is Positive, the rule body is Positive, therefore the rule head must have a logic value of positive. The eGanges heuristics go further by ensuring decidability which is captured in the truth tables of Table 2; the multi-valued versions of de Morgan's laws apply.

Table 1. Comparison of eG Rivers and equivalent AnsProlog programs. © Xenogene Gray, 2011.

<table>
<tr><td>
</td><td>
</td><td>
</td></tr>
<tr><td>C ← A, not F</td><td>C ← A</td><td>C ← A</td></tr>
<tr><td>C ← B, not F</td><td>C ← B, not F</td><td>C ← B</td></tr>
<tr><td>¬ C ← F</td><td>¬ C ← F, not A</td><td>¬ C ← F, not A, not B</td></tr>
</table>

Column 1 of Table 1 shows the case where **not** F must be explicitly included in the River as it is dominant over both A and B. In the second column **not** F only dominates B. For the third column, **not** F is an unnecessary node as the eG heuristics ensure C is unanswered by default while there is a chance of an upstream making it true. However, once there is no chance for C to be true from an upstream (in this example, both A and B are established as false), C becomes false by the eG heuristics, thereby fulfilling the (¬ C ← F, **not** A, **not** B) rule by default.

Table 2. Conjunctive and Disjunctive eG Truth table for three rules: (C ← A, B) and (D ← A) and (D ← B). ©Xenogene Gray, 2011.

	A=Negative	**A=Uncertain**	**A=Unanswered**	**A=Positive**
B=Negative	C=Negative D=Negative	C=Negative D=Uncertain	C=Negative D=Unanswered	C=Negative D=Positive
B=Uncertain	C=Negative D=Uncertain	C=Uncertain D=Uncertain	C=Uncertain D=Unanswered	C=Uncertain D=Positive
B=Unanswered	C=Negative D=Unanswered	C=Uncertain D=Unanswer ed	C=Unanswered D=Unanswered	C=Unanswered D=Positive
B=Positive	C=Negative D=Positive	C=Uncertain D=Positive	C=Unanswered D=Positive	C=Positive D=Positive

In eG, when a question is answered by the end user, the answer input transforms the predicate embedded in the node into a logic literal. For example, “Were you informed of the terms of the contract?” becomes a logic literal when the two semantic terms “you” and “the contract” are implicitly grounded by an end user's answer.

Explicit negation is embedded into the nodes' resulting literal in eG; for example, “Were you uninformed of the terms of the contract?” is the negative literal related to the above literal. The nodes, River structure and processing heuristics hide the *not by default* aspect nodes, as they do not need to be explicitly shown in the River system given how intuitive it is to work them out. Shown in Table 1 are the eG Rivers related to three possible consistent AnsProlog programs. They illustrate the management of disjunction in eG rules. A sample finance law application can be trialled at: www.grayske.com/FinLawTrial/index.html

4 Negotiation Superexpertise: NeG Normalisation for Distribution

Rather than representing a rule system, a delta River in NeG is used to graphically represent a hierarchical order of the issues to be negotiated and the complex choice of subjective values in relation to available selections. The delta structure arises from a breakdown of the conflict into its component issues, sub-issues and so on. Each issue is represented by a particular node, with the more detailed aspects of that issue (i.e. its sub-issues) represented in a River sub-branch (a collection of nodes); this River sub-branch is initiated by an issue node called the encompassing node. The nodes detailing the sub-issues of this encompassing node are in a sub-branch of the branch that contains the encompassing node as a sub-issue; a hierarchical River structure is formed from these node overlaps. This makes NeG's delta River layout structurally similar to the tributary River system of eG, but eG's antecedent nodes are replaced by NeG's sub-issue nodes, and the eG consequent node is replaced by NeG's encompassing node; in NeG, the same node can be both a sub-issue node in one branch and the encompassing node for another branch. In NeG, there may be an overlap of sub-issue nodes and encompassing nodes, whereas in eG there may be an overlap of antecedents and consequents in different rules. eG may have disjunctions where two different rules share a common consequent. However, generally NeG does not have disjunctions, since its task is to proceed through each issue sequentially to see if it can be agreed; in NeG, de Morgan's laws are not relevant.

For NeG, the subjective valuations of each node by each party, flow in the opposite direction to the logic values flow in eG. This means the subjective values flow from the encompassing node 'down' the delta River sub-branches, as shown by the arrows in Figures 1 and 2 which point away from each encompassing node. This reversal of the flow makes NeG's River paradigm like a River delta, where a single River source flow is divided among several delta-mouths; the total normalised subjective valuation of each party in conflict is divided among the various issues under negotiation.

NeG nodes deal with two aspects of an issue: 1) its relative importance to both parties, and 2) who wins on the issue. The nodes on a branch are all related as the issues further detailing the branch's encompassing node issue, so all of a branch's issues need to be given a relative weight with respect to the other issues on the same branch; once all the nodes on a single branch have been given their normalised relative valuation, by both parties, then the gains of each party can be determined by assessing who wins for each issue, given the relative subjective weights of each party.

Subjective values given must be numerical but there is a range of different numerical attributions that may be used by each party; the choice is irrelevant as NeG convert them all to a normalised fraction. For example, the node 'Human rights' in Figure 1 has two more detailed aspects of the issue of human rights listed in its sub-branch, namely 'Anti-discrimination' and 'Other human rights'. Other issues that detail 'Other human rights' may be added in a new sub-branch as the draft application is further developed. Each party may differently value each of these sub-issues, using different numerical indicators. It is mainly the nodes with no sub-branches that are in contention; these are called delta-mouth nodes and are the most detailed sub-issues in the issue hierarchy. When a delta-mouth node is answered, with normalised subjective valuations from both parties, and one of the six sorting values is agreed by both parties, it is placed in the appropriate feedback window as follows:

1) If a 'Win/Win' resolution is agreed, the label of the node appears in both the 'Party A gains' and the 'Party B gains' windows with '(Win-Win)' before the label.
2) If a 'Party A gain' resolution is agreed, the label of the node appears only in the 'Party A gains' window.
3) If a 'Party B gain' resolution is agreed, the label of the node appears only in the 'Party B gains' window.
4) If a 'Lose/Lose' resolution is agreed, the label of the node appears in the unresolved window with '(Lose-Lose)' before the label.
5) If an 'Unresolved' value is assigned to a node by the parties, the label of the node appears in the unresolved window.

Each node is given a fraction of each negotiating party's total normalised valuation. When a node is listed as a gain for that party, their valuation for that node is added to their total gains. As the negotiation proceeds, issues may be resolved cumulatively and the current gains for each party can be accessed at any time to see if there is an acceptable bargain. While a party might not gain on a specific node, NeG will record these losses, and this can be used to justify that party gaining on other issues they value as more important. Eventually, a net Win-Win for both parties may be agreed.

The NeG epistemology uses a six-valued logic for two parties, as there are only six possible outcomes. Even in multi-party negotiation, there are a finite number of possible states of an issue, and these states range over more than a single dimension of concern; there is the dimension of whether Party A gains, and a separate dimension of whether Party B gains, etc. Compacting all the possible states into a one dimensional value assignment eases processing both by the computer (with the use of appropriate multi-valued 'truth tables') and potentially by the user. For two parties NeG gives a choice of five answers, as shown in Figure 1 and the appropriate feedback window list above; both parties must agree on their single joint answer, or select the unresolved option. There is a sixth default value representing no answer; this makes NeG a six valued logic.

The normalisation of values ensures that the sum of the delta-mouth node values equals 1 for each party; normalised comparison of the values of each party for each node is then possible and gains can then be allotted and compared.

For example, in Figure 1, the node 'Human rights' is one of 18 nodes on the primary branch of the delta. 'Human rights' has a sub-branch containing the nodes 'Anti-discrimination' and 'Other human rights'. If the negotiating parties give 'Human rights' the same normalised subjective value, compared to the other nodes on the primary branch, of 0.05 then this is the weighting factor for both parties when determining the net weighting for all sub-branch nodes of the encompassing 'Human rights' node.

A node's weighting factor is the normalised value of all its upstream encompassing nodes in the hierarchy multiplied together. For instance, suppose that one party (P1), following the normalisation calculation, gives a normalised subjective fraction of 0.6 to 'Anti-discrimination' and therefore 0.4 to 'Other human rights'. This means for P1 that 0.05 (from 'Human rights') x 0.6 (from 'Anti-discrimination') = 0.03 propagates up the 'Anti-discrimination' sub-branch as P1's weighting factor of 0.03, which will be multiplied by the normalised fraction of every node in this sub-branch to determine each node's true relative valuation.

In summary, there are two reasons to propagate weighting factors downstream. The first reason is to ensure that the sum of all the delta-mouth subjective valuations by any one party always sums to 1; this normalisation to 1 requires a delta-mouth node's true relative valuation be equal to its normalised fraction multiplied by its encompassing node's weighting factor. Normalisation ensures that a true relative valuation between parties can be performed; if P2 values 'Other human rights' relatively more, compared with every other delta-mouth issue in the entire delta system, than P1, then, in the negotiation, P2 gains more of their relative valuation from being assigned 'Other human rights' than P1 would gain. This resulting difference can be used to provide both parties with more than half their total valuations.

The second reason to propagate weighting factors downstream is to calculate gains as soon as possible, instead of requiring the parties, before negotiation can begin, to assign values to all the detailed issues (in order to normalise the values); the parties can assign values simply to encompassing nodes and then know relatively how much of their valuation will be devoted to the more detailed issues downstream of the encompassing nodes. The parties' values are not permanently locked in, and they can change their relative values over the course of the negotiation. However, each change a party makes will affect the net values of other branches, as their values are always normalised. The calculation of values for the issues also provides a calculated assessment of alternatives to the BATNA.

When a party agrees to gain on a delta-mouth issue, their valuation of that issue is added to their gains total. Pressing the Current result button will result in NeG giving, in the Current result window, the total gains, thus far, for both parties.

Some issues have mutually exclusive solutions i.e. only one party can gain from the solution, for example the allotment of particular property that can not be shared. Mutual exclusivity is a matter of negotiation epistemology for further study that is outside the scope of this paper. Subjective values for mutual exclusivity, relative to shared solutions, and the implication of differences upstream for each that may vary depending on who gains, will be considered in subsequent work.

5 Conclusion

The user-friendly NeG interface obtains instructions according to its six-value sorting requirements, and receives input on the subjective values of the parties which reflect their relative interests. Mathematical normalisation of subjective values, and the propagation of those values lends legitimacy by providing an objective metric of fairness for proposed negotiation solutions. Feedback throughout interrogation appears in windows that monitor cumulative relative gains of the parties and unresolved issues. This is a basis for fostering a trust relationship between negotiators and their commitment.

NeG allows quick construction and alteration of a River system and its glosses to permit the civilisation application to be fully expressed and particularised, for negotiation in an ongoing way, as suggested by the Harvard model. The parties may record the substance of their negotiations in the Notes window. The subjective value input of both parties is also available as gloss information, and summarised in the Current result window. NeG provides only for two party conflicts, but the design could be extended for multi-party negotiations.

NeG enhances all seven aspects of Principled Negotiation. The use of a quality control representation of issues should reinforce legitimacy and foster commitment of the negotiators; as a means of complex communication and processing, it should improve the common understanding and relationship of the negotiators. The communication system of NeG assists the management of clusters of issues in a broader hierarchy as a framework for identifying delta-mouth issues which are decisive, and for expanding options at any level of the hierarchy; this assists deconstruction of complex conflicts and may reveal ways to extend Win-Win solutions.

References

1. Gray, P.: Legal Knowledge Engineering Methodology for Large Scale Expert Systems, Volumes 1 and 2, PhD thesis, University of Western Sydney (2007), http://trove.nla.gov.au/work/3901700
2. Gray, P., Gray, X.: A map-based expert-friendly shell. In: Bourcier, D. (ed.) Legal Knowledge and Information Systems. IOS Press, Amsterdam (2003)
3. Fraunce, A.: The lawyer's logic, Reproduced from the original manuscript in the British Museum, Lawiers Logike, first published by William How, London, The Scolar Press Limited, Menston, Yorkshire, England (originally published 1588, 1969)
4. Gray, P.N., Mann, S.: The Fraunce (1588) model of case-based reasoning. In: Proceedings of 9th International Conference on Artificial Intelligence and Law. ACM, New York (2003)
5. Ong, W.J.: Ramus method and the decay of dialogue. Harvard University Press, Cambridge (1958)
6. Fisher, R., Ertel, D.: Getting Ready to Negotiate. Penguin, New York (1995)
7. Fisher, R., Ury, W., Patton, B.: Getting to Yes. Random Century, London (1991)
8. Spencer, D.: Essential Dispute Resolution, Cavendish Australia, Sydney (2005)
9. Mayer, B.: The Dynamics of Conflict Resolution. Jossey-Bass, San Francisco (2000)
10. Zeleznikow, J., Bellucci, E.: Classifying Online Dispute Resolution by comparing family mediation and the Israel – Palestinian conflict. In: Proceedings of 13th International Conference on Artificial Intelligence and Law. ACM, New York (2011)
11. Gray, P., Gray, X.: Quality Controlled Government with Spherical Logic. Int. J. Interdisciplinary Soc. Sci. 5(10), 271–284 (2011), http://iji.cgpublisher.com/product/pub.88/prod.1305
12. Menkel-Meadow, C.: Toward another view of legal negotiation: the structure of problem solving. University of California Law Review 31, 755 (1984)
13. Ishikawa, K.: What is total quality control? The Japanese way, translated by David J. Lu. Prentice-Hall Inc., Englewood Cliffs (1985)
14. Blyth, T.S.: Lattices and Ordered Algebraic Structures. Springer, San Francisco (2005)
15. Lukasiewicz: O logice trojwartosciowej (On three-valued logic). Ruch filozoficzny 5, 170–171 (1920)
16. Kleene, S.C.: Introduction to Metamathematics. North Holland, Amsterdam (1952)
17. Belnap, N.: How a computer should think. In: Ryle, G. (ed.) Contemporary Aspects of Philosophy, pp. 30–55. Oriel Press, Stocksfield (1976)
18. Ginsberg, M.L.: Multivalued logics: A uniform approach to inference in artificial intelligence. Computational Intelligence 4(3), 256–316 (1992)
19. Fitting, M.: Many-valued non-monotonic modal logics. In: Nerode, A., Taitslin, M.A. (eds.) LFCS 1992. LNCS, vol. 620, pp. 139–150. Springer, Heidelberg (1992)
20. Majkić, Z.: Ontological encapsulation of many-valued logic. In: Proceedings of Convegno Italiano di Logica Computazionale (2004), http://www.cs.unipr.it/CILC04/DownloadArea/Majkic-CILC04.pdf

Author Index