Witold Abramowicz
Dominik Flejter (Eds.)

Business Information Systems Workshops

LNBIP 37

BIS 2009 International Workshops
Poznan, Poland, April 2009
Revised Papers

Springer

Lecture Notes in Business Information Processing 37

Witold Abramowicz Dominik Flejter (Eds.)

Business Information Systems Workshops

BIS 2009 International Workshops
Poznan, Poland, April 27-29, 2009
Revised Papers

Springer

Volume Editors

Witold Abramowicz
Dominik Flejter
Poznań University of Economics
Department of Information Systems
Al. Niepodległości 10, 61-875 Poznań, Poland
E-mail: {W.Abramowicz,D.Flejter}@kie.ue.poznan.pl

Library of Congress Control Number: 2009932134

ACM Computing Classification (1998): J.1, K.5, H.3.5, H.4

ISSN 1865-1348
ISBN-10 3-642-03423-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-03423-7 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12723096 06/3180 5 4 3 2 1 0

Preface

Over the last few decades, business information systems have been one of the most important factors in 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. During the last 12 years these innovations were carefully observed but also shaped by researchers attending the annual International Conference on Business Information Systems (BIS).

Recently, apart from the main conference, workshops covering specific topics in the area of business information systems have been organized. In 2007 and 2008, the BIS conference featured two and three workshops, respectively. Their proceedings were published on-line and outstanding workshop papers were included in two special issues of international journals. This year nine workshops were successfully organized in conjunction with BIS 2009, covering the topics of Deep Web (ADW), applications and economics of knowledge-based technologies (AKTB, ECONOM), SOA (SDS-SOA), legal IT (LIT), social Web and Web 2.0 (SAW, Enterprise X.0), e-learning (EeLT) and enterprise systems in higher education (ESHE). This volume contains papers that were accepted and presented at the BIS 2009 workshops. Additionally it features the BIS 2009 keynote speech by Wil van der Aalst, as well as two invited speeches presented at LIT and ECONOM / Enterprise X.0 workshops.

Workshop papers included in this volume were subject to a thorough review procedure. Each submitted paper received from two up to five reviews (a total of 208 reviews were received, with an average of 2.9 reviews per paper) by over 130 members of the Program Committees of the workshops. Out of 72 full papers, demo papers and work-in-progress reports submitted, 34 were accepted and presented at the conference (accounting for 47% of all submissions). One of the papers presented at the ECONOM workshop, which received the BIS 2009 best paper award is also included here.

On this occasion, we would like to express our thanks to everyone who made the BIS 2009 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 next year at the 13th BIS Conference in Berlin.

April 2009

Witold Abramowicz
Dominik Flejter

BIS 2009 Conference Organization

Program Chair

Witold Abramowicz

Organized by:

The Poznań University of Economics, Department of Information Systems, Poland

Organizing Committee

Elżbieta Bukowska
Dominik Flejter
Konstanty Haniewicz
Lidia Leszczyńska
Jolanta Skorwider
Monika Starzecka
Piotr Stolarski
Krzysztof Węcel
Adam Walczak

BIS 2009 Workshop Organization

ADW 2009 – Second Workshop on Advances in Accessing the Deep Web

Workshop Chairs

Flejter, Dominik	Poznan University of Economics, Poland
Kaczmarek, Tomasz	Poznan University of Economics, Poland
Kowalkiewicz, Marek	SAP Research Brisbane, Australia

Workshop Organizers

Poznan University of Economics, Department of Information Systems

Program Committee

Alvarez, Manuel	University of A Coruna, Spain
Bergamaschi, Sonia	University of Modena and Reggio Emilia, Italy
Bergman, Michael K.	Structured Dynamics LLC, USA
Celino, Irene	CEFRIEL - Politecnico di Milano, Italy
Della Valle, Emanuele	CEFRIEL - Politecnico di Milano, Italy
Guerra, Francesco	University of Modena and Reggio Emilia, Italy
Hornung, Thomas	Albert-Ludwigs-Universität Freiburg, Germany
Lucchese, Claudio	ISTI - CNR, Italy
Maurino, Andrea	Università di Milano Bicocca, Italy
Raposo Santiago, Juan	University of A Coruna, Spain
Shestakov, Denis	University of Turku, Finland
Soares da Silva, Altigran	Universidade Federal do Amazonas, Brazil
Wu, Zonghuan	University of Louisiana, Lafayette, USA

AKTB 2009 – First Workshop on Applications of Knowledge-Based Technologies in Business

Workshop Chairs

Kriksciuniene, Dalia	Vilnius University, Lithuania
Sakalauskas, Virgilijus	Vilnius University, Lithuania

Workshop Organizers

Vilnius University, Kaunas Faculty, Department of Informatics

Program Committee

Schoop, Eric	Dresden University of Technology, Germany
Gudas, Saulius	Vilnius University, Lithuania
Krebs, Irene	Potsdam University, Germany
Simutis, Rimvydas	Vilnius University, Lithuania
Meinberg, Uwe	Cottbus Technical University, Germany
Kiss, Ferenc	Foundation for Information Society, Hungary
Xavier, Adilson Elias	COPPE- Federal University of Rio de Janeiro, Brazil
Sakalauskas, Leonidas	Institute of Mathematics and Informatics, Lithuania
Bassa, Lia	Foundation for Information Society, Hungary
Augutis, Juozas	Vytautas Magnus University, Lithuania
Gustas, Remigijus	Karlstad University, Sweden
Lee, Soo-Young	KAIST, Korea
Yin, Hujun	University of Manchester, UK
Jordaan, Jaco	Tshwane University of Technology, South Africa
Zhang, Du	California State University, USA
Corchado, Emilio	University of Burgos, Spain
Shi, Yong	Graduate University of the Chinese Academy of Sciences, China
Tan, Margaret	Nanyang Technological University, Singapore
Mazeika, Arturas	Free University of Bozen-Bolzano, Italy
Bauerle, Peter	IBM Deutschland, Germany
Bergeron, Francois	Universite Laval, Canada

ECONOM 2009 – First Workshop on the Economics of Knowledge-Based Technologies

Workshop Chairs

Simperl, Elena	STI Innsbruck, Austria
Buerger, Tobias	STI Innsbruck, Austria
Filipowska, Agata	Poznan Univeristy of Economics, Poland
Tempich, Christoph	Detecon International GmbH, Germany

Workshop Organizers

Semantic Technology Institute (STI) Innsbruck, University of Innsbruck, Austria
Poznan University of Economics, Department of Information Systems, Poland
Detecon International GmbH, Germany

Program Committee

King, Nicholas	British Telecom, UK
Gomez-Perez, Jose-Manuel	ISOCO, Spain
Tolksdorf, Robert	Free University Berlin, Germany
del Carmen Suarez-Figueroa, Maria	UPM, Spain
Kowalkiewicz, Marek	SAP Research, Australia
Lopez, Ozelin	XimetriX Network Thoughts, Spain
Zapletal, Marco	Vienna University of Technology, Austria
Sure, York	SAP, Germany
Giernalczyk, Astrid	Research Institute for Operations Management at RWTH Aachen University, Germany
Cuel, Roberta	University of Trento, Italy
Maier, Ronald	University of Innsbruck, Austria
Strasunskas, Darijus	Norwegian University of Science and Technology, Norway

EeLT 2009 – Second Workshop on Emerging eLearning Web Technologies

Workshop Chairs

Grzonkowski, Slawomir	DERI, NUI Galway, Ireland
Nagle, Tadhg	UCC Cork, Ireland

Workshop Organizers

Digital Enterprise Research Institute, National University of Ireland, Galway

Program Committee

Dagger, Declan	Trinity College, Ireland
McDaniel, Bill	DERI, NUI Galway, Ireland
Westerski, Adam	Universidad Politecnica de Madrid, Spain
Luca De Coi, Juri	L3S and University of Hannover, Germany

Stankiewicz, Katarzyna	Gdansk University of Technology, Poland
van Damme, Celine	Vrije Universiteit Brussel, Belgium
Mukundanunny, Shyam Diwakar	University of Pavia, Italy
Simon, Bernd	Vienna University of Economics and Business Administration, Austria

First International Enterprise X.0 Workshop: From Web 2.0 in Enterprises Towards a Corporate Web X.0

Workshop Chairs

Mochol, Malgorzata	Free University of Berlin, Germany
Nixon, Lyndon JB	STI International Wien, Austria
Luczak-Roesch, Markus	Free University of Berlin, Germany

Program Committee

Berrueta, Diego	CTIC, Spain
Bussler, Christoph	Merced Systems, Inc., USA
Erling, Orri	OpenLink Software, UK
de Francisco, David	Telefonica, Spain
Filipowska, Agata	Poznan University of Economics, Poland
Grobelnik, Marko	Josef Stefan Institute, Slovenia
Groza, Tudor	DERI Galway, Ireland
Hoppe, Thomas	Ontonym GmbH, Germany
Lassila, Ora	Nokia, USA
Leger, Alain	France Telecom, France
Lucose, Dickson	MIMOS, Malaysia
Nowack, Benjamin	semsol web semantics, Germany
Pan, Jeff	University of Aberdeen, UK
Pellegrini, Tassilo	Semantic Web Company, Austria
Schmidt, Kai-Uwe	SAP, Germany
Shvaiko, Pavel	TasLab, Informatica Trentina, Italy
Svatek, Vojtech	University of Economics Prague, Czech Republic
Tolksdorf, Robert	Free University Berlin, Germany
Werthner, Hannes	Technical University of Vienna, Austria

ESHE 2009 – First Workshop on Enterprise Systems in Higher Education

Workshop Chairs

Gomez, Jorge Marx	University Oldenburg, Germany
Haak, Liane	University Oldenburg, Germany
Peters, Dirk	University Oldenburg, Germany

Workshop Organizers

Carl von Ossietzky University Oldenburg, Department of Computer Science, Chair of Business Informatics I / Very Large Business Applications

Program Committee

Cruz-Cunha, Maria Manuela	Polytechnic Institute of Cåvado and Ave School of Technology, Portugal
Davcev, Danco	Sts. Cyril and Methodius University, F.Y.R.O. Macedonia
Juell-Skielse, Gustaf	Royal Institute of Technology, Sweden
Kolokytha, Elpida	Aristotle University of Thessaloniki, Greece
Magnusson, Johan	University of Gothenburg, Sweden
Mylopoulos, Yannis	Aristotle University of Thessaloniki, Greece
Varajao, Joao Eduardo	University of Trås-os-Montes e Alto Douro, Portugal

LIT 2009 – Second Workshop on Legal Informatics and Legal Information Technology

Workshop Chairs

Stolarski, Piotr	Poznan University of Economics, Poland
Tomaszewski, Tadeusz	Poznan University of Economics, Poland

Workshop Organizers

Poznan University of Economics, Department of Information Systems

Program Committee

Bellucci, Emilia	Victoria University, Australia
D'Agostini Bueno, Tania Cristina	Presidente da Diretoria Executiva IJURIS, Brazil
Hoeschl, Hugo	IJURIS Research Group, Brazil
Lodder, Arno	Free University of Amsterdam, The Netherlands
Morek, Rafal	University of Warsaw, Poland
Schweighofer, Erich	University of Vienna, Austria, Austria
Stranieri, Andrew	University of Ballarat, Australia

SAW 2009 – Third Workshop on Social Aspects of the Web

Workshop Chairs

Flejter, Dominik	Poznan University of Economics, Poland
Kaczmarek, Tomasz	Poznan University of Economics, Poland
Kowalkiewicz, Marek	SAP Research Brisbane, Australia

Workshop Organizers

Poznan University of Economics, Department of Information Systems

Program Committee

Balog, Krisztian	University of Amsterdam, The Netherlands
Braun, Simone	FZI Karlsruhe, Germany
Breslin, John	DERI, NUI Galway, Ireland
Coenen, Tanguy	Vrije Universiteit Brussel, Belgium
Dietzold, Sebastian	University of Leipzig, Germany
Eynard, Davide	Politecnico di Milano, Italy
Jatowt, Adam	Kyoto University, Japan
Paprzycki, Marcin	Polish Academy of Science, Poland
Picard, Willy	Poznan University of Economics, Poland
Siorpaes, Katharina	STI, University of Innsbruck, Austria
Tang, Jie	Tshingua University, China
van Damme, Celine	Vrije Universiteit Brussel, Belgium
Zacharias, Valentin	FZI Karlsruhe, Germany

SDS-SOA 2009 – First Workshop on Service Discovery and Selection in SOA Ecosystems

Workshop Chairs

Haniewicz, Konstanty	Poznan University of Economics, Poland
Kaczmarek, Monika	Poznan University of Economics, Poland
Zaremba, Maciej	DERI Galway, Ireland
Zyskowski, Dominik	Poznan University of Economics, Poland

Workshop Organizers

Poznan University of Economics, Department of Information Systems

Program Committee

Kokash, Natalia	CWI Amsterdam, The Netherlands
Kuster, Ulrich	Friedrich Schiller University of Jena, Germany
Ludwig, Andre	University of Leipzig, Germany
Markovic, Ivan	SAP Research Centre, Karlsruhe
Palma, Raul	Universidad Politecnica de Madrid, Spain
Vasiliu, Laurentiu	DERI Galway, Ireland
Vitvar, Tomas	STI2 Innsbruck, Austria
Zaharia, Raluca	DERI Galway, Ireland

Table of Contents

BIS 2009 Keynote Speech

ADW Workshop

AKTB Workshop

ESHE Workshop

LIT Workshop

SAW Workshop

SDS-SOA Workshop

Using Process Mining to Generate Accurate and Interactive Business Process Maps

W.M.P. van der Aalst

Eindhoven University of Technology
P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands
w.m.p.v.d.aalst@tue.nl

Abstract. The quality of today's digital maps is very high. This allows for new functionality as illustrated by modern car navigation systems (e.g., TomTom, Garmin, etc.), Google maps, Google Street View, Mashups using geo-tagging (e.g., Panoramio, HousingMaps, etc.), etc. People can seamlessly zoom in and out using the interactive maps in such systems. Moreover, all kinds of information can be projected on these interactive maps (e.g., traffic jams, four-bedroom apartments for sale, etc.). Process models can be seen as the "maps" describing the operational processes of organizations. Unfortunately, accurate and interactive process maps are typically missing when it comes to business process management. Either there are no good maps or the maps are static or outdated. Therefore, we propose to *automatically generate business process maps using process mining techniques*. By doing this, there is a close connection between these maps and the actual behavior recorded in event logs. This will allow for high-quality process models showing what really happened. Moreover, this will also allow for the *projection of dynamic information*, e.g., the "traffic jams" in business processes. In fact, the combination of accurate maps, historic information, and information about current process instances, allows for *prediction* and *recommendation*. For example, just like TomTom can predict the arrival time at a particular location, process mining techniques can be used to predict when a process instance will finish.

1 The Need for Accurate and Interactive Business Process Maps

Process models are vital for the design, analysis, and implementation of information systems. Their role is similar to the role of maps for navigation systems, mashups, etc. For example, people increasingly rely on the devices of TomTom and other vendors and find it useful to get directions to go from A to B, know the expected arrival time, learn about traffic jams on the planned route, and be able to view maps that can be customized in various ways (zoom-in/zoom-out, show fuel stations, speed limits, etc.). Maps do not only play an important role in car navigation, but are also crucial for all kinds of innovative information services. Figure 1 shows two examples combining cartographic information

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 1–14, 2009.

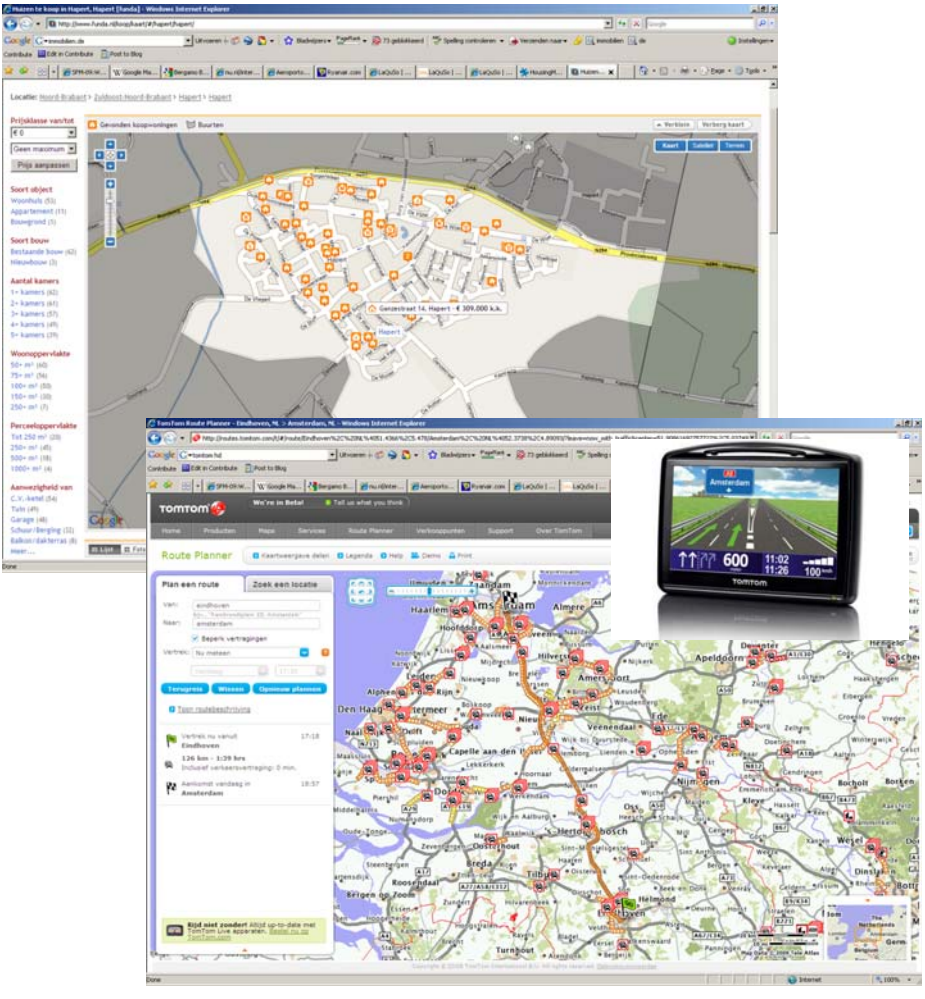


Fig. 1. The role of maps in Funda (top left) and TomTom HD Traffic (bottom right). Funda dynamically shows houses for sale in a particular area (in this case town of Hapert) meeting specific criteria (cf. `www.funda.nl`). TomTom HD Traffic is calculating the best route based on cell phone information provided by Vodafone, i.e., the locations and directions of cell phones are used to predict traffic jams (cf. `www.tomtom.com`). Both examples use a combination of high-quality maps augmented with dynamic information allowing for seamlessly zooming in and out. This paper advocates the development of such functionality for business information systems.

with dynamically changing data. However, when looking at business processes, such information is typically lacking. Good and accurate "maps" of business processes are often missing and, if they exist, they tend to be restrictive and provide little information. For example, very few information systems are able to predict *when* a case will complete. Therefore, we advocate more TomTom-like

functionality for business process management, coined "TomTom4BPM" in [2]. Besides navigation systems, there are many applications based on Google maps. For example, real-estate agencies dynamically projecting information on maps, etc. A key element is the availability of high-quality maps. The early navigation systems were using very course maps that were often outdated, thus limiting their applicability. A similar situation can be seen when looking at information systems based on incorrect or outdated process models.

In this paper, we advocate the use of *accurate and interactive business process maps obtained through process mining*. The goal is to provide a better breed of *Business Process Management Systems* (BPMSs) [1,15,29]. BPMSs are used to manage and execute operational processes involving people, applications, and/or information sources on the basis of process models. These systems can be seen as the next generation of workflow technology offering more support for analysis. Despite significant advances in the last decade, the functionality of today's BPMSs leaves much to be desired. This becomes evident when comparing such systems with the latest car navigation systems of TomTom or innovative applications based on Google maps. Some examples of functionality provided by TomTom and/or Google maps that are generally missing in contemporary BPMSs are:

- In today's organizations often *a good process map is missing*. Process models are not present, incorrect, or outdated. Sometimes process models are used to directly configure the BPMS. However, in most situations there is not an explicit process model as the process is fragmented and hidden inside legacy code, the configuration of ERP systems, and in the minds of people.
- If process models exist in an explicit form, *their quality typically leaves much to be desired*. Especially when a process model is not used for enactment and is only used for documentation and communication, it tends to present a "PowerPoint reality". Road maps are typically of much higher quality and use intuitive colors and shapes of varying sizes, e.g., highways are emphasized by thick colorful lines and dirt roads are not shown or shown using thin dark lines. In process models, *all activities tend to have the same size and color and it is difficult to distinguish the main process flow from the less traveled process paths*.
- Most process modeling languages have a static decomposition mechanism (e.g., nested subprocesses). However, what is needed are controls allowing users *to zoom in or zoom out seamlessly like in a navigation system or Google maps*. Note that, while zooming out, insignificant things are either left out or dynamically clustered into aggregate shapes (e.g., streets and suburbs amalgamate into cities). Process models should not be static but allow for various (context dependent) views.
- Sometimes process models are used for enactment. However, such "process maps" are often trying to "control" the user. When using a car navigation system, the driver is always in control, i.e., the road map (or TomTom) is not trying to "control" the user. The goal of a BPMS should be to *provide directions and guidance rather than enforcing a particular route*.
- A navigation system continuously shows a clear *overview of the current situation* (i.e., location and speed). Moreover, traffic information is given,

showing potential problems and delays. This information is typically missing in a BPMS. Even if the BPMS provides a management dashboard, TomTom-like features such as traffic information and current location are typically not shown in an intuitive manner.
– A TomTom system *continuously recalculates* the route, i.e., the recommended route is not fixed and changed based on the actions of the driver and contextual information (e.g. traffic jams). Moreover, at any point in time the navigation system is showing the *estimated arrival time*. Existing BPMSs are not showing this information and do not recalculate the optimal process based on new information.

The above list of examples illustrates desirable functionality that is currently missing in commercial BPMSs. Fortunately, recent breakthroughs in *process mining* may assist in realizing highly innovative features that are based on high-quality business process maps tightly connected to historic information collected in the form of event logs.

In the remainder of this paper, we first briefly introduce the concept process mining in Section 2. Section 3 introduces the ProM framework that aims at the generation of accurate and interactive business process maps obtained through process mining. Based on ProM and process mining it is possible to provide TomTom-like functionality as discussed in Section 4. One particular example of such innovative functionality is "case prediction" as described in Section 5. Pointers to related work on process mining are given in Section 6. Section 7 concludes the paper.

2 Process Mining

Process mining techniques attempt to extract non-trivial and useful information from *event logs* [5,9]. Many of today's information systems are recording an abundance of events in such logs. Various process mining approaches make it possible to uncover information about the processes they support. Typically, these approaches assume that it is possible to sequentially record events such that each event refers to an activity (i.e., a well-defined step in the process) and is related to a particular case (i.e., a process instance). Furthermore, some mining techniques use additional information such as the performer or originator of the event (i.e., the person/resource executing or initiating the activity), the timestamp of the event, or data elements recorded with the event (e.g., the size of an order).

Process mining addresses the problem that most people have very limited information about what is actually happening in their organization. In practice, there is often a significant gap between what is prescribed or supposed to happen, and what *actually* happens. Only a concise assessment of the organizational reality, which process mining strives to deliver, can help in verifying process models, and ultimately be used in a process redesign effort or BPMS implementation.

Some examples of questions addressed by process mining:

– *Process discovery*: "What is really happening?"
– *Conformance checking*:"Do we do what was agreed upon?"

– *Performance analysis*: "Where are the bottlenecks?"
– *Process prediction*: "Will this case be late?"
– *Process improvement*: "How to redesign this process?"

The above questions show that process mining is not limited to control-flow discovery. In fact, we identify three types of process mining: (a) *discovery*, (b) *conformance*, and (c) *extension*. We also distinguish three different perspectives: (a) the *control-flow perspective* ("How?"), (b) the *organizational perspective* ("Who?") and (c) the *case perspective* ("What?").

Figure 2 positions process mining as the technology that "sits" in-between event logs and process models. The figure also shows the three types of process mining.

The first type of process mining is *discovery*, i.e., deriving information from some event log without using an a priori model. Based on an event log various types of models may be discovered, e.g., process models, business rules, organizational models, etc.

The second type of process mining is *conformance checking*. Here the event log is used to check if reality conforms to some model. For example, there may be a process model indicating that purchase orders of more than one million Euro require two checks, while in reality this does not happen. Conformance checking may be used to detect deviations, to locate and explain these deviations, and to measure the severity of these deviations.

The third type of process mining, called *extension*, also assumes both a log and a model as input (cf. Figure 2). However, the model is not checked for correctness, instead it is used as a basis, i.e., the model is augmented with some new information or insights. For example, an existing process model could be extended by timing information, correlations, decision rules, etc.

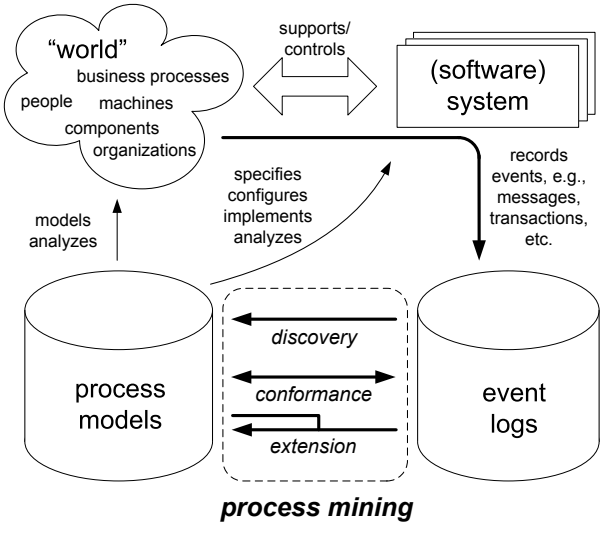


Fig. 2. Process mining as a bridge between process models and event logs

Orthogonal to the three types of mining, there are the three perspectives mentioned before. The *control-flow perspective* focuses on the control-flow, i.e., the ordering of activities. The goal of mining this perspective is to find a good characterization of all possible paths, e.g., expressed in terms of a Petri net or some other notation (e.g., EPCs, BPMN, UML ADs, etc.). The *organizational perspective* focuses on information about resources hidden in the log, i.e., which performers are involved and how are they related. The goal is to either structure the organization by classifying people in terms of roles and organizational units or to show the social network. The *case perspective* focuses on properties of cases. Cases can be characterized by their path in the process or by the originators working on a case. However, cases can also be characterized by the values of the corresponding data elements. For example, if a case represents a replenishment order, it may be interesting to know the supplier or the number of products ordered.

3 Tool Support: ProM

The ProM framework aims to *cover the full process mining spectrum* shown in Figure 2. The current version of ProM provides more than 250 plug-ins. The ProM framework has been developed as a completely plug-able environment and serves as an excellent basis for process mining initiatives.

ProM is the only comprehensive framework supporting a wide range of process mining techniques. Most other tools in this area only focus on a single perspective and/or technique. *Futura Reflect* by Futura Process Intelligence, *BPM|one* by Pallas Athena, *Comprehend* by Open Connect, *Interstage Automated Business Process Discovery and Visualization* by Fujitsu, *Process Discovery Focus* by Iontas, and *Enterprise Visualization Suite* by BusinesScape are some examples of commercial tools that offer some form of process discovery. Of these tools Futura Reflect and BPM|one are more mature as they allow for the discovery of processes with concurrency. Most of the other tools mentioned are only able to discover sequential processes or even require a-priori modeling. Commercial tools typically offer only a small subset of the functionality provided by ProM. However, the emergence of these tools illustrates the practical interest in process mining. For example, Futura Process Intelligence and Pallas Athena have been selected as "Cool Vendor 2009" by Gartner because of their process mining capabilities. Both tools use genetic process mining algorithms developed in the context of ProM [19].

The reader is referred to `www.processmining.org` to learn more about process mining and to download ProM.

4 TomTom4BPM

In [2], the term *TomTom4BPM* was coined to stress the need for the map-based functionality one can find in navigation systems (e.g., TomTom, Garmin, VDO Dayton, Mio, Magellan, etc.), Google maps, Google Street View, Mashups using geo-tagging (e.g., Panoramio, HousingMaps, FindByClick, etc.). After

introducing process mining, we revisit the desired functionalities mentioned in Section 1. Here we are particularly interested in adding innovative functionality to BPMSs.

- As indicated earlier, *good process maps are typically missing* in today's organizations. Clearly, process mining can assist here. Process discovery algorithms [6,9,10,11,13,16,27,28] are able to extract process maps from event logs. These maps are describing the way things really happened rather than providing some subjective view.
- In Section 1, we indicated that even if process models exist in an explicit form, *their quality typically leaves much to be desired*. Using process mining techniques, one can avoid depicting a "PowerPoint reality" and come closer to the quality of road maps. Moreover, based on historic information, it is possible use intuitive visual metaphors adopted from road maps. For example, we can use intuitive colors and shapes of varying sizes, e.g., the "highways in the process" are emphasized by thick colorful lines and "process dirt roads" are not shown or shown using thin dark lines. The major "cities of a process" can also be emphasized and less relevant activities can be removed. Relevance can be determined based on actual frequencies of activities in logs. Other metrics may be the time spent on activities or the costs associated with them. ProM's *Fuzzy Miner* [16] can discover processes from event logs and offers such visualizations.
- Most process modeling languages have a static decomposition mechanism (e.g., nested subprocesses) without the ability to *seamlessly zoom in or zoom out like in a navigation system or Google maps*. ProM's *Fuzzy Miner* [16] allows for such a seamless zoom. Note that, while zooming out, insignificant activities and paths are either left out or dynamically clustered into aggregate shapes (e.g., streets and suburbs amalgamate into cities).
- When "process maps" are used in an operational sense, they typically *attempt to control the users*. However, when using a car navigation system, the driver is always in control, i.e., the road map (or TomTom) is not trying to "control" the user. The goal of an information system should be to provide directions and guidance rather than enforcing a particular route. ProM's *Recommendation Engine* [23] learns from historic data and uses this to provide recommendations to the user. This way the workflow system can provide more flexibility while still supporting the user. This is comparable to the directions given by a navigation system.
- A navigation system continuously shows a clear overview of the current situation (i.e., location and speed). Moreover, *traffic information* is given, showing potential problems and delays. Since process mining results in a tight connection between events and maps, it is easy to project dynamic information on process maps. Ideas such as the ones presented Figure 1 have their counterparts in BPMSs, e.g., showing "traffic jams" in business processes.
- At any point in time the navigation system is showing the *estimated arrival time*. Existing BPMSs are not showing this information and do not recalculate the optimal process based on new information. ProM provides several

so-called *prediction engines* [7,14] to estimate the remaining flow time of a case. The next section shows an example of an application of the technique described in [7].

In this paper, we cannot present the various techniques supported by ProM in detail. Instead, we only show that event logs can be used to predict the remaining time until completion for running cases.

5 An Example: Case Prediction

As an illustration of the innovative features that can be provided by combining accurate process maps and historic event information, we briefly show an application of *case prediction* [7]. To illustrate the technique presented in [7] and implemented in ProM, we use an event log of municipality taken from a process that deals with objections (i.e., appeals) against the real-estate property valuation or the real-estate property tax. The municipality is using eiStream workflow (formerly known as Eastman Software and today named Global 360) to handle these objections.

The process considered in this case study is called "Bezwaar WOZ", where WOZ ("Waardering Onroerende Zaken") refers to the particular law describing regulations related to real-estate property valuation by municipalities. We used an event log with data on 1882 objections handled by the municipality. The log contains 11985 events and the average total flow time is 107 days while some cases take more than 200 days. Figure 3 shows the distribution of total flow times. The x-axis shows the 1882 cases and the y-axis shows the duration in days. Note that some cases take a very short time while others take much longer,

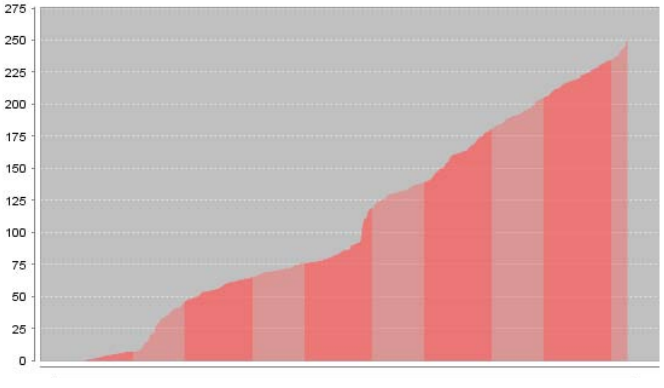


Fig. 3. The distribution of the total flow time of cases extracted using ProM. The x-axis represents the 1882 cases sorted by flow time. The y-axis shows durations in days. Note that some cases almost take 250 days.

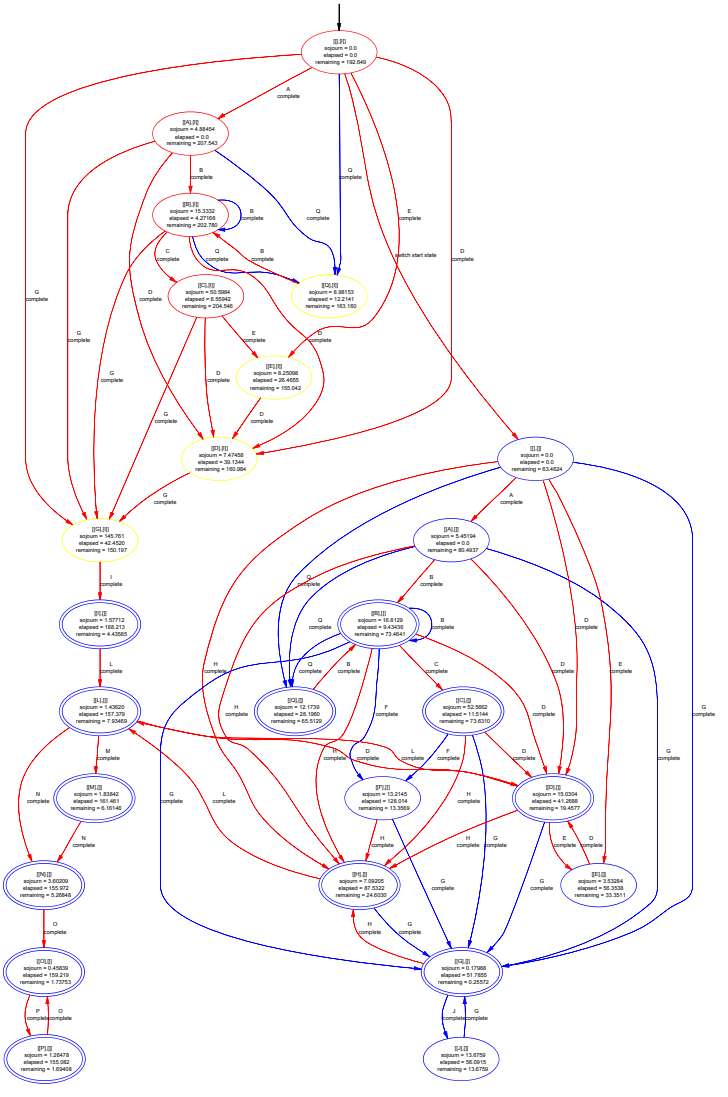


Fig. 4. An annotated transition system extracted from event log L_1. The transition system and its annotations are not intended to be readable and the activity names have been obfuscated. The transition system is learned from an event log containing information about 982 cases (objections against the real-estate property valuation/tax). Per state, historic information is used to make a prediction. For example, for the top state the predicted time until completion is 192 days, for the bottom-left state the predicted time until completion is 1.69 days, and for the bottom-right state the predicted time until completion is 13.67 days. The Mean Average Error (MAE) is 17.129 days when this annotated transition system is evaluated using another log (L_2) containing event data on 900 other objections.

thus making it difficult to predict the remaining time for cases in the system. To measure the quality of predictions, we split the log into a training set (log L_1) and a test set (log L_2). Log L_1 contains 982 cases and log L_2 contains 900 cases.

The goal is to predict, at any point in time, the remaining processing time of a case. This corresponds to the "estimated arrival time" provided by car navigation systems like TomTom. To do this, we build a so-called *annotated transition system* using the training set (log L_1). Using a variable abstraction mechanism, partial traces are mapped onto states of the transition system. Using historic information, appropriate statistics are collected per state.

Figure 4 shows an annotated transition system obtained using a particular abstraction (see [7] for other abstractions). If one is interested in the remaining time until completion of a particular case c, then the partial trace σ_c of this case is mapped onto a state s_c. Based on s_c a lookup is done in Figure 4 resulting in a prediction t_c^p, e.g., for a case where two particular steps have been executed, the predicted remaining time until completion is $t_c^p = 20.5$ days. Afterwards, it is possible to measure what the actual quality of this estimate. For example, if the real remaining time until completion turns out to be $t_c^r = 25.7$, then the error is $|t_c^p - t_c^r| = 5.2$ days.

If we use the annotated transition system shown in Figure 3 (which was derived from L_1) to predict the remaining time until completion before/after every event in L_2, then the Mean Average Error (MAE) is 17.129 days. Given the fact that there are huge variations in flow times and that the average flow time is 107 days (cf. Figure 3), this is a spectacular performance. For processes which less variation, it is possible to make even better predictions. To put the MAE of 17.129 days into perspective, it is interesting to compare the performance of the annotated transition system shown in Figure 3 with the simple heuristic of always estimating half of average total flow time (i.e., 53.5 days). The MAE of this heuristic is 61.750 days. Hence, the performance of the technique presented in [7] is much better than simple heuristics. It is quite remarkable that one can predict the remaining time until completion so accurately. This shows that using process mining techniques one can realize TomTom-like functionality like the estimated arrival time.

6 Related Work

Since the mid-nineties several groups have been working on techniques for process mining [9,6,10,11,13,16,27,28], i.e., discovering process models based on observed events. In [8] an overview is given of the early work in this domain. The idea to apply process mining in the context of workflow management systems was introduced in [10]. In parallel, Datta [13] looked at the discovery of business process models. Cook et al. investigated similar issues in the context of software engineering processes [11]. Herbst [17] was one of the first to tackle more complicated processes, e.g., processes containing duplicate tasks. Most of the classical approaches have problems dealing with concurrency. The α-algorithm [9]

was the first technique taking concurrency as a starting point. However, this simple algorithm has problems dealing with complicated routing constructs and noise (like most of the other approaches described in literature). In the context of the ProM framework [3] more robust techniques have been developed. The heuristics miner [27] and the fuzzy miner [16] can deal with incomplete, unbalanced, and/or noisy events logs. The two-phase approach presented in [6] allows for various abstractions to obtain more useful models. It is impossible to give a complete review of process mining techniques here, see `www.processmining.org` for more pointers to literature.

The approaches mentioned above focus on control-flow discovery. However, when event logs contain time information, the discovered models can be extended with timing information. For example, in [25] it is shown how timed automata can be derived. In [20] it is shown how any Petri net discovered by ProM can be enriched with timing and resource information.

The above approaches all focus on discovering process models based on historic information and do not support users at run-time. The recommendation service of ProM learns based on historic information and uses this to guide the user in selecting the next work-item [23]. This is related to the use of case-based reasoning in workflow systems [26]. In the context of ProM two prediction approaches are supported: [7] and [14]. The prediction service presented in [14,12] predicts the completion time of cases by using non-parametric regression. The prediction service presented in [7] (used in Section 5) is based on annotated transition systems and uses the abstractions defined in [6]. Also related is the prediction engine of Staffware [24,22] which is using simulation to complete audit trails with expected information about future events. This particular approach is rather unreliable since it is based on one run through the system using a copy of the actual engine. Hence, no probabilities are taken into account and there is no means of "learning" to make better predictions over time. A more refined approach focusing on the transient behavior (called "short-term simulation") is presented in [21].

The limitations related to the representation and visualization of process models mentioned at the beginning of this paper became evident based on experiences gathered in many process mining projects. It seems that the "map metaphor" can be used to present process models and process information in completely new ways [16,18]. In the context of YAWL [4,18], we showed that it is possible to show current work items on top of various maps. Work items can be shown on top of a geographic map, a process model, a time chart, an organizational model, etc. In the context of ProM, we have used the "map metaphor" to enhance the so-called Fuzzy Miner [16]. As presented in [16], four ideas are being combined in ProM's Fuzzy Miner to draw maps of process models.

– *Aggregation:* To limit the number of information items displayed, maps often show coherent clusters of low-level detail information in an aggregated manner. One example are cities in road maps, where particular houses and streets are combined within the city's transitive closure.

- *Abstraction:* Lower-level information which is insignificant in the chosen context is simply omitted from the visualization. Examples are bicycle paths, which are of no interest in a motorists map.
- *Emphasis:* More significant information is highlighted by visual means such as color, contrast, saturation, and size. For example, maps emphasize more important roads by displaying them as thicker, more colorful and contrasting lines (e.g., motorways).
- *Customization:* There is no one single map for the world. Maps are specialized on a defined local context, have a specific level of detail (city maps vs highway maps), and a dedicated purpose (interregional travel vs alpine hiking).

7 Conclusion

The paper suggests *using process mining to create accurate and interactive business process maps* for the management of business processes. The maps can be accurate because they are no longer based on outdated or subjective information, but on facts recorded in event logs. By establishing a close connection between event logs and such maps, it is possible to project information dynamically and let the user interact with such business process maps. Using ProM some of the desired TomTom functionality has been realized and there is a huge innovation potential for today's BPMSs. Using "TomTom4BPM" we can realize truly intelligent information systems.

To make things a bit more concrete, we presented a particular example of such functionality using a new method for predicting the "future of a running instance". Given a running case, our prediction approach allows answering questions like "When will this case be finished?", "How long does it take before activity A is completed?", "How likely is it that activity B will be performed in the next two days?", etc. This corresponds to the functionality we know from modern car navigation systems that give an estimate for the remaining driving time.

Essentially for all of this is that we have high-quality business process maps. Unfortunately, the quality of today's process models leaves much to be desired and the situation is comparable to cartographic information decades ago. Problems with the first navigation systems showed that incorrect maps result in systems that are not very usable. Therefore, the ability to extract maps from event logs using process mining is crucial.

Some people may argue that business processes are less stable than infrastructures consisting of roads, intersections, and bridges. Therefore, it is much more difficult to provide accurate business process maps. This is indeed the case. However, this illustrates that a continuous effort is required to keep business process maps up to date. Process mining can be used for this. Moreover, by recording and analyzing event logs on-the-fly, it is possible to offer more flexibility without loosing sight of the actual processes. Therefore, the need to enforce rigid processes is removed and, like in the context of a car navigation system, the "driver is in control" rather than some archaic information system.

Acknowledgements

This research is supported by EIT, NWO-EW, SUPER, and the Technology Foundation STW. We would like to thank the many people involved in the development of ProM (see references).

References

1. van der Aalst, W.M.P.: Business Process Management Demystified: A Tutorial on Models, Systems and Standards for Workflow Management. In: Desel, J., Reisig, W., Rozenberg, G. (eds.) Lectures on Concurrency and Petri Nets. LNCS, vol. 3098, pp. 1–65. Springer, Heidelberg (2004)
2. van der Aalst, W.M.P.: TomTom for Business Process Management (TomTom4BPM). In: Gordijn, J. (ed.) Proceedings of the 21st International Conference on Advanced Information Systems Engineering (CAiSE 2009). LNCS. Springer, Heidelberg (2009)
3. van der Aalst, W.M.P., van Dongen, B.F., Günther, C.W., Mans, R.S., Alves de Medeiros, A.K., Rozinat, A., Rubin, V., Song, M., Verbeek, H.M.W., Weijters, A.J.M.M.: ProM 4.0: Comprehensive Support for Real Process Analysis. In: Kleijn, J., Yakovlev, A. (eds.) ICATPN 2007. LNCS, vol. 4546, pp. 484–494. Springer, Heidelberg (2007)
4. van der Aalst, W.M.P., ter Hofstede, A.H.M.: YAWL: Yet Another Workflow Language. Information Systems 30(4), 245–275 (2005)
5. van der Aalst, W.M.P., Reijers, H.A., Weijters, A.J.M.M., van Dongen, B.F., Alves de Medeiros, A.K., Song, M., Verbeek, H.M.W.: Business Process Mining: An Industrial Application. Information Systems 32(5), 713–732 (2007)
6. van der Aalst, W.M.P., Rubin, V., van Dongen, B.F., Kindler, E., Günther, C.W.: Process Mining: A Two-Step Approach to Balance Between Underfitting and Overfitting. Software and Systems Modeling (2009)
7. van der Aalst, W.M.P., Schonenberg, M.H., Song, M.: Time Prediction Based on Process Mining. BPM Center Report BPM-09-04, BPMcenter.org (2009)
8. 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 and Knowledge Engineering 47(2), 237–267 (2003)
9. van der Aalst, W.M.P., Weijters, A.J.M.M., Maruster, L.: Workflow Mining: Discovering Process Models from Event Logs. IEEE Transactions on Knowledge and Data Engineering 16(9), 1128–1142 (2004)
10. Agrawal, R., Gunopulos, D., Leymann, F.: Mining Process Models from Workflow Logs. In: Sixth International Conference on Extending Database Technology, pp. 469–483 (1998)
11. Cook, J.E., Wolf, A.L.: Discovering Models of Software Processes from Event-Based Data. ACM Transactions on Software Engineering and Methodology 7(3), 215–249 (1998)
12. Crooy, R.: Predictions in Information Systems: A process mining perspective. Master's thesis, Eindhoven University of Technology, Eindhoven (2008)
13. Datta, A.: Automating the Discovery of As-Is Business Process Models: Probabilistic and Algorithmic Approaches. Information Systems Research 9(3), 275–301 (1998)

14. van Dongen, B.F., Crooy, R.A., van der Aalst, W.M.P.: Cycle Time Prediction: When Will This Case Finally Be Finished? In: Meersman, R., Tari, Z. (eds.) CoopIS 2008, OTM 2008, Part I. LNCS, vol. 5331, pp. 319–336. Springer, Heidelberg (2008)
15. Dumas, M., van der Aalst, W.M.P., ter Hofstede, A.H.M.: Process-Aware Information Systems: Bridging People and Software through Process Technology. Wiley & Sons, Chichester (2005)
16. Günther, C.W., van der Aalst, W.M.P.: Fuzzy Mining: Adaptive Process Simplification Based on Multi-perspective Metrics. In: Alonso, G., Dadam, P., Rosemann, M. (eds.) BPM 2007. LNCS, vol. 4714, pp. 328–343. Springer, Heidelberg (2007)
17. Herbst, J.: A Machine Learning Approach to Workflow Management. In: Lopez de Mantaras, R., Plaza, E. (eds.) ECML 2000. LNCS, vol. 1810, pp. 183–194. Springer, Heidelberg (2000)
18. de Leoni, M., van der Aalst, W.M.P., ter Hofstede, A.H.M.: Visual Support for Work Assignment in Process-Aware Information Systems. In: Dumas, M., Reichert, M., Shan, M.-C. (eds.) BPM 2008. LNCS, vol. 5240, pp. 67–83. Springer, Heidelberg (2008)
19. Alves de Medeiros, A.K., Weijters, A.J.M.M., van der Aalst, W.M.P.: Genetic Process Mining: An Experimental Evaluation. Data Mining and Knowledge Discovery 14(2), 245–304 (2007)
20. Rozinat, A., Mans, R.S., Song, M., van der Aalst, W.M.P.: Discovering Simulation Models. Information Systems 34(3), 305–327 (2009)
21. Rozinat, A., Wynn, M.T., van der Aalst, W.M.P., ter Hofstede, A.H.M., Fidge, C.: Workflow Simulation for Operational Decision Support Using Design, Historic and State Information. In: Dumas, M., Reichert, M., Shan, M.-C. (eds.) BPM 2008. LNCS, vol. 5240, pp. 196–211. Springer, Heidelberg (2008)
22. Schellekens, B.: Cycle Time Prediction in Staffware. Master's thesis, Eindhoven University of Technology, Eindhoven (2009)
23. Schonenberg, H., Weber, B., van Dongen, B.F., van der Aalst, W.M.P.: Supporting Flexible Processes Through Recommendations Based on History. In: Dumas, M., Reichert, M., Shan, M.-C. (eds.) BPM 2008. LNCS, vol. 5240, pp. 51–66. Springer, Heidelberg (2008)
24. Staffware. Staffware Process Suite Version 2 – White Paper. Staffware PLC, Maidenhead, UK (2003)
25. Verwer, S.E., de Weerdt, M.M., Witteveen, C.: Efficiently learning timed models from observations. In: Wehenkel, L., Geurts, P., Maree, R. (eds.) Benelearn, Benelearn, pp. 75–76. University of Liege (2008)
26. Weber, B., Wild, W., Breu, R.: CBRFlow: Enabling Adaptive Workflow Management Through Conversational Case-Based Reasoning. In: Funk, P., González Calero, P.A. (eds.) ECCBR 2004. LNCS, vol. 3155, pp. 434–448. Springer, Heidelberg (2004)
27. Weijters, A.J.M.M., van der Aalst, W.M.P.: Rediscovering Workflow Models from Event-Based Data using Little Thumb. Integrated Computer-Aided Engineering 10(2), 151–162 (2003)
28. van der Werf, J.M.E.M., van Dongen, B.F., Hurkens, C.A.J., Serebrenik, A.: Process Discovery using Integer Linear Programming. In: van Hee, K.M., Valk, R. (eds.) PETRI NETS 2008. LNCS, vol. 5062, pp. 368–387. Springer, Heidelberg (2008)
29. Weske, M.: Business Process Management: Concepts, Languages, Architectures. Springer, Berlin (2007)

ADW 2009 Chairs' Message

Dominik Flejter[1], Tomasz Kaczmarek[1], and Marek Kowalkiewicz[2]

[1] Poznan University of Economics
D.Flejter@kie.ue.poznan.pl, T.Kaczmarek@kie.ue.poznan.pl
[2] SAP Research Brisbane
marek.kowalkiewicz@sap.com

We are proud to present for the second time the proceedings of the Advances in Accessing Deep Web workshop. This issue is a collection of the papers presented during the workshop co-located with the Business Information Systems conference, held in Poznan, Poland on 27-29 of April 2009. We established the workshop last year because we identified the need for a publication and discussion forum for young researchers in the Deep Web field. Although mainstream papers in this area are often published on the top Web-related conferences (WWW, WISE, ICWE and others) we thought there is a need to bring together the researchers and stimulate collaboration, especially among young adepts in the field to discuss innovative solutions and ongoing work. A successful second edition of our workshop proves that this need has been fulfilled. The success would not be possible without work of our Program Committee which includes leading researchers in the field coming both from universities and industry. We would like to thank them for encouraging their peers and students to consider our workshop and helping us to organize and rise the scientific level of the workshop with their experience and knowledge. The Program Committee of the Workshop included 16 researchers from eight countries, specializing in different aspects of Deep Web issues. We are happy that we could work together with people that were the first to notice the Deep Web potential and are currently among the leading experts in accessing its resources.

In the last year we saw a growing awareness of the Deep Web issue not only in the research community, but also in the industry and public. Top researchers are working intensely for industry leaders (Goggle, Yahoo, Microsoft) to uncover this dormant resource and make it available to the public, potentially bringing the next breakthrough in the Web search field. At the same time we observe growing number of press releases covering the Deep Web challenges and ongoing work - this signifies both, publicly expressed need to access the Deep Web resources in their full potential and the advances in the field. This encourages us to continue with next editions of the workshop, which this year covered a vide range of research topics - practical as well as theoretical - which were expressed in the CfP[1] and spanned from modeling the Deep Web, empirical studies, practical issues of DW sources access, and applications of DW research, to data integration issues and semantics support for solving Deep Web issues.

[1] http://bis.kie.ae.poznan.pl/12th_bis/wscfp.php?ws=adw2009

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 15–16, 2009.

Three out of six submitted papers were accepted and presented during the workshop. The first paper authored by Markus Pfuhl und Paul Alpar, titled "Improving Database Retrieval on the Web through Query Relaxation" considered problem of posing queries against DW sources. Authors aim at easing the query formulation by using taxonomies for some of the attributes present in the source and later using the taxonomy to relax queries posed by the users. They conclude that their approach allows to keep the query interface simple, and yet deliver more powerful query capabilities and broader results.

The second paper, by Monika Starzecka and Adam Walczak – "Using Semantics to Personalize Access to Data Intensive Web Sources" fits into Semantic Web - Deep Web mixture in that it provides guidelines for personalizing access to Deep Web source by use of ontology to model source structure and relationships between attributes, and allowing user to specify his personal view on the ontology which is taken into account during query formulation.

In the last paper, titled "Deep Web Queries in a Semantic Web Environment", authored by Thomas Hornung and Wolfgang May, it was demonstrated how Semantic Web technologies could be used to lift the Deep Web sources to the level of databases with a precise schema and strong typing information and then to the level of Semantic Web applications. The work mainly considered modeling Deep Web sources using the MARS approach developed in the Semantic Web framework for annotating source's structure.

We would like to thank all the Authors for their contributions and discussion during the workshop and invite you to take part in next edition of ADW Workshop to be announced soon on our project website: `http://www.integror.net` >Events.

Improving Database Retrieval on the Web through Query Relaxation

Markus Pfuhl and Paul Alpar

Institut für Wirtschaftsinformatik, Philipps-Universität Marburg
Universitätsstraße 24, 35037 Marburg
alpar@wiwi.uni-marburg.de, markus@pfuhl-net.de

Abstract. Offering database content to unknown Web users creates two problems. First, users need to know about its existence. Second, once they know that it exists, they need to be able to retrieve it. We concentrate on the latter task. The same problem occurs also within an organization but there at least the skilled users can use powerful tools like SQL to find any content within the database. Web interfaces to databases are relatively simple and restricted. Even a skilled user could not define complicated queries due to their limitations. Therefore, especially databases that should be accessed via the Web should offer more "intelligence". We propose two features towards this goal. First, taxonomies should be built for selected attributes. Second, better query results should be offered by relaxing user queries based on the knowledge captured in the taxonomies. In this paper, we derive a method for query relaxation guided by the ideas of Bayesian inference. It helps to select the best attribute to relax the query in a retrieval step. The approach is applied to taxonomy-based attributes although it can be generalized to other types of attributes as well. The quality of the method is tested with data from an actual database offered on the Web.

1 Introduction

A big part of the Deep Web consists of structured data [1, 2, 3] that are usually stored in a relational database. If the database contains personal (transaction) data, like in a bank account, then the user is allowed to access his data only. He only needs to supply the right identification and password. If the data are for general use, like product information or news, then the user would like to retrieve many or all the records or documents that are relevant to his inquiry. This is where the difficulties start. Assuming the user knows about the existence of the database storing the information of interest, he has to communicate his needs to it. On the Web, users can access the data by filling out a Web form. Their entries in the Web form are usually transformed into a Structured Query Language (SQL) query which is then run against the database. Most of the Web forms let the user enter keywords. If he enters more than one then they are connected by a logical "and" by default. This quickly leads to a small and often to an empty answer set. It is very likely that some or all relevant documents are missed in this case. Even sophisticated database engines that handle stemming, synonyms etc. will probably miss some relevant information (see example below). Some Web interfaces also offer to connect the keywords with the logical "or." The answer set of an "or"-connected query may become very big. While it may contain all

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 17–27, 2009.

relevant records, it may be too big for the user to peruse it so that again some relevant records or documents could be missed. Of course, some interfaces offer more features but this tends to be too complicated for the majority of (casual) users. Research on search behavior of Web users, mostly using search engines, shows that they seldom use logical operators or advanced features [4, 5]. The average number of terms used in a search was two. "And" was the most often used logical operator.

We propose a method which does not require more input from the user than the keywords he usually provides but the method can relax his query using semantic knowledge in order to provide him with a bigger answer set. The semantic knowledge is expressed in form of taxonomies that need not be visible to the user. To illustrate the idea, let us assume that a company is looking for potential employees in a Web database of job seekers. They may look for a *controller* who has experience in the *banking* business. If such people are not available or rare, they may be willing to also consider controllers who worked in the *insurance* industry. A regular database engine would not return this record because i*nsurance* and *banking* are neither equal nor synonyms. However, they are both part of the financial industry and in that sense siblings. These are the type of relaxations that our method should allow for in order to increase small answer sets without over blowing them.

In the following, we assume that an "and"-query leads to an empty or very small result set. The query can be relaxed in one or more directions to yield a greater result set. We show that the general ideas of Bayesian Decision Theory can be used to choose the right attribute for the next relaxation step and get a greater set of hits without losing (much) precision.

2 Query Relaxation Based on Bayesian Theory

2.1 Origins of Query Relaxation

The question of query relaxation has been discussed in various research areas like Case-Based Reasoning or Cooperative Answering [6, 7, 8]. Related work is also done in the field of Query Expansion, which handles the process of including related terms in the original query, while query relaxation produce sub-queries to get better results [9]. Query Expansion could be distinguished into document-using methods, statistical methods and semantic methods [10]. Semantic methods often use ontologies to expand queries [11, 12] which are often found in structured domains. One step towards query-relaxation in structured domains was made by Shimazu et al. [13]. They introduce an Abstraction Hierarchy (AH)[1] for every attribute and also use a nearest-neighbor approach for retrieval. To relax a given query they generate a set of values neighboring the value specified by the user. Fig. 1 shows an example for operating systems were the user specified BSD4.2 which is an element in the first-order neighbor value set. BSD4.3 and BSD are elements of the second-order neighbor value set etc. In the next step all combinations of possible neighbor value sets for all attributes are assembled and the relaxed queries are formulated. The number of possible combinations depends on the number of attributes and the complexity (deepness) of the AHs. In the worst case one has to make an "and"-relation for every combination

[1] An AH represents values and not data types like a TAH.

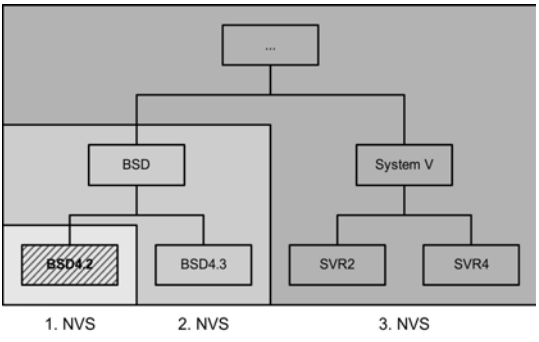


Fig. 1. Example of a Neighbor Value Set [13]

in one SQL-Query. Nevertheless, this approach is widely used in Case-Based-Reasoning-Systems as shown in [14].

An improvement of query relaxation for attributes based on numeric values or a finite set of symbolic values is given by Schumacher and Bergmann [15]. They describe cases and the query as a list of attribute-value-pairs ($A_1=a_1,\ldots, A_n=a_n$) and consider that all attribute types are of the named scales of measurement. Because of these constraints one can consider the case base as an n-dimensional Euclidian space and every query is a SQL-statement like:

```
SELECT a1,…,an FROM CaseTable
WHERE (ai1 ≥ mini1 AND ai1 ≤ maxi1)…
AND   (aim ≥ minim AND aim ≤ maxim)
```

where $a_{i1},\ldots,a_{im}$ are those attributes in the query which are not undefined, $m \leq n$, as shown in Fig. 2. Such SQL-queries are formulated and executed continuously and the lower and upper boundaries are expanded with every relaxation step. Fig. 2 shows the retrieval rings as a result of the relaxation for two attributes.

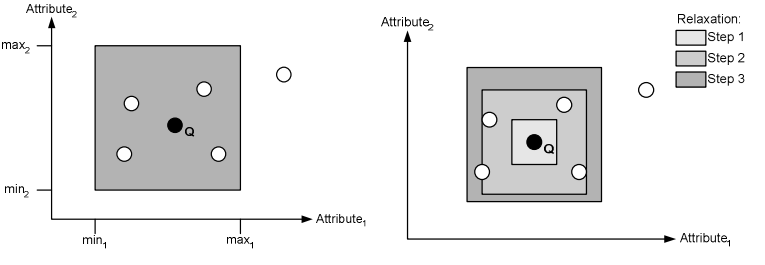


Fig. 2. Rectangle of a SQL-Query and Retrieval Rings as shown in [15]

The difficulty with this procedure is the definition of the new boundaries in every relaxation step, because it is essential for relaxation speed. Schumacher and Bergmann [15] introduce some refinements to optimize these decisions. Furthermore, their model is 'based on the idealistic assumption of a uniform distribution of the cases in the representation space'. If this assumption is violated efficiency decreases.

2.2 Bayesian Decision Theory

From the viewpoint of the query (or from the software module which has to control the relaxation) the distribution of the documents (or database entries) in the knowledge space (or data base) is unknown. At the moment of a relaxation step there is uncertainty about choosing the best attribute for the next step. General decision theory tries to judge decisions, which depend on a sampling, by their risk or their costs of an error. Bayesian decision theory uses in addition the principle of Bayes [16, 17].

Assume a set A of possible decisions (actions). The best decision leads to a parameter (or state of nature) θ which is unknown. A wrong decision a out of A leads to a loss resp. to higher costs. Hence, one could define a loss function l on $\Theta \times A$. Looking at the set X of all possible samples, which could be used to solve the decision problem, a decision rule δ is a function from X to A, which chooses a decision out of A by virtue of a sample out of X. The risk of this decision is defined as:

$$R(\theta, \delta) = E_{X|\theta}\big[l\big(\theta, \delta(X)\big)\big] \tag{1}$$

The risk defines the average loss over the values of X. Bayesian decision theory uses also the so called a priori knowledge, which exists as an a priori distribution of the unknown parameter θ. This distribution is named $\pi(\theta)$. Because of the Bayes-Principle one must choose the decision which minimizes the Bayes-Risk:

$$r(\pi, \delta) = E_{\theta}[R(\theta, \delta)] \tag{2}$$

Berger [18] shows that one can also use the Extended Bayes Decision Rule and minimize the posteriori risk. This means that one chooses a decision rule $\delta(X)$ which minimizes

$$\rho\big(\pi, \delta(X)\big) = E_{\theta|X}\big[l\big(\theta, \delta(X)\big)\big] = \int l\big(\theta, \delta(X)\big) p(\theta|X)\, d\theta \tag{3}$$

where $p(\theta|X)$ is the posteriori distribution of θ. For the further discussion we assume a discrete $\Theta = \{\theta_1, \ldots, \theta_N\}$. This leads to:

$$\rho(\theta_i|X_n) = \frac{\pi_i P(X_n|\theta_i)}{m(X_n)} \tag{4}$$

with $m(X_n) = \sum_j \pi_j P\big(X_n|\theta_j\big)$

Using equation 3 one gets the posteriori risk:

$$\begin{aligned} \rho\big(\pi, \delta(X_n)\big) &= E_{\theta|X_n}\big[l\big(\theta, \delta(X_n)\big)\big] \\ &= \sum_i l\big(\theta_i, \delta(X_n)\big) \cdot p(\theta_i|X_n) \\ &= \sum_i l\big(\theta_i, \delta(X_n)\big) \cdot \frac{\pi_i P(X_n|\theta_i)}{m(X_n)} \end{aligned} \tag{5}$$

Because of the identity of $m(X_n)$ for all decisions, it will suffice to look for a decision $\delta(X_n)$ that minimizes the following equation:

$$\sum_i \pi_i \cdot l\big(\theta_i, \delta(X_n)\big) \cdot P(X_n|\theta_i) \tag{6}$$

2.3 Using Ideas of Bayesian Decision Theory for Query Relaxation

In the following consideration, we will look at a database D which is represented by an N-dimensional space with a query-vector $q = (q^1,\ldots,q^N)$, and database entries $d = (d^1,\ldots,d^N)$. q^j resp. d^j are values of attribute j of query q resp. entry d. We also assume a set $\Theta = \{\theta_1,\ldots,\theta_N\}$ of taxonomies where every attribute of the data base is represented by a taxonomy. Figure 3 shows an example of three documents and their assignments to two taxonomies.

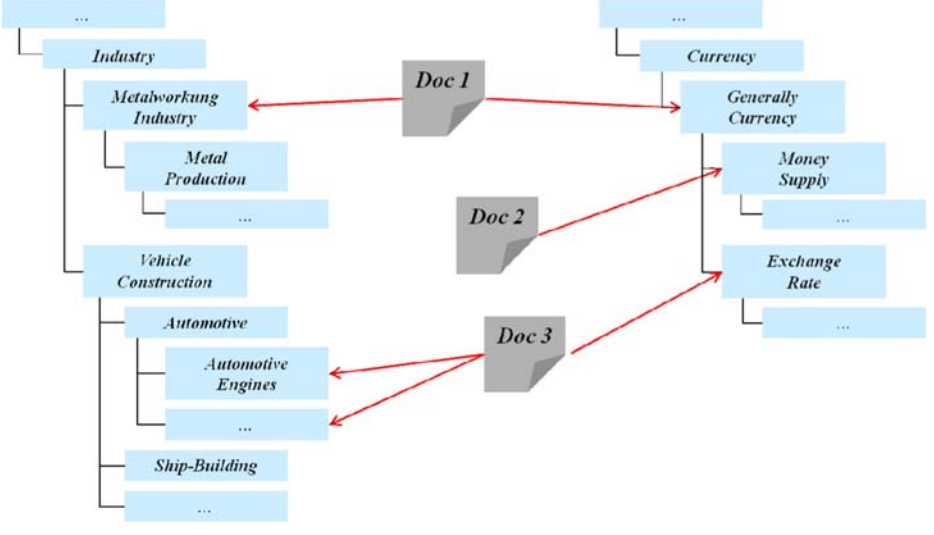


Fig. 3. Taxonomies and assigned documents

Every node of a taxonomy holds a similarity value as a lower boundary for the similarity between all successors of the node. For calculating local similarity of two nodes in a taxonomy we use the following similarity measure which is inspired by [19, 20]:

$$sim\big(node_q^t, node_c^t\big) = \begin{cases} 1, \text{ if } node_c^t \text{is a successor of } node_q^t \\ 1, \text{ if } node_q^t = node_c^t \\ s^t_{<node_q^t, node_c^t>} \end{cases}$$

where $s^t_{<node_q^t, node_c^t>}$ is the described similarity value of the lowest common predecessor of q and c. The calculation of similarity can be demonstrated with Figure 4. For example, the similarity between "Ship-Building" and "Metal Production" is determined by looking for the lowest common predecessor. This is "Industry" in the example which holds the similarity value of 0.2.

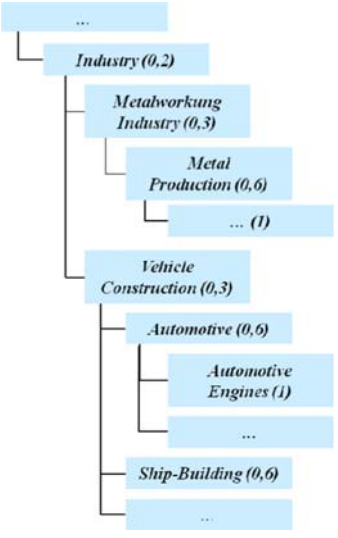


Fig. 4. Similarities and taxonomies

Relaxation of a query in a taxonomy means to get a step up in the taxonomy (generalizing the value of the attribute). Retrieval starts with looking for a hit in a taxonomy, i. e. looking for all d^j with $sim^j(q^j, d^j) = 1$. If no hit is found relaxation would be performed as described and the search starts again. The challenge is, to find the best taxonomy to perform this relaxation step.

To use Bayesian Decision Theory to relax a SQL-Query one has to search the decision rule $\delta(X_n)$ which chooses the best taxonomy under a given result set X_n. We use the similarity measure

$$SIM(q,d) = w^1 \cdot sim^1(q^1, d^1) + \cdots + w^N \cdot sim^N(q^N, d^N) \text{ with } \sum w^i = 1. \quad (7)$$

Because a given query leads always to the same result set X_n, the retrieval is a random experiment with probability $P(X_n) = 1$. Therefore, one could formulate an optimization problem based on Bayesian Decision Theory. Looking at equation (4) while using $P(X_n) = 1$ and $\sum \pi_j = 1$ one gets

$$p(\theta_i | X_n) = \pi_i. \quad (8)$$

Using equation (5) one could define the following optimization problem: search the decision $\delta(X_n)$ under a given result set X_n which minimizes

$$\sum_{i=1}^{N} \pi_i \cdot l\big(\theta_i, \delta(X_n)\big). \quad (9)$$

Next, the a priori probabilities have to be defined in such a manner, that the probability increases with an increasing number of assignments[2] of data base entries to the taxonomy:

[2] A case or data base entry is assigned to a taxonomy, if the value of the case is represented by one node of the taxonomy.

$$\pi_i = \frac{Assignments\ to\ Taxonomy_i}{Total\ Assignments\ to\ Taxonomies} =: \frac{M^i}{M} \tag{10}$$

The loss function l expresses the loss that results out of choosing θ_i instead of the best θ_j. The consequence of this decision is a smaller gain of local similarity or – caused by the weights of the similarity function – a smaller gain of global similarity.

To get the new relaxed query q_{n+1} one must move one step up in the taxonomy t of attribute t or:

$$q_{n+1}^t = node_{k_{n+1}^t}^t > node_{k_n^t}^t \tag{11}$$

So the smaller gain on local similarity is a result of

$$s_{k_{n+1}}^{\theta_j} - s_{k_{n+1}}^{\theta_i} . \tag{12}$$

As a consequence, the loss function l can be defined as:

$$l\left(\theta_j, \delta(X_n)\right) = v\left(\theta_j, X_n\right) \cdot w^{\theta_j} \cdot s_{k_{n+1}}^{\theta_j} - v(\delta(X_n), X_n) \cdot w^{\delta(X_n)} \cdot s_{k_{n+1}^{\delta(X_n)}}^{\delta(X_n)} \tag{13}$$

$$v\left(\theta_j, X_n\right) = \begin{cases} 1, \text{if } M_n^{\theta_j} = 0 \\ \frac{1}{K}, \text{if } M_n^{\theta_j} = M^{\theta_j} \\ \frac{1}{M_n^{\theta_j}}, \text{else} \end{cases}$$

where $K < \infty$ and M_n^i=Assignments to Taxonomy i in step n.

The elements of the loss function can be easily computed and the relaxation step can be executed in the taxonomy with the smallest posteriori risk of the decision rule.

A special case of this optimization problem is given when the uniform distribution of (10), or $\pi_i = 1/N$ is assumed. Assuming equal weights in the similarity function and a negligible difference in the gain of local similarity, equation (13) leads to the simplified loss function:

$$l\left(\theta_j, \delta(X_n)\right) = w \cdot \left(v\left(\theta_j, X_n\right) - v(\delta(X_n), X_n)\right). \tag{14}$$

Then the optimization problem simplifies to

$$\sum_{j=1}^{N} \frac{1}{N} \cdot w \cdot \left(v\left(\theta_j, X_n\right) - v(\delta(X_n), X_n)\right) \tag{15}$$

and one only has to consider

$$\sum_{j=1}^{N} \left(v(\theta_j, X_n) - v(\delta(X_n), X_n) \right). \tag{16}$$

With $0 < M_n^{\theta_j} < M^{\theta_j}$ one gets for all $\theta_j \in \Theta$:

$$\sum_{j=1}^{N} \left(\frac{1}{M_n^{\theta_j}} - \frac{1}{M_n^{\delta(X_n)}} \right) = -N \cdot \frac{1}{M_n^{\delta(X_n)}} + \sum_{j=1}^{N} \frac{1}{M_n^{\theta_j}} \tag{17}$$

To get the best taxonomy to use for relaxation, $\delta(X_n) = \theta_i$ needs to be found that minimizes equation (17). Obviously $M_n^{\theta_i}$ must be smaller than all $M_n^{\theta_j}$ with $i \neq j$. Therefore, the simplified rule chooses the taxonomy which has the least assignments of the result set X_n.

3 An Application

To test the approach we chose a database maintained by the company called Deutsche Gesellschaft für Ad-hoc-Publizität (DGAP). The database contains specific news that DGAP receives from public companies listed on a German stock exchange in order to publicize them widely. These are news items that public companies are required to publicize by §15 of the German law on stock trading (Wertpapierhandelsgesetz). Similar requirements are also formulated in the EU Transparency Directive that has been implemented in German law in January, 2007. The purpose of these regulations is to protect the interests of stakeholders of public companies, esp. their investors. DGAP helps its customers to comply with the regulations. Such news items are, for example, announcements of periodic reports, earnings warnings, director's dealings,

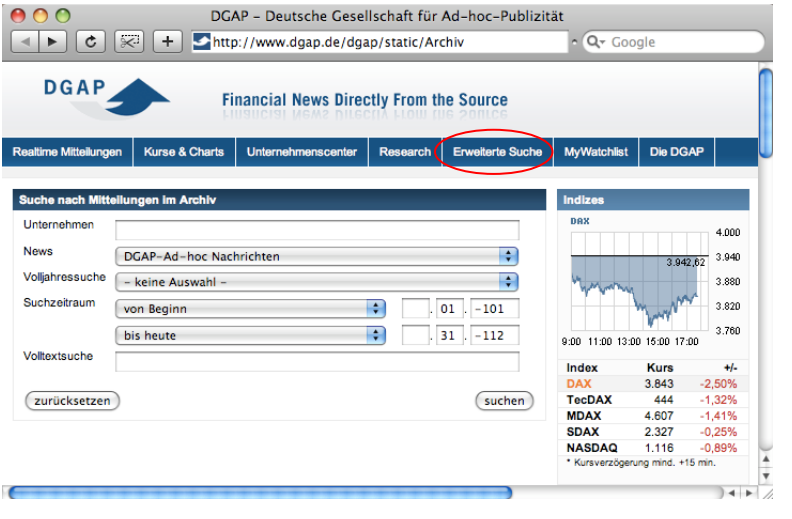


Fig. 5. Database Interface at http://www.dgap.de/dgap/static/Archiv/

or takeover bids. DGAP distributes the news upon their arrival to the stock exchanges, overseeing authorities, and the general public. Nowadays, the general public is mainly informed through the Internet. Internet users can find the information at http://www.dgap.de. The database interface is shown in Fig. 5. It is the genuine interest of DGAP that the news it receives and publishes is as well accessible as possible.

The "extended search" (in German: Erweiterte Suche) presented in the figure allows just the specification of some attributes: company name, DGAP news category, full year or time period, and a "full text search." The latter only allows the specification of one keyword.

One of DGAP owners provided us with about three months of historical data for our research. We created taxonomies for some of the attributes, e.g., industry structure. Then, we ran a number of simple SQL queries using keywords that we connected with logical "and" or "or" to mimic the use of the database via Internet (actually, the current interface does not even provide logical connectors as it was the case in the past except for the implicit "and"-connection of the terms in the Web form). The same queries were then submitted to a program that we developed to implement the search strategy described in section 2.

Tab. 1 displays the results for various queries (originally entered in German). The table contains the number of news items retrieved and, in parentheses, how many of them were relevant. The results of Query Relaxation are given for three different levels of calculated similarity (see equation (7)). As a reading example, we interpret the results for the terms "board of directors" and "earnings forecast". If the terms were connected by the logical "and" then nine news items were found which were all relevant. The "or"-connection yielded 481 news items of which only 32 were relevant. After relaxing the query with our method and requiring a similarity of 100%, twelve news items were retrieved which were all relevant. Decreasing the required similarity to the range from 60% to under 100% led to the retrieval of 35 news items of which 20 were relevant.

The table shows that our query relaxation approach returns more records than the "and"-connection but much less than the "or"-connection. This reduction in recall can lead to a reduction in precision but in a real application with a large database, users would not be able or willing to analyze all retrieved records for relevance. A ranking algorithm mainly based on keywords would also be of limited help.

Table 1. Comparison of simple query results with results of query relaxation

Query Terms	**DGAP**	**DGAP**	**Results of Query Relaxation**		
			calculated similarity		
	and	or	100%	60% to <100%	50% to <60%
"sales plan"	7 (7)	not possible	48 (28)	-	-
"credit institution" "balance sum"	-	22 (4)	1 (1)	1 (0)	18 (3)
"board of directors" "earnings forecast"	9 (9)	481 (32)	12 (12)	35 (20)	-
"CFO" "Loss" "Nemax"	-	107 (4)	-	5 (4)	-

4 Summary and Future Work

The described approach does not support general search of the Deep Web. It is designed for specific databases or vertical applications where application or semantic knowledge can be represented in taxonomies. It has been criticized that approaches of the Semantic Web where mark-up of data is necessary are spreading too slow because database owners avoid the work [21]. Our approach only requires the development of relatively stable taxonomies. Mark-up or any changes of data structures, including already archived data, are not necessary.

Future work has to handle questions of representing arbitrary attribute-value representations, e.g. continuous attributes or ordered sets of entities. In this case the definition of the a priori probabilities is equal to equation (10). For the definition of the loss function it is necessary to use the local similarity measure. Based on this, an estimation of an upper bound of the gain on local similarity can be made because the local similarity decreases in the next relaxation step. Hence,

$$s^{\theta_j}_{k^{\theta_j}_{n+1}}$$

in equation (12) could be substituted with $sim^{\delta(X_{n-1})} - \varepsilon$, where ε is an arbitrary value that controls the relaxation step and $sim^{\delta(X_{n-1})}$ denotes the minimum gain on local similarity added by the cases in relaxation step *n-1*.

References

1. He, B., Patel, M., Zhang, Z., Chang, K.: Accessing the Deep Web. Communications of the ACM (CACM) 50(5), 95–101 (2007)
2. Shestakov, D.: Search interfaces on the Web: Querying and Characterizing, TUCS Dissertations 104 (2008)
3. Shestakov, D.: Deep Web: Databases on the Web. In: Entry in Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends (2009)
4. Spink, A., Wolfram, D., Jansen, M.B.J., Saracevic, T.: Searching the Web: The Public and Their Queries. J. of the American Society for Information Science and Technology 52(3), 226–234 (2001)
5. Wright, A.: Searching the Deep Web. Communications of the ACM (CACM) 51(10), 14–15 (2008)
6. Chu, W.W., Chen, Q., Lee, R.-C.: Cooperative Query Answering Via Type Abstraction Hierarchy. Technical Report CSD-900032, Computer Science Department, University of California (1990)
7. Chu, W.W., Yang, H., Chow, G.: A Cooperative Database System (CoBase) for Query Relaxation. In: Proceedings of the 3rd International Conference on Artificial Intelligence Planning Systems, AIPS 1996, Edinburgh (1996)
8. Gaasterland, T., Godfrey, P., Minker, J.: Relaxation as a Platform for Cooperative Answering. Technical Report CS-TR-2818, Institute for Advanced Computer Studies and Department of Computer Science, University of Maryland (1991)
9. Kumaran, G., Allan, J.: Selective User Interaction. In: Proceedings of the 16th Conference on Information and Knowledge Management, CIKM 2007, pp. 923–926 (2007)

10. Carpio, G.V.G., Abrouk, L., Cullot, N.: A Query Expansion Methodology in a Cooperation of Information Systems Based on Ontologies. In: Proceedings of the 5th International Conference on Web Information Systems and Technologies, WEBIST 2009, pp. 256–261 (2009)
11. Tomassen, S.L., Gulla, J.A., Strasunskas, D.: Document Space Adapted Ontology: Application in Query Enrichment. In: Kop, C., Fliedl, G., Mayr, H.C., Métais, E. (eds.) NLDB 2006. LNCS, vol. 3999, pp. 46–57. Springer, Heidelberg (2006)
12. Schweighofer, E., Geist, A.: Legal Query Expansion using Ontologies and Relevance Feedback. In: Proceedings of the 11th Conference on Legal Ontologies and Artificial Intelligence Techniques, LOAIT 2007, pp. 149–160 (2007)
13. Shimazu, H., Kitano, H., Shibata, A.: Retrieving Cases from Relational Data-Bases: Another Strike Towards Corporate-Wide Case-Base Systems. In: Proceedings of the 13th International Joint Conference in Artificial Intelligence, IJCAI 1993 (1993)
14. Watson, I.: A Case-Based Reasoning Application for Engineering Sales Support Using Introspective Reasoning. In: Proceedings of the 17th National Conference on Artificial Intelligence (AAAI 2000) and the 12th Innovative Applications of Artificial Intelligence Conference (IAAI 2000), vol. 1, pp. 1054–1059. AAAI Press, Menlo Park (2000)
15. Schumacher, J., Bergmann, R.: An Efficient Approach to Similarity-Based Retrieval on Top of Relational Databases. In: Blanzieri, E., Portinale, L. (eds.) EWCBR 2000. LNCS, vol. 1898, pp. 273–284. Springer, Heidelberg (2000)
16. Bernardo, J.M., Smith, A.F.M.: Bayesian Theory. John Wiley and Sons, New York (1993)
17. Carlin, B.P., Louis, T.A.: Bayes and Empirical Bayes Methods for Data Analysis. Chapman and Hall, New York (1996)
18. Berger, J.O.: Statistical Decision Theory. Springer, Berlin (1980)
19. Bergmann, R.: On the Use of Taxonomies for Representing Case Features and Local Similarity Measures. In: Proceedings of the 6th German Workshop on Case-Based Reasoning (1998)
20. Bergmann, R., Stahl, A.: Similarity Measures for Object-Oriented Case Representations. In: Smyth, B., Cunningham, P. (eds.) EWCBR 1998. LNCS, vol. 1488, p. 25. Springer, Heidelberg (1998)
21. Silverstein, C., Henzinger, M., Marais, H., Moricz, M.: Analysis of a Very Large Web Search Engine Query Log. ACM SIGIR Forum 33 (1), online version, 7 p. (Fall 1999)

Using Semantics to Personalize Access to Data-Intensive Web Sources

Monika Starzecka and Adam Walczak

Poznan University of Economics, Al. Niepodleglosci 10, 61-875 Poznan, Poland
{m.starzecka,a.walczak}@kie.ue.poznan.pl
http://kie.ue.poznan.pl

Abstract. Data-intensive Web sites are an important and growing source of well-structured information on the Web. Their potential value remains largely unused as they pose a number of challenges to both machine and human users. They are dispersed and provide heterogeneous, rigid, site-specific and non-personalizable query and navigation interfaces. In this paper we present outline of method for accessing data from data-intensive Web sites in an uniform way. Our method is independent of the source and allows for personalization of the access to data. We describe how domain ontology is used for definition of personalized GUI. Then initial evaluation of described method is provided.

Keywords: Data-Intensive Web sites, Deep Web, personalization, semantics, ontologies.

1 Introduction

The contemporary World Wide Web contains unprecedented quantities of information. While basic Web technologies focus on unstructured text, huge part of Web sites (including Deep Web [7]) are *data-intensive* and contain semistructured content. Examples of such sites are on-line databases (e.g. Deep Web search engines), commercial Web sites (e.g. on-line stores or e-auctions) and Web applications (e.g. Web calendars, social networking sites). Due to size and quality of data-intensive Web sites content, they are priceless resource for both individual and organizational users.

From a user's perspective, data-intensive Web sites are quite challenging to access for several reasons. In most cases, they are dispersed and hard to find.

Web sites, even if they belong to the same domain, usually differ significantly with respect to their user interface (i.e. how data is presented to the user), navigation patterns (i.e. if navigation is form or link based) and data organization (i.e. how attributes and classes of entities are split into different types of pages). Typically navigation and data organization cannot be personalized and not necessarily correspond to user's needs and view of the domain. Advanced users build wrappers or copy the data manually to some application that allows for greater flexibility in data manipulation.

In this paper we present an approach for personalized access to data from data-intensive Web sites in an integrated and semantics-aware way.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 28–38, 2009.

2 Related Work

Research on Web data access, presentation and extraction started at the very dawn of World Wide Web, giving birth to the wrapper construction field [16]. As the Web grown, enterprises started to expose data in this environment and enable its querying. Thus Deep Web was born [5] and became a huge source of well structured Web data. There are many Deep Web systems (including [20,9,18,2]) focusing rather on accessing and indexing content than on extracting data from the sources. Merely few solutions dealing with data-intensive Web sites in a semantic way are presented. Gal et al. [23] propose a framework that supports extraction of ontologies from Web search interfaces, ranging from simple Search Engine forms to multiple-pages, complex reservation systems. OntoBuilder enables fully-automatic ontology matching, yet the solution doesn't deal with the problem of data extraction.

In [3] adding of an ontology layer to the Deep Web structure is proposed. With the use of given ontology (in provided example it is WordNet) synonyms for keywords from user's query are listed. Then, the appropriate attributes in Deep Web sources can be found, by comparison with provided keywords and listed synonyms (with given similarity rate). Similar approach is presented in [24], where the process of filling forms is automated by correlating web form labels to entries in a domain ontology. Matching the ontology concepts with appropriate web form labels is achieved through the use of a continually refined knowledge base and application of LSD system [10].

Data-intensive Web sites offer typically very simplistic and rigid query interfaces with limited querying capabilities. Such simplicity is sometimes desirable from a human point of view but comes at a cost of limited flexibility and constraints for automated querying of multiple sources. Methods for dealing with querying via interface with limited capabilities were studied e.g. in [12]. Source querying capabilities description and associated query rewriting problems were previously addressed for mediator systems working with dispersed databases (for example [19]).

So far little effort was devoted to personalizing access to the data-intensive Web sites. To certain extent all the previously mentioned work referring to wrappers or information integration touches this subject, but in fact provides just the first step - unified view on multiple sources. The second step - to provide personalized access for individual users is often neglected. We found the work of Bigham and colleagues to refer to personalized access [6], yet they too constrain themselves to extraction and navigation in the Web data sources.

3 System Overview

Building on our concept of navigating and extracting from the data-intensive Web sites using finite state machines we aim at personalizing the access to the data sources.

We adopted formal ontology as a way to describe the source and its data. The ontology is both underlying structure that is personalized according to user's

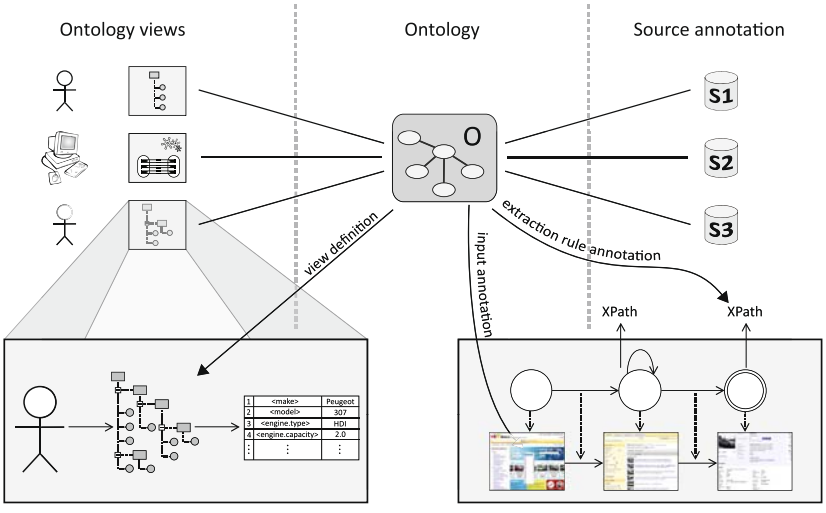


Fig. 1. Overview of System's – Conceptual Architecture

perception of the particular source, and provider of the domain data necessary for query execution. Conceptual architecture of the system shows Fig.1. The main components of our system are:

- Domain ontology - represents common domain model for all the sources considered for particular task (see Section 4).
- Views and Profile Manager - responsible for creating and managing user profiles (see Section 5).
- Source Description Manager - allows for maintenance of relations between particular source and the domain ontology e.g. referring form fields and presentation labels to the concepts in the ontology (description of site descriptions management is out of scope of this paper).
- Query planner - the component responsible for building source-specific query plans based on semantic description of individual data-intensive Web sites (detailed description of query planning is out of scope of this paper, basic idea was presented in [1]).
- Query execution engine - the component that actually navigates the data-intensive site and extracts information from relevant pages (using finite state machine model) based on the source description and its binding to the instances stored in the ontology together with user's profile (see [1]).
- Graphical User Interface - the component responsible for building profiles and queries and visualizing the retrieved data (see Section 5.2).

4 Domain Ontology

Automotive Ontology was developed from scratch for the needs of domain source description, and as a means to determine user informational needs. Its main

purpose is to provide shared conceptualization, that would enable car description standardization from the perspective of automotive Web sites. It includes all car attributes, that may be determined on the web site, or which must be provided by the user according to web sites navigational process. The system which part of we are describing in the paper is now during the stage of a prototype, therefore some simplistic assumption were made. We decided to analyze car description until the level of version definition, thus we are not handling the information concerning equipment and peripherals (such as air-conditioning, CD player, body or upholstery coloring, etc.). Essential for us is information about: make, model, body type, gearbox, engine, fuel, version and drive. Those notions constituted concepts in designed automotive ontology.

- Model - a car model is a particular brand of vehicle sold under a make. From an engineering point of view, a particular car model is usually defined and/or constrained by the use of a particular car chassis/body type combination.
- Make - a make is a brand name. For example, Chevrolet and Pontiac are marques of their maker, General Motors.
- Body type - body types are largely (though not completely) independent of a car's classification in terms of price, size and intended broad market; the same car model might be available in multiple body styles (or model ranges). Ex.: Sedan, Hardtop, Coup, Limousine
- Gearbox - gearbox provides a speed-torque conversion from a higher speed motor to a slower but more forceful output or vice-versa. For our needs we distinguish following gearbox types: manual, automatic, semi-automatic
- Engine - a car engine is a machinery in which the combustion of a fuel occurs with an oxidiser (usually air) in a combustion chamber. In a car engine the expansion of the high temperature and pressure gases (that are produced by the combustion) directly apply force to a movable component of the engine, such as the pistons or turbine blades and by moving it over a distance, generate useful mechanical energy. From our perspective, the most important attributes of an engine are: capacity, power and type (for example: TDI, CTDI, etc.)
- Fuel - fuel is any material that is burned or altered in order to obtain energy and to heat or to move an object. Fuel releases its energy through a chemical reaction means (such as combustion).
- Car Version - car version is a particular category of cars offered on the market with specific combination of attributes: model, engine, body and gearbox.

To make an ontology fully useful for determined purpose we needed to create instances of previously defined concepts. For this task we used information from data base that was facilitated to us, by our business partner from automotive industry. Current version of the ontology was developed in OWL, it has seven concepts, over a dozen of properties and approximately five hundred of instances.

5 Views and Profile Manager

It is common practice among data-intensive Web sites to provide a single, rigid user interface. Our proposal is to give the user possibility of defining his own interface independently of the one provided by source owner. Assuming that appropriate domain ontology is given, we designed a solution which enables user to create her own query interfaces and data views for data-intensive Web sites.

5.1 User Profile

In our approach, we allow user to represent her perception of the domain by selecting part of the domain ontology and saving it in *user profile*. It specifies what are the attributes typically used to access data, what is their order of importance, what combinations of attributes should be selected together to avoid too many clicks or improve readability (e.g. 'Opel Astra' - make and model together, or '2.0 TDI' for engine capacity and type) and which attributes should be retrieved in answer to a query. As a consequence, user profile defines what attributes are not important for specific user. User profile also contains the list of data-intensive Web sites that the user wants to query.

After a thorough examination of existing ontology visualization methods [22] we decided to use tree view for ontology representation in user interface. We found intuitivity and simplicity in data representation of this method advantageous. Together with numerous shortcomings pointed by researchers, this technique in few conducted tests (for example by [21,15,8,14]) proved to be more efficient in comparison with other techniques used for visualization of hierarchical structures. No particular explanation can be found for such a good tests results of this quite primitive technique. The most possible reason may be that user can find this technique intuitive, as it is similar to the way that data are represented in everyday life (for example: table of contents in a book, a list of tasks to do, etc.). For results presentation we propose to use grid view (definition of projection part of the query).

To enable personalized access to a variety of on-line data, we let the user define structures of both tree and grid. While defining tree structure, user determines the number of levels used in query definition and the list of attributes for each level. As the same user may have very different requirements in different usage scenarios, she is allowed to define multiple profiles in our prototypical implementation and switch between them when needed.

The user interface for definition of profiles is displayed on Fig.2. In this interface user may choose to apply an existing profile (by selection of a list option and pushing "OK" button) or manage her list of profiles. She may add a new empty profile ("Add Profile" button) or build a new profile starting with existing one ("Copy Profile" button). She may also modify any profile by redefining tree structure, grid structure or list of sources to be used. In this example, user has already defined the first level of the tree to contain make and model names, and is currently defining the second level.

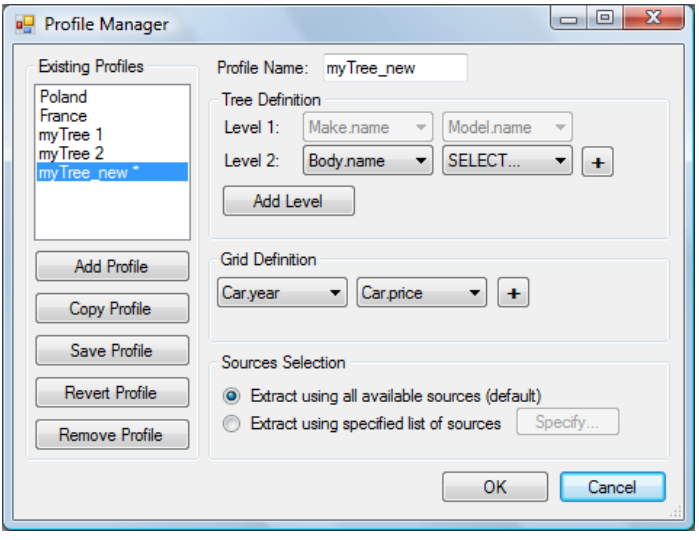


Fig. 2. Profile management and definition user interface

Common view of specific domain may reflect cultural differences between countries or organizational differences between companies. For example in automotive domain, on German sites fuel type is a very important attribute of the car, while on Polish Web sites the user is rarely asked to specify this attribute. Thus the list of profiles may also include predefined profiles for specific countries ("Poland" and "France" in Fig.2.). Predefined profiles may be good starting point for definition of new tree or grid structure.

In the remaining of this paper we assume the following definition of the user profile. The grid includes year, price, mileage, fuel type and number of bids. The tree has following three levels:

- level 1: make and model,
- level 2: body and gearbox type, and
- level 3: engine capacity.

5.2 Graphical User Interface

In our system we propose two main ways of querying data-intensive Web sites. The first approach is based on user profiles described in two previous sections and corresponds to *long-term personalized access to data-intensive Web sites*. The second approach enables dynamic, intuitive definition of user query by selecting attributes in any arbitrary order (possibly very different from Web sites navigation pattern), thus enabling *ad hoc personalized access to data-intensive Web sites*.

The user interface enabling access to data-intensive Web site in the first case is displayed in Fig. 3. The tree in this view is constructed based on tree definition in user's profile, and filled in with values acquired from domain ontology. After

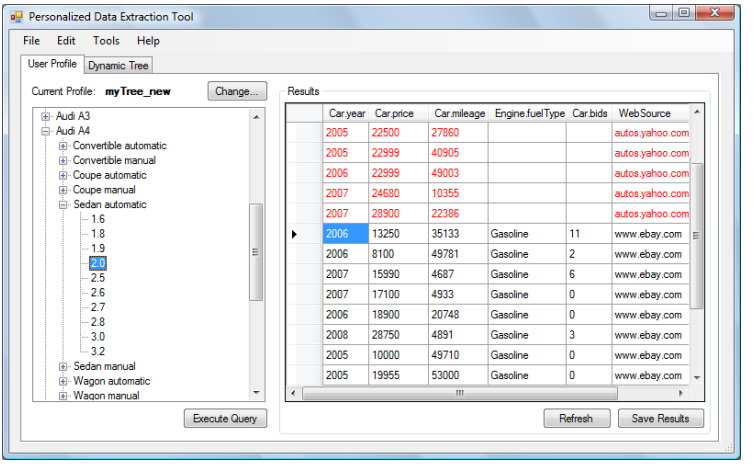


Fig. 3. Profile-based personalized view of automotive domain for profile defined in Fig.2. Data returned for specific query shown in the grid; records from sources not containing some of attributes are marked in red.

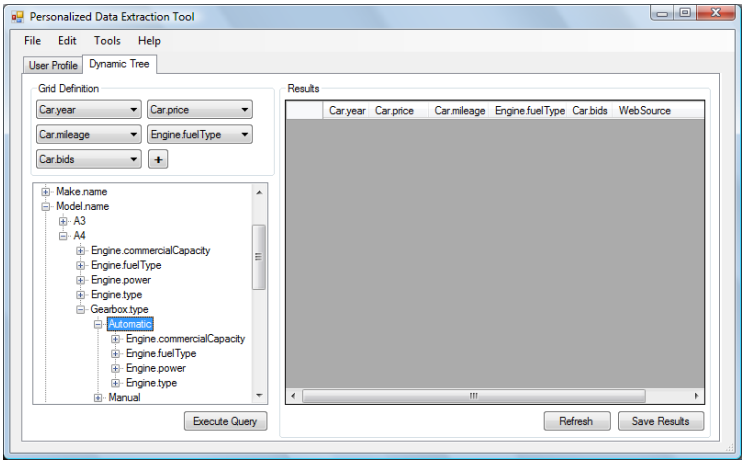


Fig. 4. Dynamic tree view of automotive domain

the user selects any item of the tree and presses "Execute" button, the query definition, rewriting and execution is performed and the result is displayed in the grid in the right part of the window. As some of data attributes may be missing or empty in some sources, some cells in the grid may remain empty.

The user interface enabling ad hoc personalization is displayed in Fig. 4. While it is similar to the previous view, few significant differences should be emphasized. Firstly, while by default the grid structure from active profile is reused, it is possible to change this structure at query time. Secondly, the contents of the tree is very different. At each odd level of the tree, user has to choose the name of

attribute that will be determined next, and at each even level she selects specific value for the attribute. The set of attributes presented in the dynamic tree is rigid and corresponds to all ontology instances. Every two steps (attribute selection and value determination) allow to determine single attribute. While, such method gives more flexibility than currently existing solutions, its usefulness might decrease when the number of used attributes is high.

6 Idea Validation

The examination described in this section was performed to verify presented idea in comparison with standard navigation in data intensive sources from a perspective of their efficiency. Assumption made during validation process was that we have a complete set of sources descriptions needed for automatic navigation through the analyzed data sources. As the efficiency measure we decided to use a number of activities that need to be performed by the user in order to reach the information she is looking for. Under the notion of activity we understand click, choice from drop-down list, radio button selection etc. as well as visual scanning of big table of results. The fewer activities need to be performed, the more efficient is navigation. The size of test set was amounted to 46 sources[1]. The web pages that composed final test set were chosen from around 300 automotive data sources indicated by practitioners as reliable and useful sources of information.

Just as example in Section 5, the examination was made for automotive industry domain. It covered two scenarios of situations when user wants to find current price of particular car. First one: the user knows exactly what kind of a car she is looking for and has defined navigational tree as: 1st level: make, model, body; 2nd level: engine, fuel; 3rd level: gear box, car version. In second scenario we assumed that the user has exact preferences regarding just three attributes of the car: make, model and body. Every one of the attributes constitutes separate level in navigational tree.

Figures 5 and 6 shows results respectively for first and second scenario. Charts represents how number of required activities grows with the number of searched sources. It can be easily seen that in both cases the examination came off better for our method.

The more sources were searched the bigger difference in efficiency between analyzed navigational methods was noticed. It is so, because in standard method user needs to determine searched car attributes for every single source separately. Our proposed method after defining tree structure and single car attributes determination uses the information for every chosen source. For all 46 sources it amounted to 202 activities for the first scenario and 1012 for the second one. For second scenario the difference was significantly bigger. The reason is inflexibility

[1] The number was calculated by the equation: $n = \left(\frac{z\sigma}{e}\right)^2$, where: n - searched test set size, z - value of cumulative distribution function for normal distribution with given statistical significance(90%), σ - standard deviation estimator, e - value of permissible error (0,6).

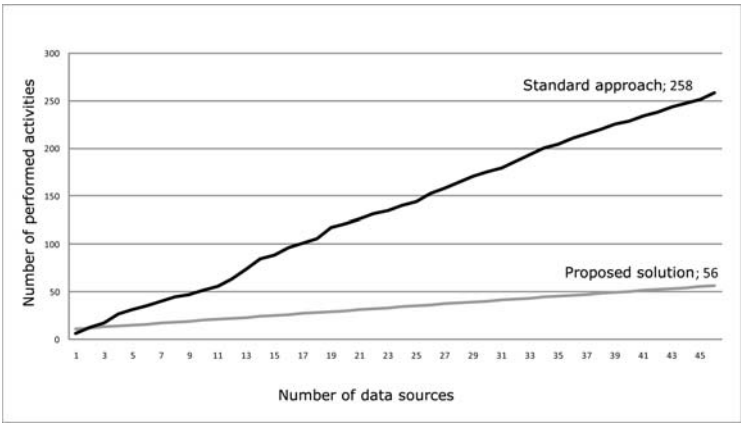


Fig. 5. Relation between number of activities required to achieve user goal and number of data sources - first scenario

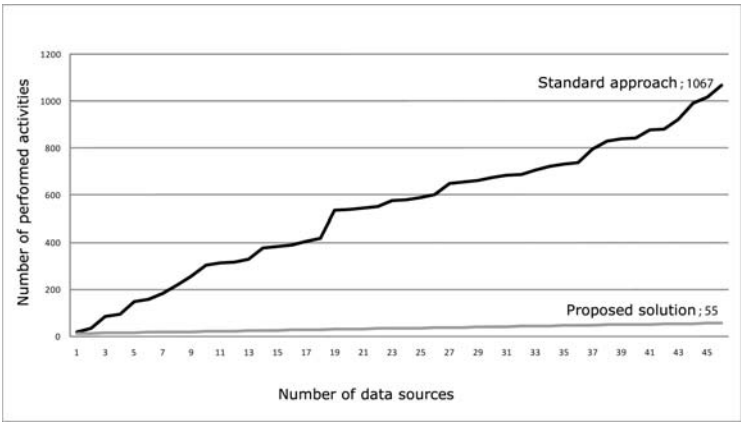


Fig. 6. Relation between number of activities required to achieve user goal and number of data sources - second scenario

of standard navigational structure provided by web sites. User has no possibility to change navigational path of the web page, and when her needs are incoherent with provided structure, then very often more activities must be performed to find needed information. In our scenario for the user fuel type was indifferent, but in some of the web pages it was second (after model) attribute to determine. In practice it generated few parallel navigational paths (one for each existing fuel type) which generated more activities necessary to perform.

Presented examination clearly shows that even if user must define tree structure and expected results in the system, while searching the same information in many different sources our method is far more efficient then the standard one. These were very simple scenarios. 38 from analyzed sources are dedicated

to only one make. On average each make has 11,36 models, every model is available with 1,46 body types. Standard navigation through the web page, when the user defined only model and body type requires on average 26,36 activities. If the user wants to check car price in all 38 sources - she would have to perform 16320 activities. Remaining 8 sources provide information about 50 makes on average. If our user would like to analyze those sources the number of performed activities would grow to more than 200000.

7 Future Work

In this paper we presented an approach that enables users to construct their personalized and uniform access interface for data-intensive Web sites in specific domain, described by an ontology. There are few more challenges that we plan to solve in our future work. Firstly, to deal with lexical and value-encoding variations, the problem of synonyms and more complex lexical relations (e.g. hyponymy or hypernymy) and data translation rules (e.g. currency or unit transformation) needs to be addressed. Secondly, while proposed simple tree interface works well for textual data attributes, support for more complex tree rules (e.g. by comparison or range operators on integer attributes) would be beneficiary. Thirdly, in order to support arbitrary-depth tree presentation of cyclic relations in domain ontologies (such as similarity between models of cars). Also, as our future work, we plan to handle the problem of emerging web technologies and navigation patterns (such as AJAX, flash).

References

1. Abramowicz, W., Flejter, D., Kaczmarek, T., Starzecka, M., Walczak, A.: Semantically Enhanced Deep Web. In: INFORMATIK 2008 Beherrschbare Systeme dank Informatik, 38. Jahrestagung der Gesellschaft für Informatik, Gesellschaft für Informatik e.V (GI), München, September, 8. bis 13, pp. 673–679 (2008)
2. Alvarez, M., Raposo, J., Pan, A., Cacheda, F., Bellas, F., Carneiro, V.: Deepbot: A focused crawler for accessing hidden web content. In: 3rd international workshop on Data engineering issues in E-commerce and services, pp. 18–25 (2007)
3. An, Y.J., Geller, J., Wu, Y.-T., Chun, S.A.: Semantic deep web: automatic attribute extraction from the deep web data sources. In: SAC 2007: Proceedings of the 2007 ACM symposium on Applied computing, pp. 1667–1672. ACM, New York (2007)
4. Anupam, V., Freire, J., Kumar, B., Lieuwen, D.: Automating web navigation with the webvcr. In: 9th International Conference on World Wide Web, pp. 503–517 (2000)
5. Bergman, M.K.: The deep web: Surfacing hidden value. The Journal of Electronic Publishing 7(1) (2001)
6. Bigham, J.P., Cavender, A.C., Kaminsky, R.S., Prince, C.M., Robinson, T.S.: Transcendence: Enabling a personal view of the deep web. In: International Conference on Intelligent User Interfaces (2008)
7. Chang, K.C.-C., He, B., Zhang, Z.: Mining semantics for large scale integration on the web: evidences, insights, and challenges. SIGKDD Exploration Newsletter 6(2), 67–76 (2004)

8. Cockburn, A., McKenzie, D.: An evaluation of cone trees. In: Proceedings of the 2000 British Computer Society Conference on Human Computer Interaction (2000)
9. Chang, K.C.-C., He, B., Zhang, Z.: Metaquerier: querying structured web sources on-the-fly. In: 2005 ACM SIGMOD International Conference on Management of Data, pp. 927–929 (2005)
10. Doan, A., Domingos, P., Halevy, A.Y.: Reconciling schemas of disparate data sources: A machine-learning approach. In: SIGMOD Conference (2001)
11. Flesca, S., Gottlob, G., Baumgartner, R.: Supervised wrapper generation with lixto. In: 27th International Conference on Very Large Data Bases, pp. 715–716 (2001)
12. Halevy, A.: Theory of answering queries using views. SIGMOD Record 29(4), 40–47 (2000)
13. Handschuh, S., Staab, S., Volz, R.: On deep annotation. In: WWW 2003: Proceedings of the 12th international conference on World Wide Web, pp. 431–438. ACM, New York (2003)
14. Katifori, A., Torou, E., Halatsis, C., Lepouras, G., Vassilakis, C.: A comparative study of four ontology visualization techniques in protege: Experiment setup and preliminary results. In: IV 2006: Proceedings of the conference on Information Visualization, Washington, DC, USA, pp. 417–423. IEEE Computer Society, Los Alamitos (2006)
15. Kobsa, A.: User experiments with tree visualization systems. In: INFOVIS 2004: Proceedings of the IEEE Symposium on Information Visualization, Washington, DC, USA, pp. 9–16. IEEE Computer Society, Los Alamitos (2004)
16. Teixeira, J.S., Ribeiro-Neto, B.A., Laender, A.H.F., da Silva, A.S.: A brief survey of web data extraction tools. SIGMOD Record 31(2), 84–93 (2002)
17. Nagao, K.: Digital Content Annotation and Transcoding. Artech House Publishers, Norwood (2003)
18. Ntoulas, A., Zerfos, P., Cho, J.: Downloading textual hidden web content through keyword queries. In: 5th ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 100–109 (2005)
19. Papakonstantinou, Y., Gupta, A., Haas, L.: Capabilities-based query rewriting in mediator systems. In: 4th International Conference on Parallel and Distributed Information Systems (1996)
20. Raghavan, S., Garcia-Molina, H.: Crawling the hidden web. In: 27th International Conference on Very Large Data Bases, pp. 129–138 (2001)
21. Rivandeneira, W., Benderson, B.B.: A study of search result clustering interfaces: Comparing textual and zoomable interfaces. Technical report, University of Maryland HCIL (2003)
22. Starzecka, M.: Nawigacja w serwisach www na podstawie ontologicznego opisu zrode. Master's thesis. Akademia Ekonomiczna w Poznaniu (2008)
23. Gal, A., Modica, G., Jamil, H.: OntoBuilder: Fully Automatic Extraction and Consolidation of Ontologies from Web Sources. In: International Conference on Data Engineering. IEEE Computer Society, Los Alamitos (1996)
24. Walny, J.: Semaform: Semantic wrapper generation for querying deep web data sources. CPSC 502 project under the supervision of Dr. Denilson Barbosa (2007), http://www.ucalgary.ca/~jkwalny/502/index.html

Deep Web Queries in a Semantic Web Environment

Thomas Hornung[1] and Wolfgang May[2]

[1] Institut für Informatik, Universität Freiburg
hornungt@informatik.uni-freiburg.de
[2] Institut für Informatik, Universität Göttingen
may@informatik.uni-goettingen.de

Abstract. Access to Deep Web sources is concerned with querying data that is hidden behind Web forms and primarily not accessible by common query languages. Web forms do not contain any type information, and it thus follows that Deep Web sources only work on string data in its rudimentary form. In this paper, we demonstrate how Semantic Web technologies can be used to first lift Deep Web sources to the level of databases with a precise schema and strong typing information and finally to the level of Semantic Web applications. A special focus in this context is on handling measurements, units and dimensions, which is an important issue when data from multiple Deep Web sources is declaratively combined for more involved querying tasks.

1 Introduction

Most of the information that is needed for daily tasks is available on the Web. The main problem is often not the general access to the information, but to access the right information, and to combine it in an appropriate way. Direct query evaluation is not always possible: most of the data is not immediately available for querying, but kept in the *Deep Web* or *Hidden Web*, which consists of dynamically generated result pages of numerous databases, which can only be queried interactively via Web forms. For the *human* user, these Deep Web sources, made visible as HTML pages, have an implicit semantics. For accessing them in an automated environment, this semantics is not available. In contrast to Semantic Web knowledge bases, and even to databases, Deep Web sources have a very primitive data model: their only concept are strings. Even WSDL specifications of (XML-based) Web services provide more information since they have an notion of "answer" and they specify what datatype is returned as response. Current use of Deep Web sources in computerized workflows very explicitly incorporates the background knowledge of a human, e.g., by explicitly programming Web data extraction processes.

For more generic computerized access, Deep Web sources must be *annotated* by metadata. In a first step, this metadata lifts them to the level of databases where the attributes are assigned with datatypes and optionally simple (range)

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 39–50, 2009.

integrity constraints. On a higher, semantical, level, annotations provide the link to the semantics of an application domain.

Note that one must distinguish between making Deep Web sources machine-*accessible* (which means the tasks of Deep Web navigation to request the hidden data as HTML contents, and to program wrappers to extract data records from these HTML pages) and making them machine-*understandable* which means to lift the extracted data on the level known from databases or even Semantic Web knowledge bases. We build our work on [16] (navigation) and [14] (extraction), that solve the accessibility issue, and we deal with the second issue in this paper.

Structure of the paper. We introduce the MARS framework that provides the environment for informational workflows using Deep Web queries in this paper in Section 2. In Section 3 we discuss annotations. In Section 4, we apply the results to develop an ontology for a comprehensive description of Deep Web sources wrt. the underlying domain ontologies. Section 5 shows how such descriptions are used to embed queries against Deep Web sources in MARS workflows. Section 6 discusses related work, and a conclusion follows in Section 7.

2 MARS: The Framework

The *MARS (Modular Active Rules for the Semantic Web)* Framework [9] provides an open framework for ECA (Event-Condition-Action) rules and for processes. The core of the MARS approach are a model and an architecture for ECA rules that use *heterogeneous* event, query, and action languages. In this paper, we consider one such language, the query language *DWQL (Deep Web Query Language)* that allows to pose queries against Deep Web sources. MARS is an open framework in the sense that arbitrary languages following this metaphor can be embedded; DWQL is such a language.

The MARS data flow through a rule or a process and to/from the processors of the constituents is based on sets of tuples of variable bindings in the style of deductive rules as illustrated in Figure 1. The state of the computation is represented by a set of tuples of variable bindings, i.e., every tuple is of the form $t = \{v_1/x_1, \ldots, v_n/x_n\}$ with $v_1, \ldots, v_n$ variables and $x_1, \ldots, x_n$ elements of the underlying domain (which is in our case the set of strings, numbers, and XML literals). Thus, for given variables $v_1, \ldots, v_n$, such a state can be seen as a relation whose attributes are the names of the variables.

Elements of constituent languages, such as DWQL queries, are represented in the MARS XML markup by elements of the form

```
<dwql:Query xmlns:dwql="http://www.semwebtech.org/languages/2008/dwql#">
  <dwql:view dwql:resource="identifying URI of the DWQL view" />
  <dwql:inputVariable name="x"  ... further annotations ... />
  <dwql:outputVariable name="y"  ... further annotations ... />
  further specification in DWQL markup as element content
</dwql:Query>
```

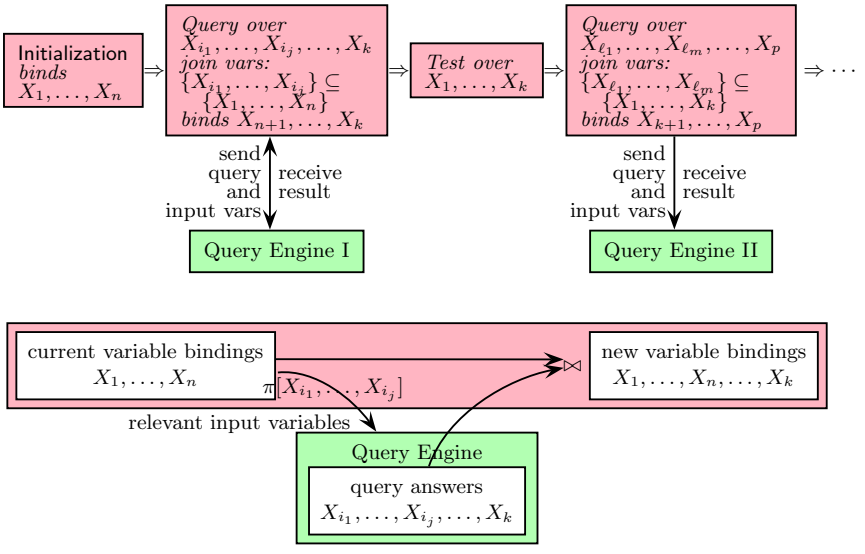


Fig. 1. Use of Variables in MARS

that contain the variable usage characteristics of the constituent. During execution, the selection of the actual services is done by a namespace-based infrastructure [5]. Based on the variable usage specification, only the relevant input variables are submitted together with the language fragment.

3 Literals, Measurements, Dimensions, and Units

Handling and combining queries against autonomous Deep Web sources requires some metadata knowledge about the values to be expected to deal with. Variables can be bound to literal values, and in RDF environments also to URIs (which are represented by strings, but represent the objects).

Schema. For handling values of variables programmatically, knowledge of the *datatypes* is mandatory. The datatypes are the same as in common programming languages and database systems; provided for the RDF world by the XML Schema simple types. In addition to the basic data types like strings and numbers, XML Schema provides `xsd:date`, `xsd:time` and `xsd:dateTime` datatypes similar to SQL. For actual usage, the syntactical representation by some format, e.g., "DD-MM-YYYY" can be specified.

Semantics. On the semantical level, the notion of *dimension* of a property is central: dimensions are e.g., the physical dimensions like *length*, *duration*, *voltage*, but also non-physical dimensions like *distance* (which is physically a length), or *price*. Values of dimensions are actually given by (value-unit)-pairs, like 100 km, or 250 €. Every dimension is associated with a set of units. In contrast to closed

applications, the units may differ between automonous sources (e.g., miles vs. kilometers, or $ vs. € or £); in such cases the conversion factors have to be known. Properties and also variables ranging over the values of a property are annotated with a dimension.

The MARS Ontology for Annotations. In MARS, processes and their constituents can be represented and annotated in RDF [13] and OWL [11]. A fragment of the ontology for dimensions, units, and conversions is shown below in Turtle [17] format. Some conversions are fixed (e.g., miles to kilometers), and some are dynamic, e.g., $ to €; for the latter, google (search e.g. for "100 USD in EUR") is used internally. Currencies are represented by fixed URIs such as <http://www.semwebtech.org/mars/currencies#EUR>.

```
@prefix : <http://www.semwebtech.org/mars#> .
@prefix dim: <http://www.semwebtech.org/mars/dimensions#> .
@prefix unit: <http://www.semwebtech.org/mars/units#> .
@prefix curr: <http://www.semwebtech.org/mars/currencies#> .
dim:Length a :Dimension;
  :hasUnits unit:meter, unit:kilometer, unit:mile, ... .
dim:Price a :Dimension;
  :hasUnits curr:USD, curr:EUR, curr:PLN, ... .
  owl:equivalentClass
  [ a owl:Restriction; owl:onProperty :hasUnits;
    owl:allValuesFrom :Currency] .
[ a :FixedConversion;
  :from unit:kilometer; :to unit:mile;  :factor 1609.3 ] .
[ a :DynamicConversion; :from curr:EUR; :to curr:USD ] .
[ a :DynamicConversion; :from curr:EUR; :to curr:PLN ] .
```

4 Deep Web Source Modeling

Deep Web sources can be considered on two levels: as data sources on the plain *database and (XML) Web level*, and wrt. their domain ontology on the *Semantic Web level*. We associate a precise source capacity description on both levels to each Deep Web source (which has to be done manually for each source).

Conceptually, every Web Data Source can be seen as an n-ary predicate $q(\overline{x}) = q(x_1, \ldots, x_n)$ (its *characteristic predicate*, which contains all input/output mappings). The first modeling step consists of *naming* the variables of this predicate by so-called *tags*. The different interaction patterns with the Web site (e.g. filling out forms, checkboxes, etc.) can be regarded as predefined views over the characteristic predicate. The modeling associates each view with a unique identifying URI (which is not the URL of the corresponding Web form, but "simply" some RDF URI) which is used (e.g. in MARS) for referring to that view. For each view v, its signature is specified in terms of one or more tags declared as input and output arguments. In the remainder of the paper we denote this signature as $\overline{out} \leftarrow v(\overline{in})$, where $\overline{in}$ and $\overline{out}$ are sets of tags. Each input argument corresponds

to an input element in the Web form, and each output argument corresponds to certain data records in the result page (cf. [7]).

Example 1 (Online Railway Schedule). *The online train schedules of railway companies are a typical example for Deep Web sources. Users can enter a start and a destination, a date and a desired departure or arrival time. The answer contains a list of relevant connections, usually together with prices.*

For the German Railway Web portal at `http://www.bahn.de`, *we tag the source with* start, dest, deptTime, arrTime, desiredDeptTime, desiredArrTime, date, duration, price. *The provided views have the signatures*

(deptTime, arrTime, duration, price) ←
germanRailwaysByDept(start, dest, date, desiredDeptTime) and
(deptTime, arrTime, duration, price) ←
germanRailwaysByArr(start, dest, date, desiredArrTime) .

A result of the first view looks as follows:

germanRailwaysByDept(
(start/"Freiburg", dest/"Göttingen", date/"03.02.2009", time/"08:00")) =
{ (deptTime/"08:57", arrTime/"13:07", duration/"4:10", price/"95.00"),
(deptTime/"09:03", arrTime/"14:48", duration/"5:45", price/"85.00"), ...}

Note that there is e.g. no view to retrieve all cities that can be reached from a given starting point within one hour traveling.

In set-oriented approaches like MARS, the input can consist of multiple tuples. For that, the answer tuples are always assumed to also contain the bindings of all input variables. With this, the results can be joined as shown in the lower part of Figure 1.

So far, there are only strings. Annotations are now made on the tag level, since the same annotations hold for each view over the source.

Datatypes and Units. According to Section 3, each tag is associated to some datatype, optionally to a specific syntactical representation, and a unit. For the specification of the format, MARS uses the one from Java's SimpleDateFormat.

Example 2 (Annotations to the Railway Source). *For the German Railways source, the tags are annotated as follows with datatypes, dimensions, syntactical representation (usually called* format*), and units.*

Tag	Datatype	Format	Unit
start, dest	xsd:string	–	–
deptTime, arrTime, desiredDeptTime, desiredArrTime	xsd:time	"HH:mm"	(internal)
duration	xsd:time	"HH:mm"	(internal)
date	xsd:date	"dd.MM.yyyy"	(internal)
price	xsd:decimal		curr:EUR

Source descriptions of DWQL wrappers for other railway portals have a similar signature, except probably the date format, and the currency: the source description of the analogue for Polish railways, `http://www.pkp.pl`*, differs only in the last entry – the unit of their price is Zloty, denoted by the URI* <*http://www.semwebtech.org/mars/currencies#curr:PLN*>*.*

Relationship with the Domain Ontology. The actual values, which are literals, have been described from the programming and data handling point of view above. From the semantical point of view, some of these literals, namely those that are only strings without dimension or datatype (in the above example: start and dest), denote entities of the according application domain.

Example 3 (Deep Web Source Description for German Railways). *In the railway example, the* start *and* end *tags are annotated to represent cities. The complete DWQL Source Description in RDF (N3) format is given in Figure 2. It lists the the tags used by the source, the views provided by the source, and for each view which tags are used in it as input or output. Note that the tag identifiers of the form "_:xxx" act only internally as identifiers. To the outside, only the tag names are known as illustrated by the SPARQL [15] query*

```
select ?U
where { ?S :providesView <bla://dwql-views/travel/germanRailwaysByDept> .
        ?S :hasTag [ :name "price"; :unit ?U ] }
```

that can be used to query the unit of the "price" slot of answers when retrieving connections by departure time. It yields the URI <*http://www.semwebtech.org/mars/currencies#EUR*>*. Such queries are used when the domains/units of variables of a process that contains a DWQL query are derived; as described in the next section.*

5 Embedding Deep Web Queries in MARS Processes

5.1 Annotation of Processes

The dataflow in MARS rules and processes is organized via tuples of variable bindings as depicted in Figure 1. The variables of MARS processes are optionally also annotated with a datatype and a dimension. This can be done automatically by analyzing the process and its variable usage if the subexpressions (here: DWQL queries) are accordingly annotated.

For annotation with units, either every single value can be annotated (which would require to store the unit as an additional column in the underlying database), or the variable is annotated once (usually based on the annotation of the source where the values originate from), and every value is transformed to that unit. For MARS, we chose the latter alternative.

Note that processes over homogeneous sources, e.g., which all use kilometers and Euro, work well even without explicit annotation. The annotation becomes important when the sources use different units.

```
 @prefix dim: <http://www.semwebtech.org/mars/dimensions#> .
 @prefix curr: <http://www.semwebtech.org/mars/currencies#> .
 @prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
 @prefix : <http://www.semwebtech.org/languages/2008/dwql#> .
 @prefix travel: <http://www.semwebtech.org/domains/2006/travel#> .

<bla://dwql-views/travel/germanRailways> a :DeepWebSource;
 :baseURL <http://www.bahn.de>;
 :providesView <bla://dwql-views/travel/germanRailwaysByDept>,
               <bla://dwql-views/travel/germanRailwaysByArr>;
 :hasTag _:start, _:dest, _:deptT, _:arrT,
         _:dDepT, _:dArrT, _:dur, _:date, _:price.

_:start a :Tag; :name "start"; :datatype xsd:string; :denotes travel:City.
_:dest  a :Tag; :name "dest"; :datatype xsd:string; :denotes travel:City.
_:deptT a :Tag; :name "deptTime"; :datatype xsd:time; :format "HH:mm".
_:arrT  a :Tag; :name "arrTime"; :datatype xsd:time; :format "HH:mm".
_:dDepT a :Tag; :name "desiredDeptTime"; :datatype xsd:time; :format "HH:mm".
_:dArrT a :Tag; :name "desiredArrTime"; :datatype xsd:time; :format "HH:mm".
_:dur   a :Tag; :name "duration"; :datatype xsd:time; :format "HH:mm".
_:date  a :Tag; :name "date"; :datatype xsd:date; :format "dd.MM.yyyy".
_:price a :Tag; :name "price"; :datatype xsd:decimal;
                               :dimension dim:price; :unit curr:EUR.

<bla://dwql-views/travel/germanRailwaysByDept> a :DeepWebView;
  :hasInputVariable _:start, _:dest, _:dDepT, _:date;
  :hasOutputVariable _:deptT, _:arrT, _:dur, _:price.
<bla://dwql-views/travel/germanRailwaysByArr> a :DeepWebView;
  :hasInputVariable _:start, _:dest, _:dArrT, _:date;
  :hasOutputVariable _:deptT, _:arrT, _:dur, _:price.
```

Fig. 2. DWQL Source Description for German Railways

Communication with Sources. As the annotations of the sources include the units that are required for the input variables, the values sent to the sources are converted accordingly (wrt. units, and also wrt. the syntactic representation, e.g., in case of time and date). As mentioned above, the returned values are also converted if required.

5.2 Embedding Deep Web Queries in MARS

The basic embedding pattern for queries has been shown in Section 2. For DWQL queries, the pattern has to be filled to contain all relevant information for communication with the DWQL service.

Usually, the variable names used in the MARS process do *not* coincide with the tag names of the DWQL views (in the same way as in programming, the variables in a method call do not coincide with the formal parameters of a method definition). As DWQL views are not positional (which wpuld mean that the arguments are ordered), but *slotted*, the pattern has to indicate how the

MARS variables are mapped to the view's variables/tags (and vice versa for the result variables).

The MARS processing model follows the idea of input and output variables for processing components of rules and processes. Thus, DWQL can be embedded in a homogeneous way. The *variable usage characteristics* of language components, i.e., a profile which variables are used, which have to be supplied as input (logic programming: negative use) and which can be bound by the evaluation of the component (logic programming: positive use) is contained in the process specification as illustrated in the example below.

5.3 Use Case: Combination of Queries against Railway Schedules

The use of the schedule of German Railways as a Deep Web source has been introduced in the above example. For international connections, for instance, from Freiburg to Poznan, the prices are not always returned. A suitable strategy is here to look up connections to the stations near to the border against the railway source in the origin country, and from these stations to the destination in the railway source of the destination country (and analogously for connections that run through three or more countries). Note that the necessity for conversion of prices naturally emerges in this situation.

We illustrate the approach using the above-mentioned connection from Freiburg to Poznan, using `http://www.bahn.de` for German Railways and `http://www.pkp.pl` for Polish Railways as Deep Web sources. The wrappers to both have been implemented based on [16,14].

The full query workflow can be specified in MARS/CCS [10,6] as shown in Figure 3 where we abstract from some parts that are not relevant for the Deep Web issues. The workflow is simplified such that it applies only to travels to direct neighbor countries. We also assume a data source that can be queried for the border stations for each pair of neighboring countries.

First, the variables start, startC, dest, destC, date, and time, are bound to their initial values, resulting in the single tuple

(start/"Freiburg", startC/"D", dest/"Poznan", destC/"PL",
date/"27.04.2009", time/"09:00") .

Then, the first query (actually evaluated against the travel database) binds the additional variable borderStation, depending on the values of destC.

In our case, there are three border stations known for traveling to Poland. Thus, three tuples are generated, namely

from	fromC	to	toC	borderStation	date	time
Freiburg	D	Poznan	PL	Szczecin	27.4.2009	09:00
Freiburg	D	Poznan	PL	Frankfurt(Oder)	27.4.2009	09:00
Freiburg	D	Poznan	PL	Görlitz	27.4.2009	09:00

With these tuples, the first DWQL query is evaluated. The tuples are projected and renamed according to the dwql:{input|output}Variable specifications (borderStation is used as dest) and the view germanRailwaysByDept is retrieved for the input tuples

{ (start/"Freiburg", dest/"Szczecin", date/"...", desiredDeptTime/"09:00"),
(start/"Freiburg", dest/"Frankfurt(Oder)", date/"...", desiredDeptTime/"09:00"),
(start/"Freiburg", dest/"Görlitz", date/"...", desiredDeptTime/"09:00") }

returning the following answer tuples:

{ (start/"Freiburg", dest/"Szczecin", date/"...", desiredDeptTime/"09:00",
deptTime/"09:49", arrTime/"18:48", duration/"8:59", price/"131.20"),
:
(start/"Freiburg", dest/"Frankfurt(Oder)", date/"...", desiredDeptTime/"09:00",
deptTime/"09:49", arrTime/"17:26", duration/"7:37", price/"127.00"),
(start/"Freiburg", dest/"Frankfurt(Oder)", date/"...", desiredDeptTime/"09:00",
deptTime/"09:49", arrTime/"17:30", duration/"7:41", price/"131.00"),

```
<ccs:Sequence xmlns:ccs="http://.../languages/2006/ccs#">
  assume variables start, startC, dest, destC, date, and time bound to initial values
  <ccsns:Query>
    binds variable borderStation by query hasBorderStation(startC, destC, borderStation)
  </ccsns:Query>
  <ccs:Query>
    <dwql:Query xmlns:dwql="http://.../languages/2008/dwql#">
      <dwql:view dwql:resource="bla://dwql-views/travel/germanRailwaysByDept" />
      <dwql:inputVariable dwql:name="start" dwql:use="start" />
      <dwql:inputVariable dwql:name="borderStation" dwql:use="dest" />
      <dwql:inputVariable dwql:name="date" dwql:use="date" />
      <dwql:inputVariable dwql:name="time" dwql:use="desiredDeptTime" />
      <dwql:outputVariable dwql:name="arrBorderTime" dwql:use="arrTime" />
      <dwql:outputVariable dwql:name="P1" dwql:use="price" />
    </dwql:Query>
  </ccs:Query>
  <ccs:Alternative>
    <ccs:Sequence>
      <ccs:Test> <ccs:Equals ccs:variable="destC" ccs:withValue="PL" /></ccs:Test>
      <ccs:Query>
        <dwql:Query xmlns:dwql="http://.../languages/2008/dwql#">
          <dwql:view dwql:resource="bla://dwql-views/travel/polishRailwaysByDept" />
          <dwql:inputVariable dwql:name="borderStation" dwql:use="start" />
          <dwql:inputVariable dwql:name="dest" dwql:use="dest" />
          <dwql:inputVariable dwql:name="date" dwql:use="date" />
          <dwql:inputVariable dwql:name="arrBorderTime" dwql:use="desiredDeptTime" />
          <dwql:outputVariable dwql:name="arrTime" dwql:use="arrTime" />
          <dwql:outputVariable dwql:name="P2" dwql:use="price" />
        </dwql:Query>
      </ccs:Query>
      calculate Price := P1 + P2
    </ccs:Sequence>
    similar <ccs:Sequence> specifications for other destination countries
  </ccs:Alternative>
</ccs:Sequence>
```

Fig. 3. Railway Connection Search as a CCS Sequence in XML Markup

:
(start/"Freiburg", dest/"Görlitz", date/"...", desiredDeptTime/"09:00",
deptTime/"10:57", arrTime/"19:27", duration/"8:30", price/"127.00"),
:
} .

Note that the result is just a set of tuples, not a set of groups of tuples and although the underlying interface does not support a set-oriented query interface, DWQL provides a set-oriented interface and iterates internally.

The tuples are then unrenamed (dest → borderStation (for joining), and new arrTime → arrBorderTime and price → P1). Then, the workflow enters the appropriate alternative for querying the railway company in the destination country. The Polish Railways page is wrapped to the same signature. For the input to query, the renaming is borderStation → start and arrBorderTime → desiredDeptTime. The query returns for each tuple the connecting trains from the respective border station to Poznan. The resulting tuples, amongst them

(start/"Frankfurt(Oder)", dest/"Poznan", date/"...", desiredDeptTime/"17:26",
deptTime/"17:33", arrTime/"19:27", duration/"1:54", price/"22.00")

are then unrenamed (start → borderStation, desiredDeptTime → arrBorderTime and price → P2) and joined with the before tuples (where the values of borderStation and arrBorderTime are the actual join condition). Finally, Price is obtained as P1 + P2, considering the different currencies as described below.

```
@prefix : <http://www.semwebtech.org/mars#> .
    ## further prefixes as in Figure 2
[ a :Process;
  useVariables _:start, _:startC, _:dest, _:destC,
       _:date, _:time, _:border, _:arrBT _:p1, _:p2, _:pr ].
_:start  a :Variable; :name "start"; :datatype xsd:string.
_:startC a :Variable; :name "start"; ## ... derived from the first query
_:dest   a :Variable; :name "dest"; :datatype xsd:string.
_:destC  a :Variable; :name "start"; ## ... derived from the first query
_:date   a :Variable; :name "date"; :datatype xsd:date;
         :format "dd.MM.yyyy".
_:border a :Variable; :name "borderStation"; :datatype xsd:string.
_:time   a :Variable; :name "time"; :datatype xsd:time; :format "HH:mm".
_:arrBT  a :Variable; :name "arrBorderTime"; :datatype xsd:time;
         :format "HH:mm".
_:arrT a :Variable; :name "arrTime"; :datatype xsd:time; :format "HH:mm".
_:p1  a :Variable; :name "P1"; :datatype xsd:decimal;
                           :dimension dim:price; :unit curr:EUR.
_:p2  a :Variable; :name "P2"; :datatype xsd:decimal;
                           :dimension dim:price; :unit curr:PLN.
_:pr  a :Variable; :name "price"; :datatype xsd:decimal;
                           :dimension dim:price; :unit curr:EUR.
```

Fig. 4. MARS Knowledge about the Railway Connection Process

5.4 Reasoning about Process Variables

As discussed in Section 3, variables in a MARS workflow are typed, including information about measurements and units. In the above example, the prices are typed and the required date formats are managed.

The MARS knowledge about the process is shown in Figure 4. It is derived completely from the process structure and the DWQL Source Descriptions. The derivation of the variables' properties is similar to *static typing* in programming languages. While most of the properties are straightforward, the prices deserve attention: P1 which is the answer from German Railways, is known to have the unit curr:EUR while P2 which is the answer from Polish Railways, has the unit curr:PLN. Price, which is derived as the sum of P1 + P2 gets also the unit curr:EUR. When computing Price := P1 + P2, for the above sample connection, P2 (e.g., 22 PLN) will be converted in 4.87 EUR before being added.

Additionally, some verification of the workflow's correctess (e.g., correct use of dimension-compatible answers) can be done based on the source annotations.

6 Related Work

Our work is related to the field of Semantic Web Services [4,2,8]. There, Web service descriptions are enhanced with semantic annotations mainly to facilitate automatic service composition [12].

In [1] an approach for annotating Web services is presented that allows to specify propositional and temporal constraints additionally to the regular input and output signature of a Web service. The constraints considered in their work are mandated by side-effects of the invocation of Web services, which is also the case for the Semantic Web service description proposals. Since our Deep Web sources are solely used for collecting information, these issues do not arise in our scenario. [3] presents a method for deriving query access plans for Deep Web sources. They describe the data sources as Datalog predicates with input and output characteristics, ranging over domain classes (i.e. movies). In our approach, we have a more detailed notion of domains, ranging from complex measurements with different syntactical representations to the possibility to use concepts of domain ontologies.

Finally, our work could benefit from complementary work on the analysis of query capabilities for deriving the data types and ranges of input arguments automatically [18,19].

7 Conclusion

For combining Deep Web data in a non-trivial way it is mandatory to assign a precise semantics to the input and output signature of the underlying source. We introduced a comprehensive formalism for annotating Deep Web sources semantically and showed how it is used for composition of different Deep Web services into a query workflow.

A prototype of the MARS framework can be found with sample processes and further documentation at http://www.semwebtech.org/mars/frontend/.

References

1. Beyer, D., Chakrabarti, A., Henzinger, T.A.: Web Service interfaces. In: WWW, pp. 148–159. ACM, New York (2005)
2. Burstein, M.H., Hobbs, J.R., Lassila, O., Martin, D.L., McDermott, D.V., McIlraith, S.A., Narayanan, S., Paolucci, M., Payne, T.R., Sycara, K.P.: DAML-S: Web Service description for the Semantic Web. In: Horrocks, I., Hendler, J. (eds.) ISWC 2002. LNCS, vol. 2342, pp. 348–363. Springer, Heidelberg (2002)
3. Calì, A., Martinenghi, D.: Querying Data under Access Limitations. In: ICDE, pp. 50–59. IEEE, Los Alamitos (2008)
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. Fritzen, O., May, W., Schenk, F.: Markup and Component Interoperability for Active Rules. In: Calvanese, D., Lausen, G. (eds.) RR 2008. LNCS, vol. 5341, pp. 197–204. Springer, Heidelberg (2008)
6. Hornung, T., May, W., Lausen, G.: Process algebra-based query workflows. In: CAiSE (to appear, 2009)
7. Hornung, T., Simon, K., Lausen, G.: Mashups over the Deep Web. In: WEBIST 2008. LNBIP, vol. 18, pp. 228–241. Springer, Heidelberg (2009)
8. Martin, D., Paolucci, M., Wagner, M.: Bringing semantic annotations to web services: OWL-S from the SAWSDL perspective. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 340–352. Springer, Heidelberg (2007)
9. May, W., Alferes, J.J., Amador, R.: Active rules in the Semantic Web: Dealing with language heterogeneity. In: Adi, A., Stoutenburg, S., Tabet, S. (eds.) RuleML 2005. LNCS, vol. 3791, pp. 30–44. Springer, Heidelberg (2005)
10. Milner, R.: Calculi for synchrony and asynchrony. Theoretical Computer Science, pp. 267–310 (1983)
11. OWL Web Ontology Language (2004), http://www.w3.org/TR/owl-features/
12. Rao, J., Su, X.: A survey of automated Web Service composition methods. In: Cardoso, J., Sheth, A.P. (eds.) SWSWPC 2004. LNCS, vol. 3387, pp. 43–54. Springer, Heidelberg (2004)
13. Resource Description Framework (RDF) (2000), http://www.w3.org/RDF
14. Simon, K., Lausen, G.: Viper: Augmenting automatic information extraction with visual perceptions. In: CIKM, pp. 381–388. ACM, New York (2005)
15. SPARQL Query Language for RDF, http://www.w3.org/TR/rdf-sparql-query/
16. Wang, Y., Hornung, T.: Deep Web Navigation by Example. In: BIS (Workshops), CEUR Workshop Proceedings 333, pp. 131–140. CEUR-WS.org (2008)
17. Turtle - Terse RDF Triple Language, http://www.dajobe.org/2004/01/turtle/
18. Wu, W., Yu, C.T., Doan, A., Meng, W.: An interactive clustering-based approach to integrating source query interfaces on the Deep Web. In: SIGMOD, pp. 95–106 (2004)
19. Zhang, Z., He, B., Chang, K.C.-C.: Understanding Web query interfaces: Best-effort parsing with hidden syntax. In: SIGMOD, pp. 107–118 (2004)

AKTB Workshop Chairs' Message

Virgilijus Sakalauskas and Dalia Kriksciuniene

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

The main goal of the workshop on Applications of Knowledge-Based Technologies in Business (AKTB) was to bring together researchers and practitioners, specialists and market analysts, to share their research expertise and advanced knowledge in modeling innovative solutions for enterprise systems and processes, analytic insights and experimental research results of designing and applying computational intelligence methods in various fields of business problems.

Workshop thematic areas were concentrated to solving complex tasks of contemporary business by applying intelligent and knowledge-based technologies expressed by these topics:

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

Total number of 25 articles was submitted to the AKTB workshop. Each paper was evaluated in the double-blind review process by at least two independent reviewers of the 18 members of the Program Committee. Each reviewer evaluated the quality of the article according to the criteria, including 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.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 51–52, 2009.

The highest ranked 12 articles were accepted for including into the post-conference proceedings and presenting during the conference. 13 articles were evaluated as not corresponding to the workshop themes or requirements. The acceptance rate for AKTB workshop was 0,48.

Statistics of AKTB workshop acceptance rate by number of countries and authors is presented in following table:

Country	*Authors*	*Submitted papers*	*Accepted*	*Acceptance rate*
Iran, Islamic Republic	2	1	1	1
Romania	1	1	0	0
United Arab Emirates	1	1	1	1
Ireland	4	2	1	0,5
France	2	2	1	0,5
Turkey	4	2	1	0,5
Lithuania	36	16	7	0,44

We would like to express our appreciation to all authors of submitted papers, members of the program committee, Department of Information Systems of the Poznan University of Economics, and the recognition of the outstanding efforts of the Organizing Committee of the 12th International conference on Business Information systems BIS2009.

Identification of Unexpected Behavior of an Automatic Teller Machine Using Principal Component Analysis Models

Rimvydas Simutis[1], Darius Dilijonas[1], and Lidija Bastina[2]

[1] Vilnius University, Kaunas Faculty of Humanities, Muitines 8, Kaunas, Lithuania
Rimvydas.Simutis@ktu.lt, Darius.Dilijonas@vukhf.lt
[2] JSC Penkių kontinentų bankinės technologijos, Kalvarijų 142, Vilnius, Lithuania
Lidija.B@5ci.lt

Abstract. Early detection of the unexpected behavior of the automatic teller machine (ATM) is crucial for efficient functioning of ATM networks. Because of the high service costs it is very expensive to employ human operators to supervise all ATMs in an ATM network. This paper proposes an automatic identification procedure based on PCA models to supervise continually the ATM networks. This automatic procedure allows detecting the unexpected behavior of the specific automatic teller machine in an ATM network. The proposed procedure has been tested using simulations studies and real experimental data. The simulation results and the first real tests show the efficiency of the proposed procedure. Currently the proposed identification procedure is being implemented in professional software for supervision and control of ATM networks.

Keywords: Automatic teller machine, principal component analysis, ATM network supervision, unexpected behavior.

1 Introduction

Automatic teller machines (ATMs) are computerized telecommunication devices which provide a financial institution's customers a method of financial transactions in a public space without the need for a human clerk. According the estimates developed by ATMIA (ATM Industry Association) the number of ATMs worldwide in 2007 was over 1.6 million. As the ATM networks expand it is very important the proper monitoring, supervision and cash management of the ATM networks [1, 2].

The crucial elements in development of efficient ATM network supervision and management system are creation of the cash demand forecasting models for every ATM and identification of unexpected behaviour of the ATMs in ATM network. The forecasting models have to be created based on historical cash demand data. The historical cash demand for every ATM varies with time and is often overlaid with non stationary behaviour of users and with additional factors, such as paydays, holidays, and seasonal demand of cash in a specific area. Cash drawings are subject to trends and generally follow weekly, monthly and annual cycles. The development of

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 53–61, 2009.

efficient cash demand forecasting models for ATMs we have introduced in earlier papers [3, 4]. Although these models generally can be used for detection of the outliers in ATMs' cash demand behaviour, they can't state the reason of these outliers. E.g., the wetter conditions can influence the cash demand of a specific ATM significantly, but this behaviour isn't anyhow connected with malfunctions of ATM or clients' illegal actions.

In this paper we propose a new computational procedure for identification of unexpected behaviour of an ATM in ATM network. The procedure is based on application of principal component analysis methods. The unexpected behaviour of an ATM can emerge from different reasons, e.g., it can be bundled with some rising obstacles in the ATM environment, with the operational problems of the ATM, or with clients' illegal actions. It is important to note, that for the identification of the unexpected behaviour of a specific ATM it is necessary to compare the ATM's behaviour with the behaviour of similar ATMs in the neighbourhood. If for some reasons (whether conditions, events in the region, etc.) disturbances are common for all ATMs in neighbourhood, then the changed behaviour of the specific ATM hasn't to be interpreted as unexpected. For the banking institutions it is crucial to identify the unexpected behaviour of an ATM as quick as possible and then act adequately to solve these problems timely. Because of the size of the ATM networks (some service institutions maintain ATM networks with over 1000 ATMs in network) human operators can't supervise efficiently the functioning of all ATMs. Therefore automatic procedures for detection of the unexpected behaviour of the ATMs have to be employed. This paper proposes a new solution for this task.

The paper is structured as follows. After short introduction of the problem in this section, the proposed identification procedure is introduced in section 2. In section 3, simulation studies using the proposed identification procedure are depicted and in section 4 practical tests are presented. Finally, the main results of this work are discussed in section 5.

2 Identification Procedure

To identify whether an ATM in ATM network shows an unexpected behaviour it is important to evaluate carefully the transactions prosecuted on the specific ATM together with the transactions prosecuted on the other ATMs with similar transactions' patterns. Based on this information the conclusions about the disturbances in behaviour of partial ATMs can be made. The proposed identification procedure consists of following steps:

a) Historical data of transactions (cash withdrawal) in ATM network have to be analyzed and clusters of the ATMs which similar behaviour must be formed. Each cluster includes specific number of ATMs. This number j can be defined by the user and in this applications was $j = 4 \div 5$;
b) For every ATM cluster a group of principal component analysis (PCA) models must be build. By development of the PCA models the historical data of ATM transactions are used. Inputs for PCA models are transactions data collected from ATMs cluster. Number of inputs for every model is $j-1$ and the

total number of PCA models for one ATM cluster is *j*. Each model uses combination of inputs which differs from inputs of other models;

c) If the new data point comes, PCA models should be used to estimate the squared prediction error (*SPE*) between the new sample and its projection into the *k* principal components. These estimations are carried out for all PCA models in ATM cluster. The *SPE* indicates how the transactions data of each ATM group conform to the designed PCA model for that ATM group;

d) If the *SPE* for the analysed group of ATM is bigger than the threshold value, then the conclusion about unexpected behaviour of ATM group is made. Advance analysis of information about the *SPE* in the other groups of the ATM cluster allows to identify the specific ATM showing the unexpected behaviour. This information is provided then to the ATM network operators.

The schema of the proposed identification procedure is presented in the Figure 1.

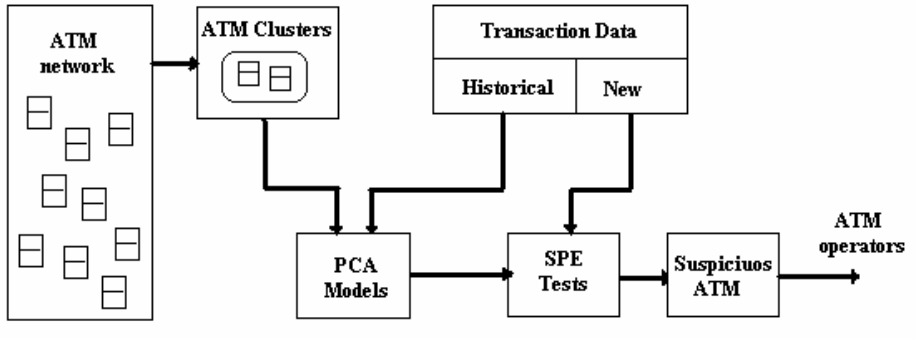


Fig. 1. Schema of the identification procedure for detection of the ATM with unexpected behaviour

In the first step of the procedure the ATM clusters have to been formed. Each cluster typically includes 4÷5 ATMs. The forming of the ATM clusters is based on correlation analysis of historical data. The ATMs with largest correlation coefficients join together in one cluster. In the second step one develops a group of principal component models for every ATM cluster. Principal component analysis is a technique for mapping multidimensional data into lower dimension with minimal loss of information and finding linear combination of the original variables with maximum variability [5,6]. PCA analysis has been extensively applied in various technical applications. Mathematically, PCA relies upon eigenvector decomposition of the covariance matrix of the original process variables. For a given data matrix $\boldsymbol{X}$ with m rows (data points, in our case - daily ATM's transactions) and n columns (variables, number of ATMs in ATM group) the covariance matrix of $\boldsymbol{X}$ is defined as:

$$\boldsymbol{R} = cov(\boldsymbol{X}) = \frac{\boldsymbol{X}^T\boldsymbol{X}}{m-1}\ , \tag{1}$$

where the columns of $\boldsymbol{X}$ have been scaled, i.e. the mean subtracted from each column and divided by the standard deviation. PCA decomposes the data matrix $\boldsymbol{X}$ into the sum of the outer product of so-called score vector t_i and so-called loading vector p_i with a residual error $\boldsymbol{E}$:

$$\boldsymbol{X} = t_1 p_1^T + t_2 p_2^T \ldots . + t_k p_k^T + \boldsymbol{E} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E}\,, \tag{2}$$

where k<m. The first principal component is that linear combination of the columns of $\boldsymbol{X}$ which describes the greatest amount of variability. In the *m*-dimensional space, p_1 defines the direction of the greatest variability, and t_1 represents the projection of each observation vector onto p_1. The second principal component explains the greatest amount of variability of the residual data. One can proceed in this manner until *k* principal components are obtained. If the variables in $\boldsymbol{X}$ are correlated, after calculating *k (k<m)* principal components most of the variation in the data set $\boldsymbol{X}$ has been explained. The score vector t_i contains information on how data points relate to each other. The loading vector p_i contains information how variables relate to each other. The columns of the loading matrix $\boldsymbol{P}$ are the eigenvectors corresponding to the *n* largest eigenvalues of the covariance matrix $\boldsymbol{R}$.

There are a number of methods that can be used to transform the input data matrix in score and loading vectors. In this case we used Non-linear Iterative Least Squares (NIPALS) method available within Mathworks's MATLAB software package [7].

In the proposed identification procedure the moving window historical data of ATMs transactions (cash withdrawal under normal operation conditions) were used to form the ATM clusters and the ATM groups. After that, the ATMs group matrix $\boldsymbol{X}$, covariance matrix $\boldsymbol{R}$, score matrix $\boldsymbol{T}$ and loading matrix $\boldsymbol{P}$ where determined. When an ATM cluster has *j* ATMs, then the number of inputs for every PCA model is *j-1* and total number of PCA models for one ATM cluster is *j*. Each PCA model in the ATM cluster uses combination of inputs (cash withdrawal from ATM) which differs from inputs of other models. Once the PCA models for ATM cluster are developed, new observation samples can be projected to the principal component space und the new ATMs data can be tested for possible disturbances and unexpected behaviour. For this purpose in the third step of the identification procedure the PCA models are used to estimate squared prediction error (*SPE*) between the new sample and its projection into the *k* principal components, also referred to as the Q statistic [8]. For the new observation vector x_{new} the *SPE* of the PCA model is estimated using equation

$$SPE = x_{new}\,(\boldsymbol{I} - \boldsymbol{P}_k \boldsymbol{P}_k^T) x_{new}^T\,, \tag{3}$$

where $\boldsymbol{P}_k$ is the matrix of the *k* loading vectors retained in the partial PCA model and $\boldsymbol{I}$ is the (*nxn*) identity matrix. The *SPE* indicates how well the new sample conforms to the PCA model, obtained with historical data. If the *SPE* of the analysed PCA model exceeds some threshold value (typical value - six squared values of standard deviation of the PCA model, developed with normal operation data), than the forth step of the identification procedure is activated. In this step it is necessary to make an advance analysis of all PCA models developed for one ATM cluster. For the reason that each PCA model has combination of inputs which differs from the inputs of other PCA models in the ATM cluster, it is possible to identify uniquely which input (ATM number) is responsible for the increased *SPE* value in PCA models.

Consequently the behaviour of this ATM is declared as unexpected and the ATM network operators are informed about this event.

3 Simulation Tests

To test the possibilities of the identification procedure to detect the unexpected behavior of the ATMs (unexpected changes in daily money withdrawal) a special simulation environment was created. An artificial ATM network with 100 ATMs was created and the daily money withdrawals from ATMs were simulated using weekly and monthly seasonality along with long term trends and special events (holiday effects). The simulation environment has imitated the daily money withdrawal from ATMs in typical ATM network in Lithuania. The simulation time was 500 days. Simulation tests and development of PCA models (subroutine *princomp*) were carried out in MATLAB programming environment. The ATM networks' simulation data were processed with correlation technique and according to the correlation coefficients the ATM clusters were formed. Typical money withdrawal patterns for one ATM cluster (4 ATMs) are presented in the Figure 2.

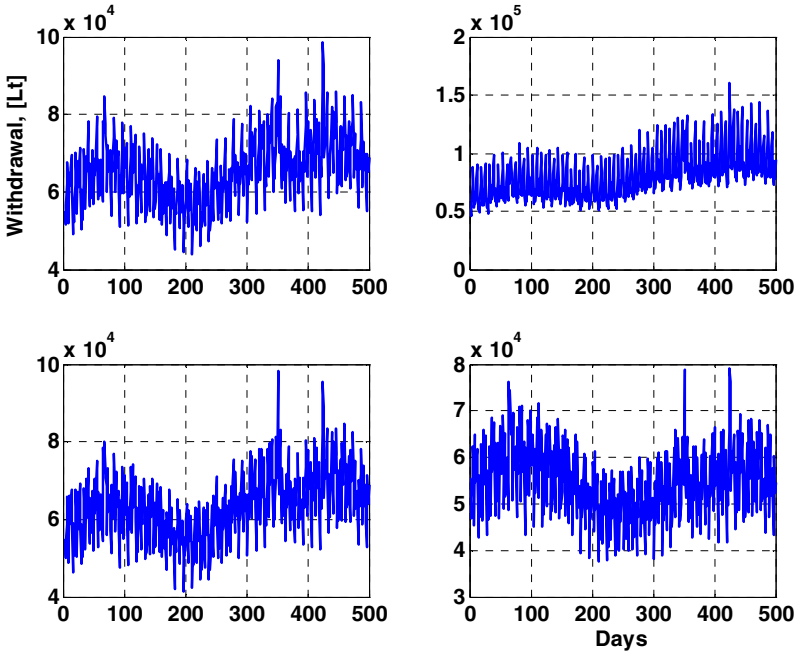


Fig. 2. Daily money withdrawal patterns for one ATM cluster (4 ATMs)

Then the identification procedure depicted above was carried out every day, based on the last 50-days moving window data. For every ATM cluster the four PCA models were build. Each PCA model has three inputs. They are scaled daily money withdrawal data of each ATM. Two principal components are used to describe the

variability of the process. After that, the developed PCA models were used to project the new next day observations to the principal component subspace and to estimate the squared prediction error (*SPE*) of the PCA model. The squared prediction errors of PCA models for one ATM cluster with normal operation conditions (without unexpected disturbances) are presented in the Figure 3.

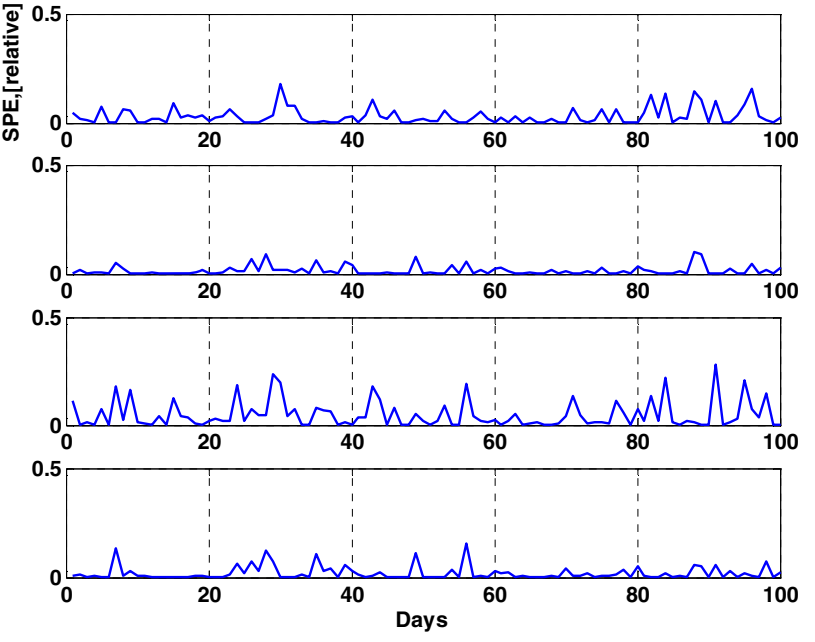


Fig. 3. Squared prediction errors of PCA models for one ATM cluster (4 ATMs) with normal operation conditions

For the same ATM cluster artificial disturbances (money withdrawal disturbances) were imitated. The work of every ATM was disturbed with additional money withdrawal equal to the average daily cash withdrawal. The first ATM was disturbed at *t=15*, second at *t=25*, third at *t=35* and forth at t= 45 days. The squared prediction errors of PCA models for this ATM cluster are presented in the Figure 4. If the *SPE* of the analysed PCA model exceeds the fixed threshold value (six squared values of standard deviation of developed PCA model in normal operation conditions) one can state that unexpected behaviour in the ATM group is observed. The next step is to identity the specific ATM responsible for this behaviour. Since every PCA model in ATM cluster has combination of inputs which differs from inputs of other PCA models it is easy to determine the ATM with unexpected behaviour. For example, in Figure 4, the *SPE* of the PCA models exceeds the threshold value for three ATM groups at time t=15 day. Only for the first ATM group *SPE* is normal at this time. It let to conclude that the unexpected behaviour shows the ATM which isn't included in this group. In this case it is the ATM with Number 1.

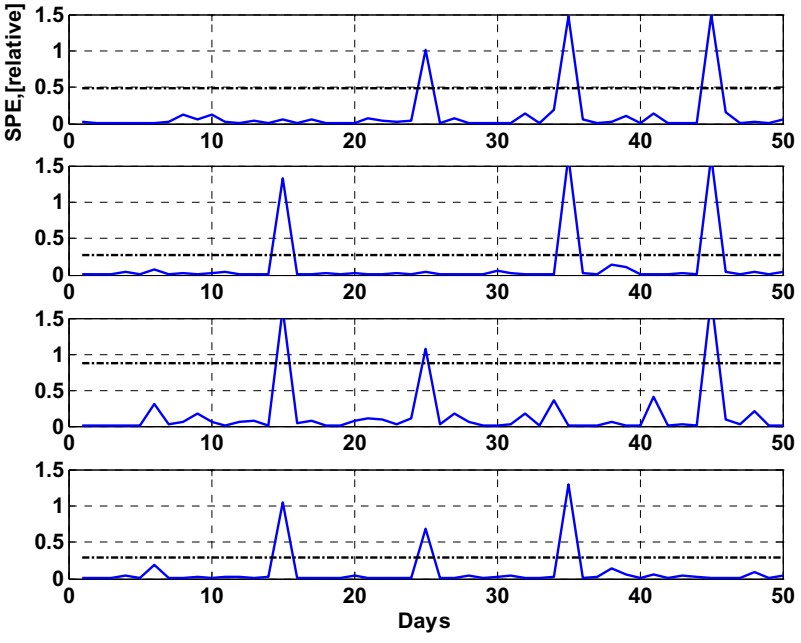


Fig. 4. Squared prediction errors of PCA models for ATM cluster (4 ATMs) with disturbed operation conditions. Dashed lines show the threshold values which are used to define the suspicious ATM groups.

In similar way we can identify easy the ATMs with unexpected behaviour at time t=25, t=35 and t=45 days. In these cases they are the ATMs with Numbers 2, 3 and 4. The simulation results shown in Figure 4 illustrate the efficiency of the proposed identification procedure. Similar results are obtained for the ATM clusters with other transaction patterns. These results confirm that the proposed procedure can be valuable in supervision of real ATM networks.

4 Test in Real ATM Network

The proposed identification procedure was partially tested with operating data from real ATM network. Because of the confidentiality of the problem only the basic information about the test is given here. Daily money withdrawals from 1900 ATMs in time range between 2-3 years have been analyzed and ATM clusters were build. Then the PCA models were developed using the ATM data collected in normal operation conditions for all ATM clusters. Later the PCA models were tested for operation conditions where big disturbances in the functioning of the ATMs had been occurred. The proposed identification procedure allowed detects the unexpected behavior of ATMs approximately in 80% of the real disturbed ATM cases. Figure 5 presents the typical *SPE* patterns by detection of unexpected behavior of

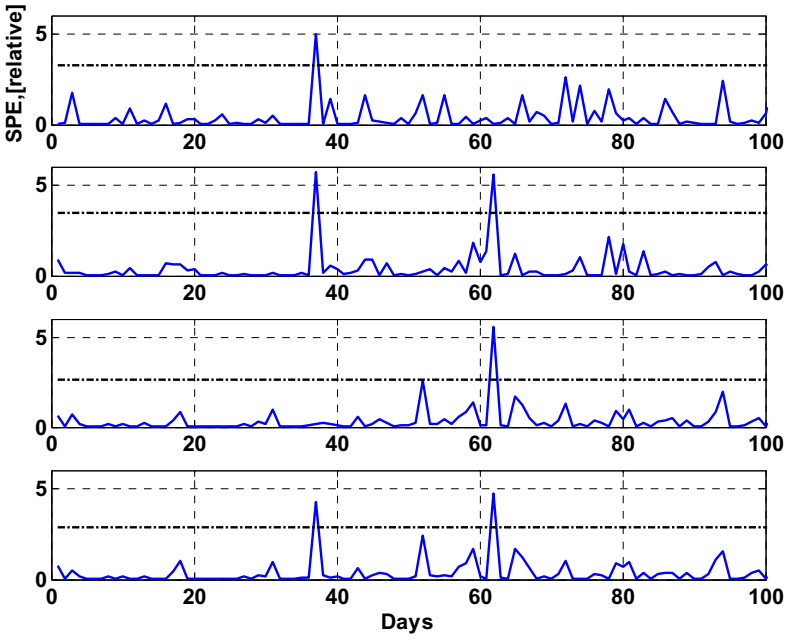


Fig. 5. Typical *SPE* pattern by detecting the unexpected behavior of ATM in ATM cluster. The data comes from a real ATM network. Unexpected behavior is detected for ATM Nr.3 (t=37) and for ATM Nr.1 (t=62).

ATM using PCA models for one ATM cluster. On the day t=37 the identification procedure detected unexpected ATM behavior for ATM Nr.3 and on the day t=62 unexpected behavior for ATM=1 (ATM Nr. 3 isn't included in ATM group 3, and ATM Nr.1 isn't included in ATM group 1). In both cases the detected behaviors match with real functioning disturbances at these ATMs. The performed real tests confirmed the efficiency of the proposed identification procedure. In the further research we will compare these tests with the results obtained using emerging data analysis technique - Exploratory Projection Pursuit (EPP) algorithms [9]. Currently the proposed procedure is being implemented in professional software for supervision and control of ATM networks.

5 Conclusions

Principal component analysis finds and eliminates linear correlation in the data. Here we analyze the possibilities of the application of the PCA models for supervision of ATM network. Early detection of the unexpected behavior of the ATM machines is crucial for efficient functioning of ATM networks. Because of the service costs it is very expensive to employ human operators to supervise continually the ATM network. This paper proposes an automatic identification procedure which is based on PCA models. This procedure allows detecting the unexpected behavior of the specific

automatic teller machine in an ATM network. The proposed procedure was tested using simulations tests and real experimental data. The simulation results and the first real tests showed that supervision of ATM network using PCA models is an efficient approach for identification of the unexpected behavior of the specific ATM. Currently the proposed identification procedure is being implemented in professional software for supervision and control of ATM networks.

Acknowledgments. This research work has been supported by means of EU structural funds (Project ASOMIS).

References

1. Snellman, H., Viren, M.: ATM networks and cash usage. Research Discussion Papers. Bank of Finland Nr. 21, 1–33 (2006)
2. Bounie, D., Francois, A.: Cash, Check or Bank Card: The effect of transaction characteristics on the sue of payment instruments. Telecom Paris, Working papers in Economics and Social Sciences, ESS-06-05 (2006)
3. Simutis, R., Dilijonas, D., Bastina, L., Friman, J., Drobinov, P.: Optimization of cash management for ATM network. Informacinės technologijos ir valdymas = Information technology and control 36(1A), 117–121 (2007)
4. Simutis, R., Dilijonas, D., Bastina, L.: Cash demand forecasting for ATM using neural networks and support vector regression algorithms. In: EurOPT 2008: the 20th International Conference Euro Mini Conference on Continuous Optimization and Knowledge-Based Technologies, Neringa, Lithuania, Vilnius, May 20-23, pp. 416–421 (2008) (selected papers), ISBN 978-9955-28-283-9
5. Roffel, B., Betlem, B.: Process Dynamics and Control. John Wiley & Sons, Ltd, Chichester (2006)
6. Jackson, J.E.: A User's Guide to Principal Components. John Wiley, Chichester (2003)
7. Nomikos, P., MacGregor, J.: Multivariate SPC Charts for Monitoring Batch Processes. Technometrics 37, 41–59 (1995)
8. Qin, S.J.: Statistical Process Monitoring: Basics and Beyond. Journal of Chemometrics 17, 480–502 (2003)
9. Friedman, J.H.: Exploratory projection pursuit. J. Am. Statist. Assoc. 82, 249–266 (1987)

Business Process Transformation Grid: An Empirical Model for Strategic Decision Making Towards IT Enabled Transformations

Dhrupad Mathur

Director(Industry Interface), S.P.Jain Center of Management, P.O. Box 502345, Block 5, Dubai International Academic City, Dubai, UAE
dhrupad.mathur@spjain.org

Abstract. The business process transformation grid postulated here is an outcome of empirical studies carried out in the areas of IT enabled transformations and e-business. It proposes a three dimensional view of a business system and creates an integration of desired momentum across these three axes. Most of the work on the business processes and e-business models is centered around developing the models capitalizing largely on the customer related processes which typically exposes the firm to the risk of having just a functional approach. Whereas, the process transformation grid focuses on three primary clusters of business processes and hence is more flexible and appropriate way of representing a business in its totality, as the approach is three-dimensional contrary to the prevalent approaches that due the lack of a structured strategic framework tend to become unidirectional.

Keywords: Transformation, e-Business, IT-enablement, Process, Grid.

It has always been a puzzle to for the decision-makers to initiate the Business Process Transformations and more so when they are IT-enabled. In spite of various available frameworks, it requites a lot of keen judgment and due diligence for one to figure out where to hold the organizational system from. At times, the outcome-centric frame of mind doesn't allow the possibility of exploring the dependencies thereby giving rise to unidirectional transformations of business processes. We all have heard about and have known transformations by different names: 'Corporate office initiative', 'Developers' view', 'their program', 'Consultants interest' etc. reflecting the perceived one-sidedness of such moves. While there is a lot of behavioral theory to be churned before one deserves to talk about the issues in totality, it is appropriate to mention that not many organization-wide transformations are even conceived holistically. This creates a partial movement of various process clusters existing in an organization without projecting or addressing their connections. At times by design but the symptomatic treatment to organizations ailments result in partial transformation creating process failures.

The business process transformation grid is an empirical approach towards determining the critical success factors for a business model. Every axis of a primary process shows the magnitude of the existing transformation, also corresponds to the degree of other factors which are critical for the success of an IT-enabled business.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 62–68, 2009.

This would help the managerial decision making and attaining a competitive advantage in many ways:

1. Re-visiting existing business model
2. Benchmarking IT-enabled business
3. Determining the aspiration level.
4. Determining transformation potential
5. Finally drawing a transformation roadmap integrating the three process areas.

Hence, making the IT-enabled transformation journey simpler.

Business process transformation grid proposed here has the potential of being developed as an effective empirical tool for managerial decision making in view of for business process transformations. This simple tool will help the managers to meditate on the nature of the business processes eventually leading to streamlined processes and inducing transformations.

Limitation: Since the grid considers various strategic elements and is dependent on the judgment of the manager for administering it, therefore the element of subjectivity can not be ruled out completely. However, it enables the manager to clearly demarcate the areas for transformation across three axes and prevents the pitfalls of isolated transformations. Number of inference-points in the grid can be developed.

1 Postulating the Business Process Transformation Grid

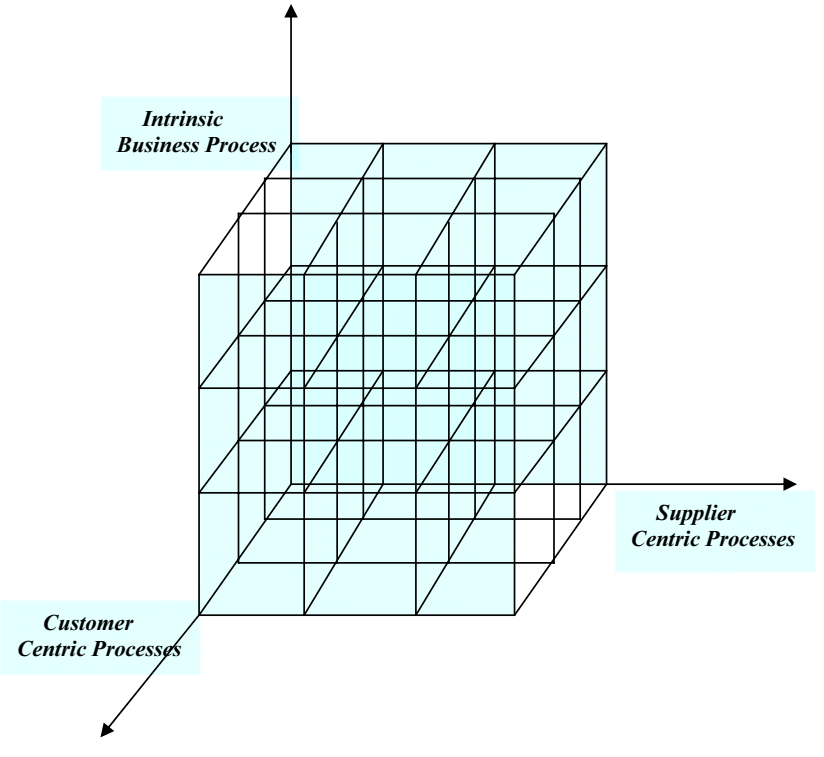


Fig. 1.

2 The Business Process Transformation Grid

Business Process Transformation Grid postulates 3 basic types of Business Processes, which are common to any e-business (or business for that matter):

1. Intrinsic Business processes (I). – Internal Business Processes and such clusters
2. Customer centric processes (C). – Demand-side Processes and such clusters
3. Supplier centric processes (S). – Supply side processes and such clusters

The observations and subsequent empirical analysis based on the secondary data and literature survey of e-Business entities highlights the fact that all the three processes together in different magnitudes give rise to a unique business model. Therefore the Business Process Transformation Grid is an empirical representation towards seeding, describing and classifying an e-Business model.

2.1 Significance and Scope of Business Process Transformation Grid

To achieve any kind of practicable business model, it is very essential to have a good understanding of the constituent components and processes. The business model ontology is the study of developing an incorporating framework, which suitably describes or gives a proper point of reference from where the e-business can be configured. This research therefore strives to achieve an empirical framework, which would be polymorphic in nature so as to take care of different dimensions where the business process exist and can be used for rightfully describing any kind of existing as well as forthcoming business model in any industry. One of the prime significance of this framework would be that contrary to the most popular approaches, which are centered on customer centric processes, it would help in understanding the other dimensions of e-business as well.

2.2 Characteristics of Business Process Transformation Grid

1. Transformation Volume

$$\text{Transformation Volume (T)} = I * C * S$$

Transformation Volume in its simplest sense is the amount of Business Process Transformation, which has already taken place. In order to determine the transformation volume, the extents of e-Business process transformation in a business entity has to be determined at a given point of time and considering the environmental factors to be nearly constant. (technology, strategy, innovation etc). Further, industry specific transformation benchmarks are developed similarly by identifying the best of the lot practices across industry for each axis to be considered as the aspiration level-benchmarks, against which the grid facilitates comparison.

It is proposed that a ***10-point scale index*** for the representation of each primary process axis be identified and the business model be graded on this scale. Therefore, the maximum Transformation, which can take place for a Business Model in a particular industry, can be:

Maximum Transformation Volume = T_{max}
Where, I = C = S = 10 (all maximum, benchmark case)
So, T_{max} = 10 x 10 x 10
T_{max} = 1000 (benchmark)
Benchmarks can either be existing or even futuristic.

Implications of Transformation Volume:

1. Transformation volume gives a 3 dimensional projection to the researcher or a manager as to which area is to be explored more for business process transformation.
2. It indicates the position of a business with respect to a particular industry therefore serving as a powerful empirical tool for business model development; as this relies on a benchmarking of the business under consideration with respect to the industry benchmarks.
3. It highlights the weaker axis of the e-business hence helps in identifying better integration areas of all the three forms of primary business processes.
4. It is proportional to firm's integration with Information & Communication Technology, which requires and enables higher levels of support of firm's robust IT infrastructure and customers' and suppliers' e-readiness.

2.3 Transformation Potential

Transformation Potential signifies, "what is to be done" in an e-business. It is the difference between the ***'desired state'*** and the ***'actual state'***. Therefore, any business which aspires to become an e-business has to look at the industry specific Transformation Potential, which indicates the total amount of transformation volume which is required at a given stage to achieve growth towards a desired e-business state.

Hence,

Transformation Potential (TP) = $T_{max} - (C * S * I)$
$TP = T_{max} - T$

Transformation potential is a relative term & has relationship with the extent of transformation achieved in a *particular industry*. Using the same ***10-point scale indexing method*** **f**or marking the Transformation Potential on a scale of 10 maximum, the difference between the desired and actual state is determined.

Therefore, Maximum Transformation Potential = TP max = 1000
Where, C = S = I = 0 (actual stage-startup for a particular business)
So, Transformation Volume = 0 x 0 x 0
T = 0 (minimum, as in a business start-up)

$TP_{max} = T_{max} - T$
$TP_{max} = 1000 - 0$
$TP_{max} = 1000$

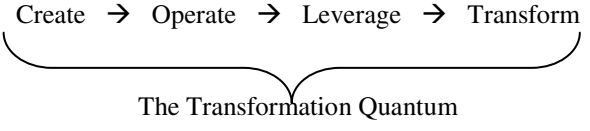
TP max would be case in those businesses that are purely traditional or are just using a fraction of e-business process as compared to the industry. The desired stage for a business would be, C = S = I = 10 (maximum) i.e. T_{max} equals to T thus,

$TP_{min} = 0$

The transformation volume of two models may be same yet there might be differences in the orientations on the three axes. But the over all impact of the volume remains same in terms of what amount of transformation has been or has to be achieved. However, this has to be studied in context of the three axes so as to get the clear picture of what is to be done tactically and operationally. So this gives broad strategic directions and strategic options but has to be administered systemically.

2.4 Transformation Quantum

The discussion on the business process transformation is incomplete until we understand how this transformation takes place and what is the microscopic constituent of the incremental transformation in the e-business. We will call it the Transformation Quantum.



The Transformation Quantum concept proposes that the transformation is an ongoing process. This happens because of the ongoing operations of a business, as an e-business model evolves and is translated to reality, various issues come out, which is a learning and knowledge gathering process. This knowledge is further leveraged to transform the e-business having a bearing on further model creation and upgradation. Hence, resulting in the next level of transformation. This can be visibly understood by the following figure as has also been illustrated by the IBM e-Business cycle.

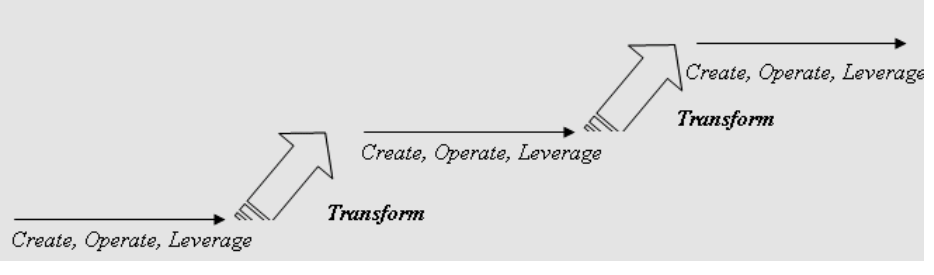


Fig. 2. Transformation Quantum

The above three concepts of Transformation Volume, Transformation Potential and Transformation quantum will help a management decision maker in developing a high-level transformation roadmap. Subsequently, giving rise to prioritization of process areas to be addressed through IT-enabled transformations.

Most of the work on the business processes and models is centered around developing the models on the basis of the customer related processes or demand side i.e. on the face of it having a functional approach. Whereas, this process transformation Grid focuses on three primary types of business processes and hence is more flexible and

appropriate way of representing a business in its totality, as the approach is three-dimensional contrary to the prevalent unidirectional approaches. Moreover, the Business Process Transformation Grid is also an empirical approach towards determining the Critical Success Factors of a business model. Every axis of a primary process shows the magnitude of the existing transformation, also corresponds to the degree of other factors which are critical for the success of a business. For example – If the Customer centric processes of a business are transformed to a greater extent, then it implies that the customer should have a certain level of e-readiness. So, the awareness and understanding on the customer's side become crucial for the e-business. After all how many customers are really keen and comfortable with various customized applets and supporting programs on Internet.

The Business Process Transformation Grid also highlights transformation from Bricks to Clicks i.e. from traditional process based business to an e-business. Yet again taking example of Dotcoms, which according to most researchers is the manifestation of e-business; is not the only entity, which makes up a business model. According to the process transformation grid, Dotcoms are those business entities that have their customer related processes transformed to a greater magnitude. So, only Dotcoms are not e-Businesses. Thus, we also have businesses, which are more transformed on the other two axes i.e., intrinsic business processes and suppler centric processes.

It has been greatly argued that the e-business model development is not an exact science and involves a good deal of judgment. However, the Business Process Transformation Grid proposed here forth can be applied as an effective empirical tool for managerial decision making when it comes to determining the IT- enabled transformation for any business model with respect to a specific industry.

E-Business strategy is about the uncertain future and therefore tends to be based on assumptions, premises and beliefs about customer priorities, technology evolution, competition and the core competencies that will be needed to compete. There are two types of E-Business strategy planning: top-down, analytic planning and bottom-up, "just do it" tactical planning. Top-down planning take a broad view of the environment, identifies options and then defines the organizations' mission and direction. A tactical operation takes a more focused or narrow view of the environment and performs the necessary activities required to produce short-term results.

(Kalakota, Robinson, 2000)

References

1. Harry, M.: Business Information: A systems approach. Prentice Hall Financial Times (2001)
2. Jupiter Media Matrix, What consumers buy on the web, Jupiter media (2003), http://www.jup.com
3. Kalakota, R.: E-Business road map for success. Addison Wesley, Reading (2000)
4. Kearney, A.T.: Line 56 E-Business Investment Benchmarking Study: August 2003, Line 56- AT Kearney joint research (2003), http://www.line56.com

5. McKenna, R.: Real Time, p. 52. Harvard Business School Press, Cambridge (1997)
6. Osterwalder, et al.: An Ontology for Developing e-Business Models. DSIage (2001)
7. David, P., Weber, Jonathan: "Clicks and Mortar", The Industry Standard (2000), http://thestandard.com/articles
8. Raphael, et al.: Value Creation In E-Business. Strategic Management Journal Strat. Mgmt. J. 22, 493–520 (2001), doi:10.1002/smj.187
9. UCLA, The UCLA Internet Report: Surveying the Digital Future Year Three, UCLA Center for Communication Policy (February 2003), http://www.ccp.ucla.edu
10. IBM Redbooks, http://www.redbooks.ibm.com
11. Zikmund, d.: Marketing Creating and keeping customers in an e-commerce world. South Western Thomson Learning (2001)

Research of the Calendar Effects in Stock Returns

Virgilijus Sakalauskas and Dalia Kriksciuniene

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

Abstract. In this article we investigate the problem of detection of the statistically significant dependences of stock trading return, which occur in particular days of the month and which could be important for creating profitable investment strategies. This problem is formulated as two hypotheses, stating that the stock trading return of the last five days of the month is greater than the average total monthly return, and the return generated over the first half of the month is significantly larger than that of the second half. By using the advanced methods of statistical analysis we researched the indications of these calendar effects for 24 stocks of the Vilnius stock exchange. The investigation did not fully confirm any of the hypotheses, but found out strong relation of risk level to the researched periods of the month. We explored the dependency of this effect to the volatility and volume of the traded stocks. The research results revealed that stocks of small and moderate volume have high volatility on the last days of the month, and the stocks of high volume have high volatility on the first part of month.

Keywords: calendar effect, F-test, mean return, Kolmogorov-Smirnov test, stock market.

1 Introduction

The profitablity of the financial markets is one of the most complicated scientific problems, which attract the attention of numerous researchers. The two main research directions include technical and fundamental analysis. The methods of technical analysis investigate influence of historical prices deviation and price shape regularity [1]. The supporters of the fundamental analysis concentrate attention to development of the financial indicators, which could evaluate stock price changes, and reveal the underlying reasons of the stock price fluctuations [2].

Any of these approaches can be given priority by their forecasting results, achieved by numerous researchers. The observed phenomena of price dynamics or their anomalies can be explained only by the integrative application of both methods.

One group of such phenomena is based of exploring various calendar effects, which could be employed for modelling profitable investment strategies and reducing risk of investment.

The biggest attention of the researchers is aimed at the influence of the day-of-the-week anomaly. Most researchers' state that on Mondays mean returns are lower, contrarily to Fridays, when the bigger returns are more likely, comparing to the other days of the week. [3-13].

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 69–78, 2009.

By analysing the mean return anomalies related to the monthly periods, researchers have notified several anomalies: the January effect, distinguished by the largest stock returns as compared to the returns of the other months; the turn-of-the-month effect, where the average return of the last days of the month is greater than the average total monthly return; the intramonth effect with the significantly larger mean return of the first half of a month than the mean return of the second half [14-17].

A great variety of the statistical analysis techniques have been employed by numerous researchers in order to detect such anomalies, and to use them for profitable investment strategies. The most prevalent of them are traditional statistical investigation methods, such as t-test, ANOVA, regression analysis or some advanced methods on influence of higher moments of mean return distribution [8-11]. In recent years the methods of discriminant analysis and artificial neural networks were applied for the identification of calendar anomalies [3,4].

A number of scientific papers [10,18-20] have disclosed that from 1990's calendar effect has faded away. Yet this tendency was not supported by the results obtained of the research of the young emerging stock markets with less available historical data and bigger fluctuations of the financial indicators. Aggarwal and Rivoli [21] studied four Asian emerging stock markets (Hong Kong, Singapore, Korea, Taiwan) and found that the day-of-the-week effect existed in all four markets. The extensive study of 21 emerging stock markets by Syed A. Basher and Perry Sadorsky [13], could not confirm the effect for all the markets, but most of them exhibited strong day-of-the-week effects even for the research model with the conditional market risk included. In [3] the research by the authors revealed that in the Vilnius Stock Exchange only approximately 30% of stocks are influenced by the day of the week effect.

In this work we examine the influence of turn-of-the-month and intramonth effects for the stocks traded in the Vilnius Stock Exchange [22], by analysing Index and return data of 24 most actively traded equities from the time interval 2003-01-01 to 2008-01-14. The computational methods include traditional statistical analysis, based on research of differences of mean return and standard deviation, calculated for the corresponding parts of the month, and the methods based on analysis of the higher moments. The following three main research tasks were aimed:

- to reveal the presence of turn-of-the-month and intramonth effects for various stocks traded in the Vilnius stock exchange,
- to analyse whether the strength of effect depends on daily turnover of the security and
- to investigate influence of the higher moments to the mean return distribution on this effect.

In the next section the organization of research data set is presented, and the investigation methodology is defined. Then in section 3 the research results of the application of traditional analysis methods, such as t-test, one-way ANOVA, Levene and Brown-Forsythe test of homogeneity of variances, are analysed. Also this section covers the results of application of nonparametric statistics. The significance of the turn-of-the-month and intramonth effect is evaluated by using tests of Kolmogorov-Smirnov and Mann-Whitney U. The research outcomes and conclusions are covered in Section 4.

2 Data and Methodology

The data for empirical research was driven from information base of Vilnius Stock Exchange which belongs to NASDAQ OMX Group (NASDAQ OMX, 2009 [20]). The NASDAQ OMX stock exchanges in Tallinn, Riga and Vilnius compound the Baltic Market for implementing the core idea to minimize to the possible extent the differences between the three Baltic markets, facilitate cross-border trading, and to attract more investments to the region [22]. Vilnius Stock Exchange belongs to the category of small emerging securities markets as it is expressed by its main financial indicators: market value of 7 billion EUR, share trading value- near 2 million EUR per business day, number of trade transactions per business day of approximately 600, the total equity list consisting of 44 shares.

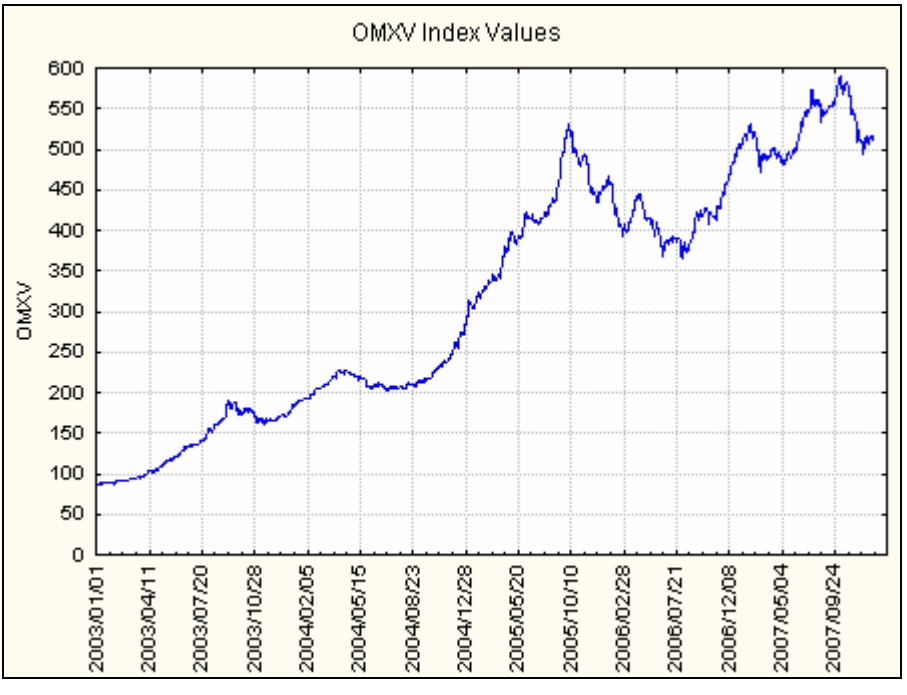


Fig. 1. OMXV Index Values from 2003.01.01-2008.01.14

The Vilnius Stock Exchange is mirrored by the OMX Vilnius Stock Index, which is a capitalization weighted chain linked total-return index. It is calculated on a continuous basis using the most recent prices of all shares that are listed on the Vilnius Stock Exchange. The dynamics of the stock prices during the period of analysis can be summarized by the profile of OMX Vilnius Stock Index (Fig.1), where big variety of economical situations is well reflected. In Fig.1 the period of 2003.01 to 2005.10 can be characterized by the increase of average stock prices, which went up to about 5 times with moderate fluctuations. Starting from 2005.10 till 2006.08 was the period of quite harsh decrease, followed by significant increase of price level during the whole

following year. At the same time the price volatility increased as well. The stock market price crisis of the end of 2007 had major influence on theVilnius Stock Exchange in Lithuania. The recent period of more stable stock price level still shows quite big price fluctuations.

To identify the presence of turn-of-the-month and intramonth effects we used the data of daily return values of actively traded 24 shares (out of 44 listed), from the period of 2003-01-01 till 2008-01-14. The selected shares represent the variety of the Vilnius Stock Exchange equity list by capitalization, daily turnover, trade volume, return and risk. The selected time span of historical development for designing the database of the stock trading records substantiates the validity of the experimental research, as it covers sufficient amount of financial data and wide variety of real trading situations in the analysed financial market.

In the paper the logarithmic understanding of return is applied $R_t = \ln(P_t / P_{t-1})$, where return R_t of time moment *t,* is evaluated by logarithmic difference of stock price over time interval [t-1,t]. P_t indicates stock price at time moment *t*.

The return values of 24 equities were assigned to the variables, named correspondingly to their symbolic notation of Vilnius Stock Exchange.

The data cleansing procedures of the stock information time series included removal of non-trading records during the holidays or weekends, and the records of the trading days with zero number of deals. After processing the data set, the average number of daily trading records for each share was approximately 1100, thus ensuring the necessary amount of experimental data for getting significant findings. For mining the data and calculations we used STATISTICA and MS EXCEL software [23].

We prepared two sets of stock trading data. The first list (I) is used for the research of turn-of-the-month effect, and it contains data of each stock, assigned to two variables. The computed values of average daily return of the starting 25 days of the month are assigned to the first variable, and the values of average daily return of the last five days of the month make the second variable. The other list (II) is designed in a similar way for validation of the intramonth effect. The first variable of this list denotes the average daily return computed for the first half of the month and the second variable – mean value computed for the last half of the month. The corresponding variables of the first list will be marked by the symbolic notation of the stock, followed by 1, and adding the 2 for the variables of the list II. E.g. Index1_2 denotes average daily return of the OMXV index values during last 5 days of the month, and the TEO2_1 denotes average daily return of the first half of the month, calculated for the stocks of TEO company.

Our goal was to check, if the turn-of-the-month and intramonth effects had influence on the profitability and volatility of stocks, and if these effects could be related to the trading volume of stocks, by observing the occurrence of the effects in groups, formed of stocks according to different trading volumes.

The difference in trading volume of the stocks was quite evident, as the daily turnover of the stocks LEL and KBL was up to 15 thousand LTL (1 EUR=3.45 LTL), and the TEO stock trading turnover reached 700 thousand LTL. Therefore the set of stocks was sorted, and arranged into three groups according to their daily turnover data (Table 1).

The initial analysis by summary statistics of the data set revealed quite big differences of average daily volatility among the variables of the list I, as contrarily to list II, where the differences among values of mean return and the standard deviation were insignificant for both parts of the month (Table1).

Table 1. The summary statistics of occurrence of turn-of-the-month and intramonth effects

Variable	Descriptive Statistics							
	I part Mean	II part Mean	I part Std.D.	II part Std.D.	I half Mean	II half Mean	I half Std.D.	II half Std.D.
Index	0,122	0,097	0,258	0,731	0,149	0,089	0,316	0,334
LEL	0,249	-0,130	0,700	1,540	0,289	0,147	1,029	0,953
KBL	0,032	-0,394	0,798	2,716	-0,043	-0,235	2,121	1,229
LNS	0,003	-0,021	1,000	2,340	-0,156	0,068	1,246	0,963
LEN	0,118	-0,123	0,588	1,757	0,040	0,191	0,620	0,885
VBL	0,175	-0,061	0,697	2,105	0,235	-0,058	0,940	0,957
LJL	0,120	0,042	0,490	1,990	0,084	0,145	0,604	0,743
LLK	0,104	0,810	0,768	2,318	0,005	0,272	2,042	1,730
Average I	0,115	0,017	0,720	2,109	0,065	0,076	1,229	1,066
KJK	0,270	0,349	1,349	2,641	0,097	0,415	2,161	2,212
ZMP	0,112	-0,327	0,680	1,297	0,070	0,081	0,840	1,054
LDJ	0,087	-0,099	0,453	1,212	0,025	0,094	0,595	0,514
RST	0,132	0,045	0,456	2,339	0,228	0,036	0,635	0,746
UTR	-0,205	0,152	0,817	2,025	-0,031	-0,128	0,773	1,057
NDL	0,504	-0,499	2,261	3,381	0,032	0,437	3,279	2,318
SAN	0,163	0,248	2,557	1,471	0,426	-0,000	1,795	2,682
KNF	0,020	-0,054	0,472	0,923	0,020	-0,018	0,543	0,622
PTR	0,440	0,482	1,284	3,278	0,765	0,425	1,865	2,159
PZV	0,159	0,071	0,456	1,227	0,157	0,118	0,554	0,461
Average II	0,168	0,037	1,079	1,979	0,179	0,146	1,304	1,382
APG	0,185	0,363	1,592	1,649	0,105	0,579	2,782	1,571
MNF	0,234	0,262	0,560	1,455	0,256	0,169	0,978	0,684
SNG	-0,154	0,154	1,326	1,384	-0,609	-0,043	5,127	0,604
UKB	-0,261	0,212	3,983	2,440	-0,869	0,226	8,318	1,204
RSU	-0,005	-0,249	1,139	1,081	-0,365	0,030	3,746	0,585
LFO	0,535	-0,218	2,705	2,461	0,446	0,462	3,593	2,385
TEO	0,081	0,065	0,392	0,893	0,090	0,061	0,419	0,575
Average II	0,088	0,084	1,671	1,623	-0,135	0,212	3,566	1,087

The results mean that the turn-of-the-month effect does not directly influence neither OMXV index, nor mean return values of separate stocks. But it is quite obvious that average volatility of the last days of the month is much higher, as compared to the starting 25 days of the month. This observation is not so evident for the stocks with

high trading volume, where there was only slight difference among the mean values of the standard deviation calculated for the last days and the remaining part of the month. The analysis of the list II variables showed that the trading risk on the first half of the month comparing to the second half was evidently more high for the stocks with the highest trading volume (Table 1). Those outcomes confirmed the general observation that the biggest speculative trading transactions in the Vilnius Stock Exchange were performed for the stocks with the highest turnover and best liquidity, and they tend to occur in the first half of the month.

The main research tasks and the influence of the initial observations are further discussed and analysed by applying methods of the statistical analysis.

3 Statistical Investigations

The analysis of the impact of the turn-of-the-month and the intramonth effects included the application of the traditional statistical methods, such as t-test, one-way ANOVA, Levene and Brown-Forsythe test of homogeneity of variances. Further the nonparametric Kolmogorov-Smirnov test, used for analysis of impact of the higher moments of return distribution was applied.

The initial analysis by the application of the t-test was used to check presence of the turn-of-the-month and intramonth effects for daily mean return. As it was already hinted by the results displayed in Table 1, where there was only a very slight difference among the values of average mean return, the t-test did not denote any significant results for any of the 24 stocks from the data set and for any of the explored effects. Only one stock (ZMP), was marked for the possible impact of turn-of-the-month effect with the significance level $p=0.0125$. The possibility to apply this analysis criteria was checked by the Shapiro-Wilk W test for normality.

The following step of the research was to analyse the impact of the turn-of-the-month and intramonth effects for the volatility of the variables. The summary table of F-test results indicating significance of the differences of standard variation is presented in Table 2. In the Table 2 the occurrence of the significant effect is marked in bold, therefore we can summarize that the turn-of-the-month effect has impact for all the stock from the groups of low and medium trading volume, but only for the few stocks which belong to the group of high volume.

The opposite results were revealed by analysing the intramonth effect. The significant difference of volatility was indicated only for the group of high volume stocks. Only few stocks of lower volume group were marked as affected by this calendar anomaly, according to the higher variance in the first half of the month (the stocks KBL, LEN, UTR and SAN).

How can changes in volatility and the periods of the month be interrelated, it is not clearly evident merely by data analysis. There could be a psychological interpretation of the investors' behaviour which is supported by the insight that during the starting part of the month there is a prevailing tendency to invest to less risky stocks with sufficient liquidity, which further cause such big difference in the standard variance of the first and the second parts of the month. Nevertheless this insight can explain the visibility of the turn-of-the-month effect for stocks which have higher risk, low volume and high volatility at the same time. On the turn of the month, their popularity among traders together with their level of risk raises significantly (Table 2).

Table 2. The turn-of-the-month and intramonth effects for variance

	F-test for Variance Marked tests are significant at p <,05000				
Part 1 vs. Part 2	F-ratio Varian	p Varian	Half 1 vs. Half 2	F-ratio Varian	p Varian
Index1_1 vs. Index1_2	**8,047**	**0,000**	Index2_1 vs. Index2_2	1,116	0,675
LEL1_1 vs. LEL1_2	**4,839**	**0,000**	LEL2_1 vs. LEL2_2	1,166	0,561
KBL1_1 vs. KBL1_2	**11,569**	**0,000**	KBL2_1 vs. KBL2_2	**2,978**	**0,000**
LNS1_1 vs. LNS1_2	**5,475**	**0,000**	LNS2_1 vs. LNS2_2	1,675	0,052
LEN1_1 vs. LEN1_2	**8,938**	**0,000**	LEN2_1 vs. LEN2_2	**2,037**	**0,008**
VBL1_1 vs. VBL1_2	**9,127**	**0,000**	VBL2_1 vs. VBL2_2	1,038	0,889
LJL1_1 vs. LJL1_2	**16,476**	**0,000**	LJL2_1 vs. LJL2_2	1,513	0,117
LLK1_1 vs. LLK1_2	**9,121**	**0,000**	LLK2_1 vs. LLK2_2	1,393	0,222
KJK1_1 vs. KJK1_2	**3,832**	**0,000**	KJK2_1 vs. KJK2_2	1,048	0,866
ZMP1_1 vs. ZMP1_2	**3,637**	**0,000**	ZMP2_1 vs. ZMP2_2	1,575	0,087
LDJ1_1 vs. LDJ1_2	**7,153**	**0,000**	LDJ2_1 vs. LDJ2_2	1,337	0,272
RST1_1 vs. RST1_2	**26,305**	**0,000**	RST2_1 vs. RST2_2	1,380	0,223
UTR1_1 vs. UTR1_2	**6,147**	**0,000**	UTR2_1 vs. UTR2_2	**1,867**	**0,035**
NDL1_1 vs. NDL1_2	**2,235**	**0,018**	NDL2_1 vs. NDL2_2	2,002	0,016
SAN1_1 vs. SAN1_2	**3,021**	**0,000**	SAN2_1 vs. SAN2_2	**2,231**	**0,003**
KNF1_1 vs. KNF1_2	**3,816**	**0,000**	KNF2_1 vs. KNF2_2	1,311	0,351
PTR1_1 vs. PTR1_2	**6,514**	**0,000**	PTR2_1 vs. PTR2_2	1,340	0,299
PZV1_1 vs. PZV1_2	**7,236**	**0,000**	PZV2_1 vs. PZV2_2	1,443	0,165
APG1_1 vs. APG1_2	1,073	0,800	APG2_1 vs. APG2_2	**3,136**	**0,000**
MNF1_F vs. MNF1-L	**6,748**	**0,000**	MNF2_F vs. MNF2_L	**2,046**	**0,010**
SNG1_1 vs. SNG1_2	1,089	0,750	SNG2_1 vs. SNG2_2	**72,169**	**0,000**
UKB1_1 vs. UKB1_2	**2,664**	**0,000**	UKB2_1 vs. UKB2_2	**47,707**	**0,000**
RSU1_1 vs. RSU1_2	1,112	0,694	RSU2_1 vs. RSU2_2	**40,957**	**0,000**
LFO1_1 vs. LFO1_2	1,208	0,529	LFO2_1 vs. LFO2_2	**2,269**	**0,004**
TEO1_1 vs. TEO1_2	**5,183**	**0,000**	TEO2_1 vs. TEO2_2	**1,884**	**0,017**

In order to confirm validity of the obtained results, an important requirement that the variances in the different groups are equal (homogeneous) were tested. For this research two powerful and most commonly used tests for exploring this assumption were applied: Levene test and Brown-Forsythe modification of this test. The latter performs the analysis on the deviations of the group medians, instead of means as in Levene test. The results of using Levene test and the application of the Brown-Forsythe test gave equivalent results: the hypothesis of homogeneity could not be rejected for none of the variables [23].

The significance of the turn-of-the-month and intramonth effects to the mean return of investment was explored by applying two nonparametric tests. Kolmogorov-Smirnov test was used to verify the hypothesis, if two samples were drawn from the

same population. The Mann-Whitney U test was used to explore the location characteristics of two samples. The Kolmogorov-Smirnov test is generally applied for testing the influence of higher moments for the distribution [23], and it is sensitive to the differences of the general shapes of the distributions of the two samples [9] (expressed by differences of dispersion, skewness, kurtosis etc.).

Table 3. Results of application of Kolmogorov-Smirnov tests for the turn-of-the-month and intramonth effects

	Kolmogorov-Smirnov Test Marked tests are significant at p <,05000						
Part 1 vs. Part 2	Neg Differ.	Pos Differ.	p-level	Half 1 vs. Half 2	Neg Differ.	Pos Differ.	p-level
Index1_1/Ind1_2	**-0,283**	**0,167**	**p < .02**	Index2_1/Ind2_2	-0,050	0,167	p > .10
LEL1_1/LEL1_2	**-0,102**	**0,271**	**p < .05**	LEL2_1/LEL2_2	-0,051	0,136	p > .10
KBL1_1/KBL1_2	-0,145	0,200	p > .10	KBL2_1/KBL2_2	-0,068	0,169	p > .10
LNS1_1/LNS1_2	**-0,121**	**0,276**	**p < .02**	LNS2_1/LNS2_2	-0,153	0,068	p > .10
LEN1_1/LEN1_2	**-0,153**	**0,305**	**p < .01**	LEN2_1/ LEN2_2	-0,169	0,136	p > .10
VBL1_1/VBL1_2	-0,208	0,226	p > .10	VBL2_1/VBL2_2	0,000	0,224	p > .10
LJL1_1/LJL1_2	-0,121	0,224	p > .10	LJL2_1/LJL2_2	-0,085	0,068	p > .10
LLK1_1/LLK1_2	**-0,362**	**0,170**	**p < .05**	LLK2_1/ LLK2_2	-0,107	0,125	p > .10
KJK1_1/KJK1_2	-0,130	0,109	p > .10	KJK2_1/KJK2_2	-0,130	0,074	p > .10
ZMP1_1/ZMP1_2	-0,051	0,237	p < .10	ZMP2_1/ZMP2_2	-0,102	0,119	p > .10
LDJ1_1/LDJ1_2	-0,086	0,224	p > .10	LDJ2_1/LDJ2_2	-0,136	0,034	p > .10
RST1_1/RST1_2	-0,190	0,138	p > .10	RST2_1/RST2_2	-0,034	0,203	p > .10
UTR1_1/UTR1_2	-0,277	0,085	p < .10	UTR2_1/UTR2_2	-0,063	0,208	p > .10
NDL1_1/NDL1_2	-0,081	0,243	p > .10	NDL2_1/ NDL2_2	-0,118	0,059	p > .10
SAN1_1/SAN1_2	-0,056	0,167	p > .10	SAN2_1/SAN2_2	-0,136	0,068	p > .10
KNF1_1/KNF1_2	-0,120	0,200	p > .10	KNF2_1/KNF2_2	-0,061	0,122	p > .10
PTR1_1/PTR1_2	-0,102	0,102	p > .10	PTR2_1/PTR2_2	-0,058	0,154	p > .10
PZV1_1/PZV1_2	-0,143	0,187	p > .10	PZV2_1/PZV2_2	-0,085	0,085	p > .10
APG1_1/APG1_2	-0,113	0,208	p > .10	APG2_1/APG2_2	-0,193	0,035	p > .10
MNF1_1/MNF1-L2	-0,189	0,133	p > .10	MNF2_1/MNF2_2	-0,091	0,164	p > .10
SNG1_1/SNG1_2	-0,155	0,103	p > .10	SNG2_1/SNG2_2	-0,051	0,186	p > .10
UKB1_1/UKB1_2	-0,204	0,130	p > .10	UKB2_1/UKB2_2	-0,086	0,155	p > .10
RSU1_1/RSU1_2	-0,053	0,175	p > .10	RSU2_1/RSU2_2	-0,034	0,153	p > .10
LFO1_1/LFO1_2	-0,043	0,174	p > .10	LFO2_1/LFO2_2	-0,196	0,039	p > .10
TEO1_1/TEO1_2	-0,153	0,186	p > .10	TEO2_1/TEO2_2	-0,102	0,102	p > .10

In our case Kolmogorov-Smirnov test was applied to reveal the impact of the turn-of-the-month and intramonth effects for the stocks belonging to the groups with different trading turnovers. The results of the investigation are presented in Table 3.

As we can see from Table 3, only the variable Index and the stocks LEL, LNS, LEN, and LLK (marked in bold) indicated the presence of the turn-of-the-month effect.

Therefore we can reject the hypothesis of the equal mean return distributions only for those particular stocks. The main characteristic of these stocks is their very low trading volume. By applying Mann-Whitney U test for exploring the location characteristics of the two samples (means, average ranks, respectively) we observed significant results for the same stocks.

4 Conclusions

In this research the two types of calendar effects, such as turn-of-the-month and intramonth, were explored for stocks drawn from the Vilnius Stock Exchange. The research methods were selected in order to disclose if these anomalies affect certain stocks, and to explore their dependency from trading volume and volatility. The research outcomes could confirm the hypotheses of the presence of the turn-of-the-month and intramonth effects only by their impact on the volatility of stocks. We can state that the direct impact of these effects to the mean return indicator of the stocks was not visible and could not be confirmed by any applied methods.

The strongest relationship of the explored effects to the trading volume could be explained only by changes in volatility. It was detected that the turn-of-the-month effect for the volatility of the variables was significant for all stocks which belong to the group of low and medium turnover. Only few stocks from the high trading volume group were marked as affected by these calendar anomalies. The analysis of the intramonth effect gave opposite results, as the significant difference of volatility could be observed only for high volume stocks. Only four stocks of the low volume group had significantly higher variance in the first half of the month.

The analysis of the impact of the higher moments to the calendar effects for the mean return was performed by applying Kolmogorov-Smirnov test. Only several stocks (LEL, LNS, LEN, and LLK) showed positive impact of the turn-of-the-month effect to the mean return.

The effectiveness level of the Vilnius stock exchange market is quite high, therefore the effective trading strategies can be based only on quite sensitive methods, able to discover various trading anomalies. The research of the turn-of-the-month and intramonth effects revealed, that they have to be differently interpreted for groups of stocks accordingly to their trading volume. The trading strategies should be based on the analysis of stocks risk, expressed as variance, because the mean return indicator was not directly affected by the explored calendar anomalies.

References

1. Achelis, S.B.: Technical Analysis from A to Z, 2nd edn. McGraw-Hill, New York (2000)
2. Thomsett Michael, C.: Getting started in fundamental analysis, p. 232. Wiley, Chichester (2006)
3. Sakalauskas, V., Kriksciuniene, D.: Statistical investigation on the day-of-week effect in emerging stock markets. In: Artificial intelligence and applications: proceedings of the international conference, Innsbruck, Austria, February 11-13, pp. 146–151 (2008) ISBN 978-88986-709-3

4. Sakalauskas, V., Kriksciuniene, D.: Neural networks approach to the detection of weekly seasonality in stock trading. In: Fyfe, C., Kim, D., Lee, S.-Y., Yin, H. (eds.) IDEAL 2008. LNCS, vol. 5326, pp. 444–451. Springer, Heidelberg (2008)
5. Sullivan, R., Timmermann, A., White, H.: Dangers of Data-Driven Inference: The Case of Calendar Effects in Stock Returns, Working Paper, University of California, San Diego (1998)
6. Balaban, E., Bayar, A., Kan, O.B.: Stock returns, seasonality and asymmetric conditional volatility in World Equity Markets. Applied Economics Letters 8, 263–268 (2001)
7. Flannery, M.J., Protopapadakis, A.A.: From T-bills to common stocks: investigating the generality of intra-week return seasonality. Journal of Finance 43, 431–450 (1988)
8. Brooks, C., Persand, G.: Seasonality in Southeast Asian stock markets: some new evidence on day-of-the-week effects. Applied Economics Letters 8, 155–158 (2001)
9. Tang, G.Y.N.: Day-of-the-week effect on skewness and kurtosis: a direct test and portfolio effect. The European Journal of Finance 2, 333–351 (1998)
10. Sakalauskas, V., Kriksciuniene, D.: The Impact of Taxes on Intra-Week Stock Return Seasonality. In: Bubak, M., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (eds.) ICCS 2008, Part II. LNCS, vol. 5102, pp. 504–513. Springer, Heidelberg (2008)
11. Kamath, R., Chusanachoti, J.: An investigation of the day-of-the-week effect in Korea: has the anomalous effect vanished in the 1990s? International Journal of Business 7, 47–62 (2002)
12. Reschenhofer, E.: Unexpected Features of Financial Time Series: Higher-Order Anomalies and Predictability. Journal of Data Science 2(2004), 1–15 (2004)
13. Basher, S.A., Sadorsky, P.: Day-of-the-week effects in emerging stock markets. Applied Economics Letters 13, 621–628 (2006)
14. Thaler, R.: The effect. Journal of Economic Perspectives (1), 197–201 (1987a)
15. Thaler, R.: Seasonal movements in security prices II: weekend, holiday, turn of the month, and intraday effects. Journal of Economic Perspectives (1) (2), 169–177 (1987b)
16. Mills, T.C., Coutts, J.A.: Anomalies and calendar affects in the new FT-SE indices. European Journal of Finance (1), 79–93 (1995)
17. Sullivan, R., Timmermann, A., White, H.: Dangers of data mining: the case of calendar effects in stock returns. Journal of Econometrics (105), 249–286 (2001)
18. Kohers, G., Kohers, N., Pandey, V., Kohers, T.: The disappearing day-of-the-week effect in the world's largest equity markets. Applied Economics Letters 11, 167–171 (2004)
19. Steeley, J.M.: A note on information seasonality and the disappearance of the weekend effect in the UK stock market. Journal of Banking and Finance 25, 1941–1956 (2001)
20. Brooks, C., Persand, G.: Seasonality in Southeast Asian stock markets: some new evidence on the day-of-the-week effects. Applied Economics Letters 8, 155–158 (2001)
21. Aggarwal, R., Rivoli, P.: Seasonal and day-of-the-week effects in four emerging stock markets. Financial Rev. 24, 541–550 (1989)
22. NASDAQ OMX Group (2009), http://www.nasdaqomxbaltic.com/
23. StatSoft Inc. Electronic Statistics Textbook. StatSoft, Tulsa (2006), http://www.statsoft.com/textbook/stathome.html

The Issues Concerning the Application of Multiple Evaluation Methods for the Projects in Lithuanian Companies

Gerda Zigiene[1] and Egle Fiodoroviene[2]

[1] Vilnius University Kaunas Faculty of Humanities, Department of Finance and Accounting, Asoc. Prof, PhD
zigiene@vukhf.lt

[2] Vilnius University Kaunas Faculty of Humanities, Department of Finance and Accounting, Master
egle.fiodoroviene@gmail.com

Abstract. In this paper the methods for financial resource management related to project planning, its implementation and control reaching for goals set, are analyzed. The accomplished empirical research allowed to assess the implementation tendencies of evaluation methods for projects in Lithuania. The application of economical-financial evaluation methods, the methods of financial resource planning and budgeting in Lithuanian companies was determined, identifying the reasons for which these principles are kept or not and distinguishing their application problems as well.

Keywords: principles of project management, GANTT, PERT, CPM project evaluation methods, project earned value.

1 Introduction

In worldwide practice, the principles of project management are widely applied in companies belonging to both private and business sectors. Normally, these methods allow companies to adapt to environmental changes easier, what is especially relevant in the context of globalization process. The implementation of several principles of project management in the company activity is currently being applied more and more widely in Lithuanian companies as well. Partially, such a tendency is related to that it quite a large part of economical subjects are seeking to use funding from European Union funds and give applications for the projects. Commonly, the activities of these projects are required to be distinguished from the general context of company activity and to give elaborate reports about the progress of the project, cost, income, impact on the company and its external environment. However, the project management is not only related to the received funding. The growing number of companies which begin to resolve their activity in separate projects in everyday activity is noticeable.

When implementing the principles of project management in the companies, the distribution of financial resources becomes an important problem, because company‘s

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 79–90, 2009.

financial resource management is one of the most important and ``sensitive" elements of business management. When the methods of project financing are rightly chosen, the appropriate financial resource planning, thoroughly evaluated investments and their profit can condition the business development and success.

The Aim of the Paper: to analyze the aspects of application of evaluation methods for the projects in Lithuanian companies' activity, in terms of the application of multiple evaluation methods for the projects.

- To analyze the methods for financial resource management related to project planning, its implementation and control, reaching for goals set.
- To do an empirical research which would allow to reveal the implementation tendencies of project evaluation models in Lithuania.
- To determine the project economical-financial methods and methods of financial resource planning, budgeting in Lithuanian companies, identifying the reasons for which these principles are done or not and distinguishing their application problems as well.2 Paper Preparation.

2 Financial Resource Planning and Distribution

A rational enterprise resource planning (ERP) is one of the most important factors in seeking for leading positions in the market and determining the development of company activity. The resource distribution and their effective management is related to many processes, which take plane in the company's everyday life [12].

According to White [26], the financial resource management is one of the most important aspects in project management, because it is directly related to the formation of the activity and finance distribution within time and place. Choosing the method of finance management, a company can begin to plan the finances needed for project management. There exist several methods of financial resource management, which allow to determine when and what resources are needed for the company implementing the project. The resource planning is related to financial resource distribution [10], [12], [11], because it is necessary not only to invest in goods and services being purchased, but also to evaluate additional cost such as conveyance, storage, insurance and so on. In order to set the activities within time, and at the same time to anticipate when and what resources will be needed, Gantt and PERT methods are used.

Henry Gantt's visual method, created in 1920 approximately, was the originator of planning methods which focused on the specific tasks over a particularly determined period. The creation of Gantt table was a first step in developing many other visual and organizational methods for business projects. Gantt table compares time on x axis with different finances needed in order to do the tasks put on y axis [8]. Gantt's merit is the first time line tables, however separate hypothesis still bring forward an idea that similar tables were used for the building of Egyptian pyramids [8].

Another medium for putting activities within time- PERT scheme (Program Evaluation and Review Technique) - was created in 1959 with the purpose to simplify the planning of big and complex projects. In this scheme the explanations can be entered, giving the opportunity to plan the project not knowing exactly the details and durations of all the activities. This scheme is more oriented towards separate cases rather than recurring projects. It is used in the investigative, developmental type of projects.

PERT scheme is a simplified PERT method for project management, which ideally applies to the early stage of project management, when time for works has not been set yet accurately; then works are considered as random variables with a supposed distribution of probability. PERT chart can be optimized relating two types of cost: (a) direct cost for each work, which increases when finances increase, budgeted to reduce the work time; (b) indirect cost which is related to a complete duration of the project. This cost is directly exponential to the general duration to finish the project.

The alternative to PERT is named the critical path method (CPM) of DuPont Corporation, which was created in 1957 [19]. In both PERT and CPM methods for project planning and management, a network is used which pictures correlation between the works related to the project. In this case, networks become useful or necessary in order to create empty works and ascertain the priority connection between the activities. The difference between these two methods is that using CPM method, critical path usually can be determined by tracing the works which reserved time is equal to zero in order of the accomplished proceedings. The sum of the periods for the accomplishment of works in critical path shows the shortest possible period to finish the project [17].

3 The Reciprocity of Project Management Processes in Terms of Limitation of Financial Resources

The implementation of principles of project management in practice is more often related to the efficiency of financial resource management in private business companies, which reduces their cost and improves the results. As Romberg [21] notices, in his analyzed companies in 1990-1998, when starting to implement the principles of project management, despite the new technologies, the budget for the projects was exceeded in most cases (almost in 89% of companies). However, according to the author, when the American concern "Gartner Group" related the principles of project management to the control of cost distribution, it achieved especially good results, allow to increase company's sales by 20%.

Financial resource is an essential condition for continuous company activity and development [9], [1]. One of the main tasks, initiating project management processes for a company, is „a sufficient" amount of financial resources that ensure a succession of these processes. Financial resource management gives company an opportunity to comply with project terms, realization of production size and its cost price, because this guarantees profitability of a project [11]. A received amount of resources is influenced by the company's control, payment forms and methods, financial situation and discipline of a purchaser, production quality and other factors.

Financial methods for project management. In theory, usually several project evaluation methods are distinguished. Normally, all companies' projects are expressed in financial indexes already during the planning stage [11]. As American Strategic Management Institute states, the project evaluation methods are one of the main factors which determine further decisions concerning project implementation. [15], [23], [6], [7] distinguish three main methods for project financial resource evaluation and use: Net Present Value (NPV), Internal Rate of Return (IRR) and

Payback Period methods. When applying NPV method, all the forthcoming money flows for the planned project are recalculated to the equivalent amount of money at the moment of current time, withdrawing initial investments and evaluating interest rate of the credits proposed by the banks. Since, according to the definition, profitability rate is the interest rate to which net present value is equal to zero; counting with this rate, discount costs of a project are equal to discount income. A Company, applying IRR method, determines "preferential profitability rate" which is least acceptable and compared to the practicably receivable profit. A project, of which profitability rate is higher than the comparable one, is acceptable to a company and it is analyzed further. If calculated profitability rate is less than the comparable one, then it is considered unacceptable. Payback period shows a relative attractiveness of investment offer. It determines what will be the amount of periods needed to reimburse an initial investment. It determines how many periods will be needed that cumulative usefulness of investment project would equal to its cumulative cost. In this method, it is being operated with money flows that are both for project usefulness and its cost is expressed in money flows. For each of the project alternatives, payback period is calculated and received values are compared with each other. Then an alternative is chosen which payback period is the shortest.

Multiple methods for project management. According to Loshi, Al-Mudhaki and Bremser [14], budgeting involves the process of whole information processing, which however, is essential, in order to evaluate cost for the planned activities. In many analyzed cases [14], during the performance time of mentioned processes, project cost is examined. However, there is another performance method of this process which involves project earned value (EV) evaluation as a medium for the progress measurement.

This method shows a direct relationship between project earned value and the percentage degree of accomplishment. Earned value has three main features: (a) it is a simple and consistent measure to evaluate project progress, (b) earned value method allows to analyze project activities constantly, (c) allows to compare cost between project activities. Usually, in the earned value analysis, when evaluating the recipiency of accomplished work, monetary and time units are used. By earned value method budget is formed, so called planed budget of planned work, which directly determines the amount of company's used financial resources related to planned activities [13]. The comparison process is done between:

Budgeted Cost of Work Scheduled (BCWS), which is equal to project Planned value (PV) [16]. In it, budget confirmed in the beginning of a project is given, which is directly related to the works scheduled in regard to time and cost [20].

Actual Cost of Work Performed (ACWP) or Actual Cost (AC) [16], which is described as actual cost for the activities performed [20].

Budgeted Cost of Work Performed (BCWP), which is equal to project earned value (EV) [16]. It is evaluation of actual performed activities, considering the schedule of works planned [4]. In this case a simple calculation is done- earned value (EV) is factually equal to actual budget cost or Budget at Completion (BAC) multiplied from a percentage of work completed (%, COMP)(EV = BAC*COMP %) (this value shows actual value of actually performed works over planned period of time and is measured by monetary units).

The use of Earned value method allows companies (a) to manage the performance of the activities and update schedule, (b) to determine and register current cost for every activity performed, (c) to calculate total earned value for every activity and project and (d) to analyze data and prepare the report of performance and progress.

When drawing earned value system it is important: (a) to make a work breakdown structure (WBS) [3], which divides project into its constituent activities (or merely into manageable parts), (b) to calculate and distribute cost for each activity, (c) to plan activities in separate periods of times and (d) to budget [20], scheduling the cost planned and works performed.

In Aramvareekul, Seider [4], given Cost-Time-Risk Diagram (CTR) is another method for project planning and management. The authors expect that Cost-Time-Risk Diagram will help project managers to evaluate project risk problems by managing implementation of time and cost distribution in one diagram [4].

Cost-Time-Risk Diagram shows the relationship between cost and time of the current project, related to possible rinks. This diagram is an advanced analysis of Earned Value Management (EVM). This diagram is as a map of project implementation control. It helps project manager to make more effective decisions during the project implementation period. By using this method, project managers can watch the performance of time and cost distribution in the same time diagram which is important for making the decisions related to the combination of time and cost. Without combination of time and cost, Cost-Time-Risk Diagram distinguishes risky and useful points for each period of project implementation. Risky and useful points are evaluates and used in Cost-Time-Risk Diagram in order to ensure timely project implementation not exceeding the budgeted finances [4].

Earned value analysis (EVA) and earned value management (EVM) is a main method for project management in concordance with Cost-Time-Risk Diagram. Cost-Time-Risk diagram is a combination of earned value analysis and risk management.

In accordance with the authors, risk management is an important aspect of quality security, using the main analysis methods and performance measurements in order to ensure that risks are properly identified, evaluated and controlled.

This method can become an effective medium to project managers when determining the deviations of project implementation in time and budget planning. The graphs in Cost-Time-Risk Diagram give the schedule of current project with integrated risk evaluation for each period of time. Risk evaluation is an important aspect of project planning and management, improving the decision making and reducing the possible indetermination [4]. If risk is not being controlled effectively, the unplanned occurrences could seriously threaten the projects; then, as a result, the overrun of budget, the inadequacies in time schedule, quality problems and/or unrealized goals are possible. Cost-Time-Risk Diagram is also an effective medium for management of decision making, especially for compatibility of time and cost.

Using this method in the first stage- the work breakdown structure is made. Then, similarly as in earned value method, three types of cost are compared: (a) budgeted cost for work scheduled (BCWS), (b) actual cost of work performed and (c) budgeted cost of work performed. This comparison, with the help of software, is “transferred”

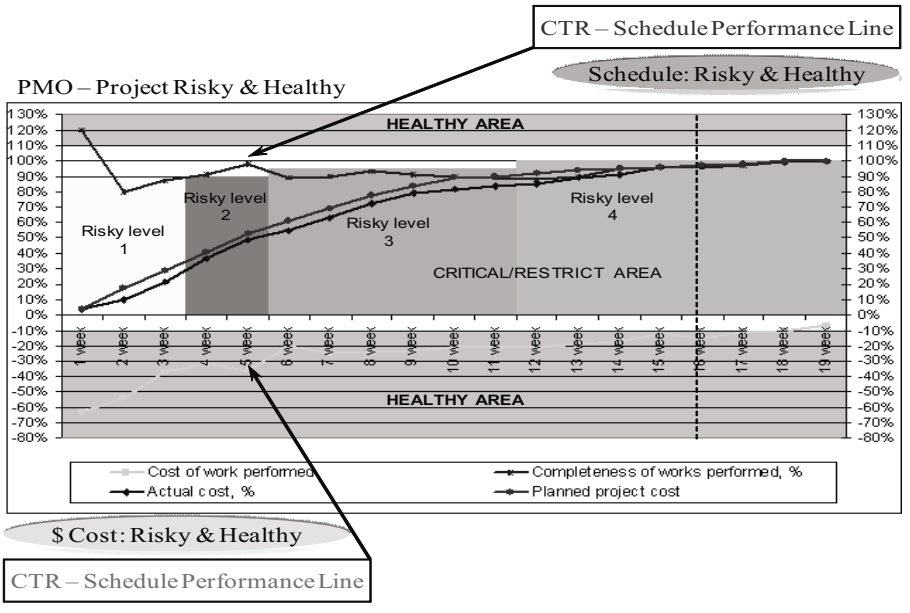


Fig. 1. The interaction between the project budget, schedule and CTR method. *Source*: Aramvareekul, Seider, 2006.

to a complicated Cost-Time-Risk Diagram which allows to clearly determine the actual cost value [4]. This is illustrated in figure 1:

As it is seen in the figure, two main functions are drawn which respectively are (a) a degree of completeness of works scheduled in every evaluation period (schedule performance line, CTR-S), expressed in percentage (%) and (b) the sum of cost compatible to budget scheduled if ever evaluation period expressed in percentage (%) (Cost performance, CTR-C). Both of these functions are attributed to three comparison indexes of activity, budget plan and results achieved. According to these data, CTR diagram is divided into two main fields- safety zone and danger zone, in which project cost (lower function) and works planned and performed in the project (upper function) are reflected. These zones differ dependently on the period of time that is whether in the initial or final stage of the project. When project activities are soon to be finished, even marginal deviations (just of the few percents) from the schedule can get into the risk zone.

In many cases, the diagrams of project implementation (example in the picture) involve a graphic depiction of company's tasks and their correlation [25]. According to Amar, Berman, Amornsawadwatana [2], Cost-Time-Risk Diagrams allow to reveal a connection between cost and time related to the possible risks. This method can become an effective medium to project managers when determining the deviations of project implementation in time and planning the performance of budget [2].

4 A Research and Analysis of Application of Project Management Methods in Lithuanian Companies

The principles of project management involve not only organizational solutions [22], but also financial resource management, relating them with strict and constantly controlled distribution of cost. Since direct application of western management theories under Lithuanian circumstances is partially methodologically risky, because the acceptability of such theories is not secured [24], the empirical research was made (questionnaires) the aim of which is to ascertain whether the implementation of project management methods is relevant in our country.

Scope of research. According to the data from department of statistics[1], the innovation implementation in Lithuanian companies is distributed irregularly and major part of companies is in the three largest regions of Vilnius, Kaunas and Klaipėda. Because of the easiest accessible information, 69 public companies which, according to the data from department of statistics, were registered in Kaunas region and performed their activity in the beginning of 2008 were chosen.

In order to get the results of 95% of reliability, the amount of respondents needed was calculated in concordance with Paiotto formula:

$$n = 1/(\Delta 2+1/N). \quad (1)$$

where Δ is a fixed value varying dependently on the needed reliability of result, in this case Δ= 0,05;

N- The general value of totality, in the case of the research 69.

Having done the calculations, the initial scope was obtained:

$$N=1/(\Delta^2+1/N)= 1/(0{,}05^2+1/65)= 56 \quad (2)$$

Thus, in this research, the questionnaires were sent via e-mail to 56 companies in Kaunas region. From 56 questionnaires sent via e-mail, 51 were filled in properly, what constitutes 91% of sent questionnaires. It is important to pay attention to the results of the research allow to analyze the received information by quantitative method and to draw the conclusion only about questioned respondents. The data of quantitative analysis can not be treated as representative on Lithuanian scale. The results only reveal the possible tendencies.

When making a questionnaire for the research, in some questions respondents are asked to evaluate the importance of given statements that is attitude towards a particular phenomenon. The answers were evaluated according to Likert scale. Analyzing data with the help of Likert scales, each evaluation was given the expression in numbers, what helped to determine positive or negative attitude towards the subject or phenomenon being researched and arithmetical average of the results (evaluations) which reflect a general attitude of respondents.

[1] Department of Statistics. Access via internet: <www.stat.gov.lt>

The questionnaire is divided into two main groups: (1) with the first group of questions it is sought to ascertain how often the principles of project management are used in Lithuanian companies and how this implementation is related to European Union funding, (2) with the second group of questions it is sought to determine the cost planning, budgeting, success and risk evaluation methods in the companies implementing the principles of project management. Because the aim of this paper is to analyze the application of project management methods in Lithuanian companies, the second part of the questionnaire will be analyzed further on:

The determination of Cost planning and control methods in Lithuanian companies

Oke, Charles-Owaba [18] notices that companies often form their cost planning methods by themselves on the basis of Gantt and PERT, so there is a possibility for the companies to answer "other" and tell in detail what systems and methods they applied to calculate cost. In order to reveal the control systems for cost planning more accurately, next question is related to the use of these methods. All the companies in accordance with activity type use Gantt method most often that is 78% companies (40); significantly more rare is PERT method, it is only used by 41% respondents (21); other methods are used by 22% companies (11) in Kaunas region. The results of the research done by Oke, Charles-Owaba [18] confirm this and reveal that in man cases companies use all the methods together, applying various software to plan and control the cost for separate projects. The results also revealed that PERT method was not used separately as cost planning and control method. This can be related to many reasons: lack of knowledge, imperfection of software or lack of knowledge using IT technologies and that.

Economical-financial methods for project evaluation in Lithuanian companies. Dedi, Orsag [7] studying economical-financial methods for project evaluation has determined that the main methods are of: (a) Net present value (NPV), (b) internal rate of profit value, (c) project payback period. In practice, the companies having determined the cost ad income of initiated project, evaluate its benefit by economical-financial methods which, according to American Strategic Management Institute (2006), is a rational method to choose an optimal decision when determining perspective projects. In the questionnaire, economical-financial methods for project evaluation, most often written down in scientific literature [9], [7] are given.

From 51 companies which participated in the research, 90% (46) use "income and cost analysis" method for project management, 92% (47% –"project payback period" method, 75% (38)- "net present value" (NPV) method and "internal rate of benefit" method is used by 69% companies (35). Analyzing the economical-financial methods for project management being applied in practice, according to Graham, Harvey [9], Dedi, Orsag, [7], a bigger part of project managers in American companies use Net present value (NPV) method (76%) and internal rate of return value (IRR) (75%) methods, comparing them to project payback period method (56%) [9], which is more popular in European countries. The authors studying scientific literature emphasizes that research results in many European countries revealed that one of the most popular economical-financial methods for financial evaluation in Europe is payback period method, which is used by 63% project managers in Great Britain, 76% in Finland and 61% in Germany. However Net present value (NPV) and Internal rate of return value (IRR)

methods are used more rarely in Europe than in America. Respectively, 43% companies in Great Britain, 52% in Sweden and 45% in Germany evaluates their projects using net present value (NPV) method; and an internal rate of return value method is used b 57% companies in Great Britain, 23% in Sweden and 36% in Germany.

A comparison between planned and actual cost. According to Christensen, [5], Marshal [16], a company's risk controlled with earned value method involves many areas in project activities related to (1) project organization, (2) cost planning and budgeting, (3) project accounting, (4) complete project analysis and control, (5) timely identification of mistakes and its correction. Due to this reason, a large part of the companies operating in western countries not only calculate and compare (a) budgeted cost of work performed with (b) actual cost of work performed and (c) planned budget cost at the moment of evaluation, but also draw diagrams of indexes mentioned. This encouraged to give question i the questionnaire related to diagram drawing. Since Christensen [5] reveals that companies could draw diagrams in some projects and do not do so in smaller projects, the opportunity to answer "partially" is left to respondents, which would deny or confirm Christensen's [5] statement.

It could be stated that budget planning and performing involve process of information analysis related to project financial resources. This process includes planned cost and income of actually performed work. In the case studied, project cost is actualized by distributing them into budgeted cost of work performed; actual cost of work performed and planned budget cost at the moment of evaluation:

Table 1. A planned calculation of frequency of budgeted cost of work performed, actual cost of work performed and planed budget cost at the moment of evaluation in Lithuanian companies

Statement	Index estimation	
	Arithmetical average of evaluation	Standard deviation
Budgeted cost of work performed or project value	5,67	1,69
Actual cost of work performed or actual project cost	6,11	1,43
Planned budget cost at the moment of evaluation or project earned value cost	4,67	2,23

In the table, respondent companies' averages of answers are given, which show if these companies estimate indexes of budgeted cost of work performed, actual cost of work performed and planned budget cost. As it is seen from the table, the evaluation of these indexes, even if they are estimated, show that it is not done by every company in Lithuania. In most cases, actual cost of work performed is estimated (a.a. 6.11), and in most rare cases planned budgeted cost at the moment of evaluation is estimated (a.a. 4,67).

Companies which implement the principles of project management only in the activities funded by European Union do not estimate these indexes. Because Anbari [3], Aramvareekul, Seider, [4] and Marshall [16] suggest to compare these indexes by drawing earned value and time diagrams, which allow to evaluate the risk of project

implementation, related to cost and time, graphically, this question was also given to participants. Only 14% (7) of companies draw earned value and time diagrams, 51% (26) draw them partially that is use graphic evaluation methods created in their company and 35% (18) do not draw diagrams at all.

5 Conclusions

Analyzing scientific literature, limitation theory was distinguished, which principles state that in all the processes of project management, cost has to be constantly controlled and analyzed by using company's limited financial resources. By analyzing scientific literature it was also revealed that not only budgeting is important, but also a constant budget control as well. Analyzing the project management methods, the methods analyzed mostly by the foreign authors were distinguished at theoretical level. In the paper were analyzed the multiple methods for project management of (a) budgeted cost of work performed, (b) actual cost of work performed, (c) planned budget cost at the moment of evaluation. With the help of these methods, companies applying the principles of project management, can evaluate budget, follow its implementation efficiency, reduce the risk to exceed budget and to deviate from the schedule.

In order to do an empirical research, a questionnaire was formulated; by which witch the implementation tendencies of project principles in Lithuania are revealed.

In order to evaluate how the principles of project management are implemented in Lithuanian companies, 51 average and small public companies were questioned (scope is based on Paniotto formula):

In public companies, when planning and controlling financial resource management, Gantt method is used most often (78%), which is a practical analytical medium, because it gives view of all project activities with elaborated periods of time, what facilitates the determination of necessary resources and periods of time, controls and plans the costs. However, large part of companies, dividing their activity into projects, have answered that when analyzing their project cost, they use Gantt and Pert methods together;

For the economical and financial evaluation of the project, many participated companies use „project payback period" (92%) and „income and cost analysis" (91%) methods. In contemporary practice, the main „net present value"(75%) and „internal rate of return"(69%) methods are distinguished, which allow to evaluate the influence of inflation on project value. These current methods are applied slightly more rarely.

Companies, being researched in the context of contemporary economical situation, implementing the principles of project management, should constantly follow and identify the tendencies in both global and local market changes, determining the possible risk which reduces the usefulness of project to a company.

References

1. Akalu, M.M., Turner, R.: Investment Appraisal Process in the Banking & Finance industry. A Case Study. Publications in the ERIM Report Series Research _ in Management ERIM Research Program: Organizing for Performance (2002), http://studentweb.tulane.edu/~mtruill/dev-pert.html

2. Ammar, A., Berman, K., Amornsawadwatana, S.: A Review of Techniques for Risk Management in Projects. Benchmarking: An International Journal 14(1), 22–36 (2007)
3. Anbari, F.T.: Earned Value Project Management Method and Extensions. Project Management Journal, 12–23 (December 2003)
4. Aramvareekul, P., Seider, D.: Cost-Time-Risk Diagram: Project Planning and Management. Cost Engineering 48(11), 12–18 (2006)
5. Christensen, D.S.: The Costs and Benefits of the Earned Value Management Process. College of Business, Southern Utah University (1998), http://findarticles.com/
6. Danielson, M.G., Scott, J.A.: The Capital Budgeting Decisions of Small Businesses. Journal of Applied Finance (2006), http://findarticles.com/
7. Dedi, L., Orsag, S.: Capital Budgeting Practices: A Survey of Croatian Firms (2007), http://web.ebscohost.com/
8. Finch, N.: The Motivations for Adopting Sustainability Disclosure. MGSM Working Paper No. -17 (2005), http://ssrn.com/abstract=798724
9. Graham, J.R., Harvey, C.R.: The Theory and Practice of Corporate Finance: Evidence from the Field. Journal of Financial Economics 60, 187–243 (2001)
10. Judgev, K., Mathur, G.: Project Management Elements as Strategic Assets: Preliminary Findings. Journal: Management Research News 29(10), 604–617 (2006)
11. Koh, S.C.L., Simpson, M.: Could Enterprise Resource Planning Create a Competitive Advantage for Small Businesses? Benchmarking: An International Journal 14(1), 59–76 (2007)
12. Law, C.C.H., Ngai, E.W.T.: An Investigation of the Relationships Between Organizational Factors, Business Process Improvement, and ERP Success. Benchmarking: An International Journal 14(3), 387–406 (2007)
13. Locker, K., Gordon, J.: Project Management and Project Network Techniques. Financial Times, p. 272. Prentice Hall, Essex (2005) ISBN-13: 978-0273614548
14. Loshi, P.L., Al-Mudhaki, J., Bremser, W.G.: Corporate Budget Planning, Control and Performance Evaluation ir Bahrain. Managerial Auditing Journal 18(9), 737–750 (2003)
15. Maccarone, P.: Organizing the Capital Budgeting Process in Large Firms. Management Decision 34(6), 43–56 (1996)
16. Marshal, R.A.: The Contribution of Earned Value Management to Project Success on Contracted Efforts: a Quantitative Statistics Approach within the Population of Experienced Practitioners. International Journal of Managing Projects in Business 1(2), 288–294 (2006)
17. Modell, M.E.: A Professional's Guide to Systems Analysis, 2nd edn., p. 312. McGraw Hill, New York (1997), http://studentweb.tulane.edu/~mtruill/dev-pert.html
18. Oke, S.A., Charles-Owaba, O.E.: A Sensitivity Analysis of an Optimal Gantt Charting Maintenance Scheduling Model. International Journal of Quality & Reliability Management 23(2), 197–229 (2006)
19. Project Management Institute, Inc. Making Project Managment Indispensable for Business Results (2008), http://www.pmi.org
20. Raby, M.: Project Management via Earned Value. In: Work study, pp. 6–10. MCB UP Ltd. (2000), http://www.emeraldinsight.com/
21. Romberg, D.: Project Managment Tools Cannot Guarantee Success. Computing Canada 24(42), 29 (1998)
22. Soderlund, J.: Developing Project Competence: Empirical Regularities in Competitive Project Operations. International Journal of Innovation Management 9(4), 451–480 (2005)

23. Stungurienė, S.: Solving Problems of Optimal use of Resources in Business Organizations: Practical Application Methods. Organizacijų vadyba: sisteminiai tyrimai (2005), http://web.ebscohost.com/ ISSN 1392-1142
24. Šaparnis, G., Merkys, G.: Kokybinių ir kiekybinių metodų derinimas mokyklinės vadybos diagnostikoje: hipotezė ir pirmieji rezultatai. Socialiniai mokslai: Kauno technologijos universitetas 2(23), 43–55 (2000)
25. Tavares, L.V.: A Review of the Contribution of operational Research to Project Management. European Journal of Operational Research 136(1), 1–18 (2002)
26. White, M.: Information Technology: A Mandatory Role in Construction Project Management. Cost Engineering 49, 18–20 (2007), http://web.ebscohost.com/

Control View Based Elicitation of Functional Requirements

Audrius Lopata[1,2] and Saulius Gudas[1,2]

[1] Kaunas University of Technology, Information Systems Department,
Studentu St. 50, Kaunas, Lithuania
Audrius.Lopata@ktu.lt
2 Vilnius University, Kaunas Faculty of Humanities,
Muitines St. 8, Kaunas, Lithuania
Gudas@soften.ktu.lt

Abstract. The paper deals with knowledge-based enterprise management modeling and user requirements acquisition. The enterprise management is considered from the control point of view, management function is formally predefined as Elementary Management Cycle (EMC). Few new types of specialized Workflow models are deployed for domain knowledge elicitation and BP analysis. Two types of logical gaps are identified by transformations of empirical BP model to knowledge based BP model.

Keywords: knowledge-based enterprise management modeling, control point of view, Elementary Management Cycle, enterprise management function, types of specialized Workflow model, knowledge-based BP analysis, user requirements, Enterprise Meta-Model, Use Case Model.

1 Introduction

Many mistakes in the area of business process (BP) modeling and user requirements acquisition can be avoided when applying formalized (algorithmic) methods of business domain analysis, BP modeling and user requirements refinement. However, the integration of Enterprise modeling techniques into the information systems development process is still not sufficient [13], [14]. A characteristic feature of CASE methods is their empirical nature, because the project models repository of CASE system is composed on the basis of empirical information - information about real world is not verified through formalized criteria. The problem domain knowledge elicitation process relies heavily on the analyst and user; therefore it is not clear whether the knowledge about this problem domain is adequate [16].

Usually user requirements specification starts from the interactive construction of the Use Case Model (UCM). In this case UCM is constructed by IS designer which takes BP model as a source of knowledge about problem domain without examining such empirical BP model according to some formal or formalized criteria.

From our point of view only verified and validated BP model (i.e. knowledge-based Enterprise Model which is examined through formalized criteria domain knowledge) should be stored in the knowledge repository of CASE tool, and should be used to control development of IS project solutions. The set of components types

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 91–102, 2009.

and relationships types of knowledge-based Enterprise Model should be regulated by formalized specification, which is called Enterprise Meta-Model [5].

There is a great number of Enterprise modeling methods and approaches, which defines the Enterprise components (such as CIMOSA, GERAM [3], IDEF suite, GRAI), standards (ISO 14258, ISO 15704, PSL, ISO TR 10314, CEN EN 12204, CEN 40003) [10], UEML [12].

The human (an expert, user) plays the pivotal role in domain knowledge acquisition, and few formalized methods of knowledge acquisition control are taken into consideration. Therefore, gaps of IS engineering process occur due to the human factor.

This paper deals the knowledge-based approach to enterprise management modeling and integration with computer-aided specification of user functional requirements. The similar (in some extent) approaches are described in [1], [2], [11] and [15].

The approach to use case modeling through business modeling is presented in [11]: the use cases are elicited and structured on the basis of the business processes of the organization, represented by Role Diagram, Sequence Diagram and process diagram.

The steps of BP model transformation to functional requirements specification, specified in the form of use case diagrams is described in the [1], [2]. This approach is meta-model based since the meta-models for use case diagrams and for business process models are defined and the mapping between these two meta-models is defined.

The peculiarity of our approach to knowledge-based business process (BP) modeling is the modeling of any enterprise management function as formally defined unit - Elementary Management Cycle (EMC) [5], [7]. So, the *enterprise management functions* (information processing and decision making) and *enterprise processes* (primary activities, product development processes) are considered from the management (control) point of view as predefined structure, namely – the EMC.

The modified Workflow modeling notation is used for representation of the components of *enterprise management function* required by definition of the EMC [7]. The specialized Workflow models (WFM) of six types are developed and deployed for representation of knowledge-based user requirements acquisition process [9].

2 The Principles of Knowledge-Based Enterprise Management Modeling

Systems analysis of IS development methods and tools refines the trend towards the knowledge–based IS engineering systems, it shows the cause of feasible changes in architecture of CASE tools. The principles of knowledge-based IS engineering process were refined in [4]. The knowledge-based CASE process is defined and constructed on the basis of domain knowledge accumulated in the Knowledge Base of enhanced CASE system [6]. A Knowledge Base of enhanced CASE system consists of two layers: Enterprise Model (EM), Enterprise Meta-Model (EMM). The Enterprise Model (EM), Enterprise Meta-Model (EMM) and formal Enterprise model (i.e. some theoretically defined Enterprise Management Framework [7]) are the major concepts of any knowledge-based CASE process and obligatory components of knowledge-based CASE tool. The peculiarity of this approach is as follows - to BP

modeling is focused on the *enterprise management modeling*. The concepts of *enterprise management function* and *enterprise process* could be illustrated, for instance by analysis of Value Chain model [17] from control point of view [5]. The traditional *enterprise management functions* of Value Chain Model are support activities as follows: financial policy, accounting, human resource management, technology development, procurement, etc. [17]. *An enterprise management function* {Fj} is identified in the structured Value Chain Model as a type of support activities and *enterprise process* (Pi) is identified as a type of primary activities (see Fig. 1).

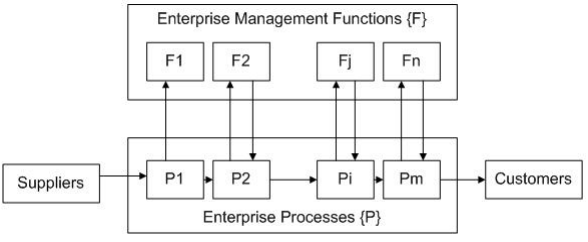


Fig. 1. The Structured Value Chain Model

An enterprise management activity is considered from the control point of view, any enterprise management function {Fj} is formally predefined as Elementary Management Cycle (EMC) [5, 7].

The inside structure of any enterprise management function {Fj} (information architecture of management function) is predefined as closed cycle of information transformation (see Fig. 2) according to control point of view [8].

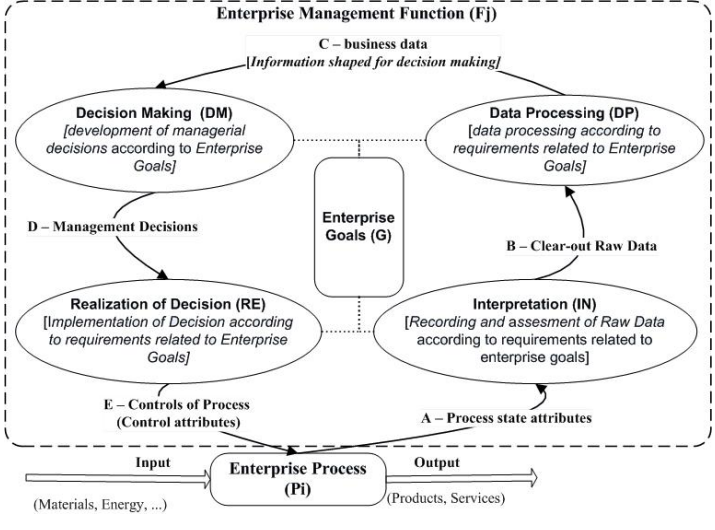


Fig. 2. The inside structure of any enterprise management function {Fj} is considered as closed cycle of information transformation

An enterprise management Function (Fj) consists of the predefined sequence of mandatory steps of information transformation (Interpretation (IN), Data Processing (DP), Decision Making (DM), Realization of Decision (DR)); these steps compose a closed management cycle (a feedback loop). A definite types of attributes (Process state attributes (A), Clear-out Raw Data (B), Business data (C), Management Decisions (D), Controls of Process (E)) are formed and transmitted during each management cycle step [5].

The workflow modeling (WFM) notation is used for enterprise management modeling in the paper. Six types of modified WFM are defined and deployed for presentation of initial (empirical) business process model and it's transformations into enterprise management model.

3 Approach to Elicitation of Domain Knowledge

The elicitation of user requirements is the initial stage of traditional IS development life cycle, starting with enterprise modeling. Therefore, the user and the analyst are two sources of information in traditional IS engineering. Most of user requirements acquisition techniques are based on empirical information provided by the user (business domain expert). Problems occur when empirically acquired problem domain information (enterprise model, requirements) has to be verified and validated.

The Enterprise Knowledge Repository of CASE system is considered to be the third source of domain knowledge for IS development stages – both for user requirements analysis and specification and for other IS development life cycle stages.

The presented BP modeling and user requirements acquisition process is developed from management (control) point of view [5]. The workflow model notation is selected for representation of BP models. This knowledge–based approach includes transformations of few types (modifications) of the workflow model:

- Workflow Model of Business Processes (VP_WFM), this is the traditional workflow model aimed to specify an expert knowledge (empirical information) about problem domain processes, material and informational flows and actors;
- Workflow Model of Processes (P_WFM) is a part of (VP_WFM) and includes only material processes, material flows and related actors of the problem domain;
- Workflow Model of Functions (F_WFM) - and includes only informational flows and related actors of the problem domain;
- Workflow Model of Processes without Gaps and Workflow Model of Functions without Gaps are intermediate results in transformations from empirical workflow model (VP_WFM) to knowledge-based workflow model (FS_WFM);
- Workflow Model of Functional Composition (FS_WFM) is developed from control point of view and specifies the composition of definite business management function in accordance with definition of EMC [7], [5].

The refinement of formally correct enterprise management function is a sequence of BP model transformations. The algorithms of four types for transformation of the empirical BP model (VP_WFM) to formally correct enterprise management function (FS_WFM) are developed already:

A1. The algorithm which identifies informational activities and material processes (presented in empirical workflow model (VP_WFM) and separates VP_WFM into Workflow Model of Processes (P_WFM) and Workflow Model of Functions (F_WFM);

A2. The algorithm which identifies and eliminates logical gaps in the P_WFM;

A3. The algorithm which identifies and eliminates logical gaps in the F_WFM;

A4. Validation of the composition of particular management function model (FS_WFM) according to the formal definition of enterprise management function (predefined as Elementary Management Cycle (EMC) [5]).

The major steps of problem domain analysis and knowledge acquisition are presented in Figure 3:

Step 1. Acquired problem domain knowledge is presented as traditional workflow diagram (VP_WFM), i.e. business process model (empirical one).

Step 2. Workflow diagram (VP_WFM) is transformed into P_WFM and F_WFM when separation algorithm is performed. Yet, in the transformation process logical gaps may occur. A logical gap is a semantic discontinuity between the elements of the problem domain model (for instance, workflow model).

Step 3. Logical gaps in the P_WFM and F_WFM models are identified by the algorithms of the P_WFM and F_WFM analysis and eliminated by the analyst. The application of these algorithms requires an additional analysis of the problem domain. The result of logical gaps elimination algorithms are P_WFM and F_WFM without logical gaps. In such eliminating process VP_WFM is also updated.

This process is called the first quality assuring cycle of computerized problem domain knowledge and it is based on formal enterprise management cycle model EMC defined in [7].

Step. 4. The algorithm defining functional composition for particular management function Fj is performed at this step domain knowledge analysis. Completeness of management function Fj (against the definition of EMC) is verified and validated. The possibly lacking activities are identified using the predefined structure of knowledge, i.e. Enterprise Meta-Model as a criterion. The algorithm A4 of functional composition indicates activities and flows, which possibly exist in the enterprise problem domain, but are not specified yet in the model (F_WFM) of business function Fj.

This process is called the second quality assuring cycle of domain knowledge acquisition process and it is based on formal enterprise management cycle model (EMC) defined in [5, 7]. The result of functional composition defining algorithm is validated model (FS_WFM) of definite management function Fj.

Step 5. The Enterprise Model of definite problem domain is refershed by verified and validated knowledge (i.e. presented in model FS_WFM) about definite management function Fj.

3.1 The Initial Workflow Model of Business Process

Primary knowledge about problem domain is acquired to initial BP model, represented as workflow model VP_WFM when designing Workflow Model of Business Processes (VP_WFM). The VP_WFM represents empirical information - user and

analyst knowledge about problem domain. The Workflow Model of Business Processes is represented using notation of traditional workflow model. The main components of traditional workflow model are *Actors, Activities* and *Flows*. In graphical notation *Activities* and *Flows* are signed by symbols without reference what nature (material or informational one) business process and flow belong to. In order to make the process of problem domain knowledge acquisition more effective it is advisable to modify traditional workflow model by establishing flows of two types: material flow and informational flow. The modified workflow model is called Workflow Model of Business Processes (VP_WFM).

Two types of VP_WFM flows are identified and used for VP_WFM decomposition into P_WFM and F. Each *Business Process* of VP_WFM, except initial and final ones, has material and (or) informational input and output. *Business Process* can be of either material or informational nature. A *Business Process* which is related to material flows is defined as business process of material nature, while a business process that is related to informational flows is defined as business process of informational nature.

The component *Business Process* of VP_WFM is defined as the sequence of organizational actions, which transform inputs into outputs. *Material Flow* is a material input and (or) output of business process, supplying material resources necessary to perform the process. Material input (output) of business process is not a mandatory element of each business process. The component *Information Flow* is informational input and (or) output of business process, intended to control it. VP_WFM actors are human, group of humans or organizational unit, which perform business process and

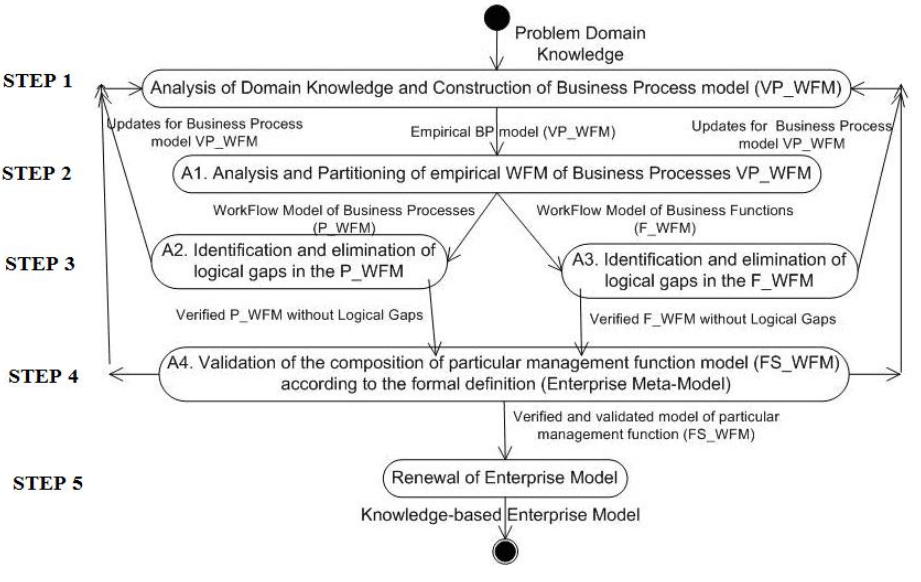


Fig. 3. The workflow modelling based elicitation of domain knowledge

are responsible for its successful performance. The prototype of VP_WFM modeling tool is developed on the base of MS "VISIO 2000" tool and MS "Access 2000" data base management system.

3.2 Model of Processes: Material Transformation Activities

The Workflow Model of Processes (P_WFM) specifies material processes (they are singled out of the VP_WFM, and have material inputs and (or) outputs) and actors, who implement them. A Process is a partially ordered set of steps, which can be executed to achieve some desired material end–result. A Process consumes material resources (it is an input of the process) and produces some material output – products or services. The internal components of the Process are sub-processes, tasks and operations. Definitions of material flow and actor of the model P_WFM are like that of the model VP_WFM.

3.3 The Model of Functions: Information Transformation Activities

Information Activities, Information Flows and *Actors* are the components of Workflow Model of Functions (F_WFM). *An Informational Activity* is enterprise function, or its component, which processes information flows when changing information input into information output. Each material process is controlled by at least one function, which consists of informational activities and information flows, linking these activities. Material transformation *Process* is totally controlled by enterprise function, while activity controls this process partly. Definitions of *Information Flow* and *Actor* in the Model of Functions (F_WFM) are like that in Model of Process (VP_WFM).

3.4 Principles for Separation of Information Flows and Material Flows

The initial verification stage of acquired domain knowledge is a process, which separates the empirical workflow model (the VP_WFM) into two workflow models: the P_WFM concerning the material flow transformations and the F_WFM concerning the information flow transformations.

When the VP_WFM is decomposed into P_WFM and F_WFM, the some inconsistency of problem domain specifications (informational gaps in the workflows) are identified. This is a step of BP model verification. For decomposition of VP_WFM into model of material transformation activities (P_WFM) and model of information transformation activities (F_WFM) the main three rules are applied. The first rule – if *Business Process* input and (or) output (specified in VP_WFM) are *Information Flows*, this *Business Process* will be specified as *Information Activity* with input and (or) output flows in the F_WFM. Second rule – if *Business Process* input and (or) output (specified in VP_WFM) are *Material Flows*, this *Business Process* will be specified as *Process* with material input and (or) output flows in P_WFM. Third rule – an *Actor* is specified as an *Actor* in the F_WFM activity or in the P_WFM. Figure 4 gives an example of decomposition of the some particular model VP_WFMp.

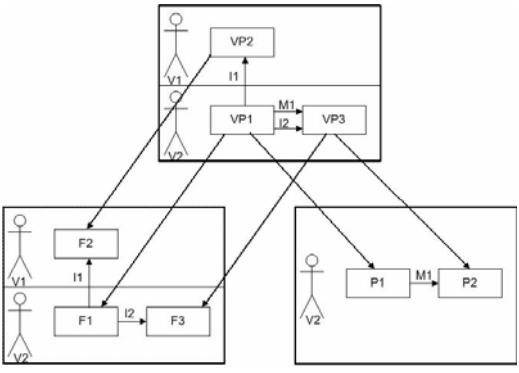


Fig. 4. Decomposition of empirical BP model

3.5 The Elimination of Logical Gaps

A logical gap is a semantic discontinuity among the elements of the enterprise management function, presented in the notation of workflow model. Identification of semantic discontinuity among elements of management function is based on the formal definition of Elementary Management Cycle [5], [7]. The logical gaps could appear when problem domain knowledge is acquired incompletely. To eliminate gaps of P_WFM, logical gaps detecting and eliminating algorithm is applied. Logical gaps of P_WFM are identified during the analysis of input and output flows of each material process. A logical gap in the P_WFM and F_WFM is identified if some *Process* or *Activity* is not related to input or output flow. Except the first and the last *Processes* of the workflow model each *Process* of the P_WFM must be related to at least one input *Material Flow* and one output *Material Flow*, in the same as each *Activity* of the F_WFM must be related to at least one input *Information Flow* and one output *Information Flow*.

The logical gaps elimination algorithm for model F_WFM is analogical to that of model P_WFM. The main difference in these algorithms is the nature BP model nodes (activities, processes) and relationships (flows): the activities and flows of model F_WFM are informational one, and flows and processes of model P_WFM are material one. If input and output of some activity in the FS_WFM are information flow "Process Output", incorrect type of activity (Impossible) is identified. The components *Activities* of the FS_WFM, according to composition of Enterprise Meta-Model, cannot have informational input and output flows of the same type. The components *Activities*, which have information input and output flows ("Process Output", "IP Input", "IP Output", "Process Input") of analogical type, can exist neither. If input of some *Activity* is "Process Output" and output of *Activity* is "IP Input", this *Activity* will be a component of management function, it is called *Interpretation* [5]. The component *Interpretation* of management function is set of rules, intended to transform information flow "Process Output" into "IP Input", which is prepared for IP processing. *Interpretation* is a necessary component of management function, because "Process Output" information flow can mismatch data format, determined for functional IP element input "IP Input". If input of *Activity* is "IP Input" and output of

Activity is “IP Output”, this *Activity* is component *IP* of management function. The component *IP* is mainly intended to control process of information processing and decision making. If input of Activity is “IP Output” and output is “Process Input”, this *Activity* is a part of management function called *Realization* [5]. The component *Realization* transforms “IP Output” data (processed in IP stage) into “Process Input” format (aimed to direct control of *Process*). *Activity* input “IP Input” indicates two possible types of outputs: “IP Output” and “Process Input”, while enterprise output “Process Input” indicates activities IP and Realization as well as information flow “IP Output” (which links IP and Realization). Activity input “Process Input” and output “Process Output” signal an error in F_WFM, thus such type of activity is impossible.

3.6 Knowledge-Based Model of Management Function

The result of functional composition verification algorithm is knowledge-based model of enterprise management function, represented in the workflow modeling notation as FS_WFM (Workflow Model of Functional Composition). The Meta-Model of the knowledge-based management function FS_WFM is presented in Figure 5.

The model FS_WFM specifies only one management function, which controls one or more processes, specified in the model P_WFM. In accordance with the internal structure of management function (defined by Enterprise Meta-Model), there are three types of F_WFM activities: Information activity of interpretation, information activity of processing and decision making (IP), Information activity of realization. Each activity of the model F_WFM can correspond to one of the above mentioned component parts of management functions. Algorithm, which defines functional composition of management function, determines what part of management function activities belong to and what material process do they control in the model F_WFM.

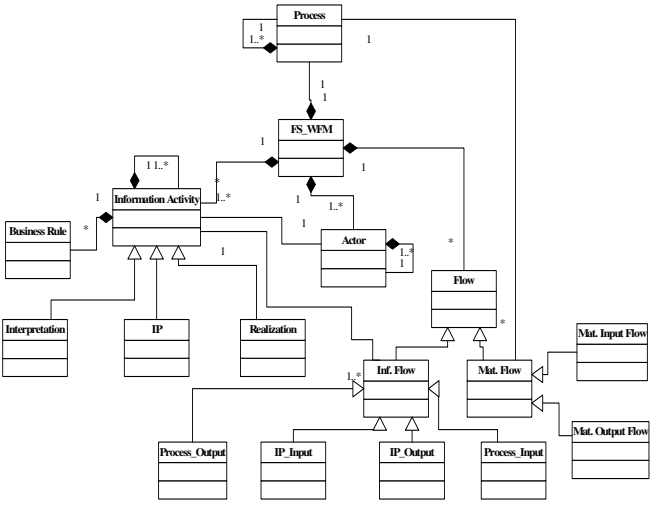


Fig. 5. Meta-Model of knowledge-based management function (FS_WFM)

4 Principles of Functional Requirements Generation

Traditionally, the UCM is aimed to specify functional requirements for particular task (information processing activity). Using identifiers of Enterprise Model elements, it is possible to single out knowledge, related to that task, and depict them according to UCM designing rules.

Enterprise knowledge repository of CASE system is an active "participant" of knowledge based IS engineering process. It is an extra source of domain knowledge besides user and analyst. In traditional IS engineering only user and analyst stand for the source of knowledge about problem domain. On the basis of this method, interactive user requirements (Use Case models) generating algorithms were created. They control user requirements specification process. Such opportunity (to control generation process) is ensured by CASE tool enterprise knowledge repository. Domain knowledge (stored in the repository) becomes a criteria, which is used to control decisions of user and analyst, i.e. such verification reduces human factor influence in the user requirements acquisition, analysis and specification stage of IS engineering. Principled possibilities of knowledge-based elicitation of user requirements (specified as UCM) are illustrated in Figure 6.

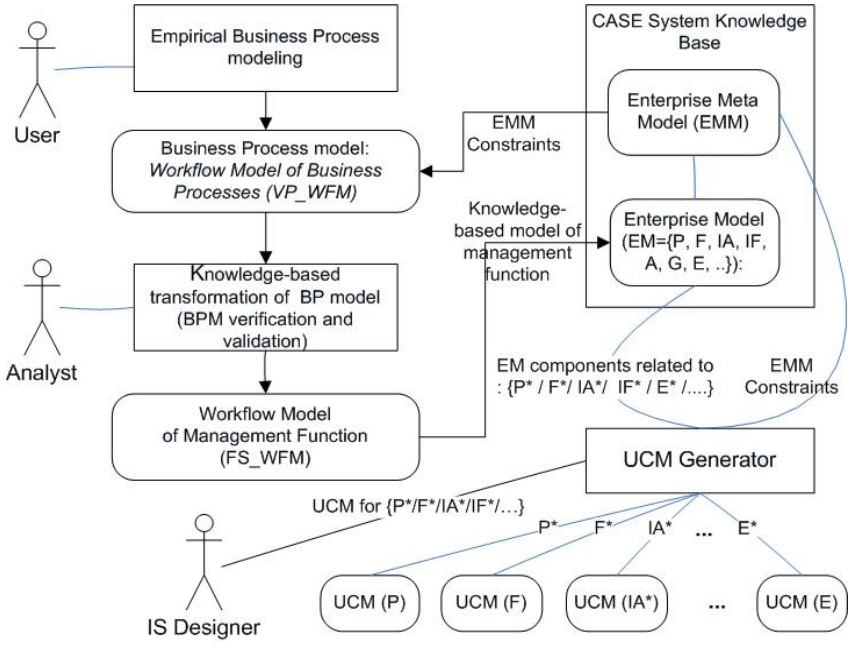


Fig. 6. Knowledge-based BP modeling and alternatives of user requirements generation

Notionally, UCM can be generated according to any user selected Enterprise Model component (Class of EM): EM= (Process (P), Function (F), Information Activities (IA), Information Flows (IF), Actor (A), Event (E), Goal (G), Business Rule,

etc.) [5]. In case of UCM generation for EM Class Process (P*), material processes (existing in definite EM*) and related actors are specified in such resulting UCM for Process (P*). This type of UCM is called the UCM of processes.

The UCM of function (see UCM (F*) in Fig. 6) is aimed to specify the components of definite enterprise management function and related actors. The components of Use Case model of Function are as follows: (*Function (F), Information Activities (IA), Information Flows (IF), Actors (A)*).

Likewise the UCM, generated for the EM Class Actor (A*), specifies material processes, functions and informational activities related with definite Actor (A*). The UCM generated for EM Class Goal specifies Functions, related to organizational goals, and their components (informational activities). The UCM could be generated as well for such EM Classes as Material flow, Information flow, Information Activity and Business Rule.

5 Conclusions

The problems of the knowledge-based BP modeling and user requirements design have been discussed. The Enterprise model is considered as the major source of knowledge in BP modeling and IS development. The formalized Enterprise model defines the structure and behavior of business enterprise, improve the process of specify requirements elicitation, IS design and implementation. The Enterprise Meta-Model is used as the "normalized" knowledge architecture to control the process of construction of an Enterprise model for the particular business domain. The usage of such Enterprise model facilitates the automation of the whole IS development process. Some work in this area has already been done [5].

This approach ensures the knowledge-based specification of user functional requirements, which are verified and validated against Enterprise model, acquired on the basis of formalized Enterprise Meta-Model. The knowledge-based requirements specification (Use Case model) generation principles are presented. The Enterprise Meta-Model is the source of formalized knowledge for Enterprise modeling and requirements specification. User requirements specification algorithms were created to generate Use Case models for user required components of Enterprise Model: enterprise management function, enterprise process, actor, goal and activity.

References

1. Dijkman, R.M., Joosten, S.: Deriving use case diagrams from business process models. CTIT technical reports series, 08 (02) (2002) ISSN 13813625
2. Dijkman, R.M., Joosten, S.M.M.: An Algorithm to Derive Use Cases from Business Processes. In: 6th IASTED International Conference on Software Engineering and Applications (SEA 2002), Cambridge, MA, USA, November 4-6, pp. 679–684. Acta Press (2002) ISBN 0-88986-323-7
3. GERAM Generalized Enterprise Reference Architecture and Methodology, Version 1.6.3. IFIP–IFAC Task Force on Architectures for Enterprise Integration (1999)
4. Gudas, S.: Žiniomis grindžiamos IS inžinerijos metodų principai. In: Konferencijos pranešimų medžiaga Informacinės technologijos 2004, Kaunas, Technologija, vol. T.2, pp. 713–717 (2005)

5. Gudas, S., Lopata, A., Skersys, T.: Approach to Enterprise Modelling for Information Systems Engineering. INFORMATICA 16(2), 175–192 (2005)
6. Gudas, S., Skersys, T., Lopata, A.: Framework for Knowledge-based IS Engineering. In: Yakhno, T. (ed.) ADVIS 2004. LNCS, vol. 3261, pp. 512–522. Springer, Heidelberg (2004)
7. Gudas, S.: Organizational System as a Hierarchy of Information Processes - Applications of Artificial Intelligence in Engineering VI. In: Proceedings of the 6TH International Conference on Artificial Intelligence in Engineering (AIENG 1991), Oxford, England, June 1991, pp. 1037–1050. Computational Mechanics Publications, Southampton (1991)
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.: Enterprise model based computerized specification method of user functional requirements. In: International conference 20th EURO mini conference Continuous optimization and Knowledge-based Technologies (EuroOpt 2008), Neringa, Lithuania, May 20-23, pp. 456–461 (2008) ISBN 978-9955-28-283-9
10. Mylopoulos, J.: Representing Software Engineering Knowledge. Automated Software Engineering 4, 291–317 (1997)
11. Molina, J.G., Ortin, M.J., Moros, B., et al.: Towards use case and conceptual models through business modeling. In: Laender, A.H.F., Liddle, S.W., Storey, V.C. (eds.) ER 2000. LNCS, vol. 1920, pp. 281–294. Springer, Heidelberg (2000)
12. Schekkerman, J.: How to survive in the jungle of Enterprise Architecture Frameworks, Trafford (2003) ISBN 1-4120-1607-x
13. Stephen, J., Kendall, S., Uhl, A., Weise, D.M.: MDA Distilled: Principles of Model-driven Architecture. Addison-Wesley Pub. Co., Reading (2004)
14. Vernadat, F.: UEML: Towards a Unified Enterprise modelling language. In: Proceedings of International Conference on Industrial Systems Design, Analysis and Management (MOSIM 2001), Troyes, France (2001), http://www.univ-troyes.fr/mosim01 (2001-04-25/27)
15. Ceponiene, L., Nemuraite, L., Vedrickas, G.: Separation of Event and Constraint Rules in UML&OCL Models of Service Oriented Information Systems. Information Technology and Control 38(1), 29–37 (2009)
16. Kapocius, K., Butleris, R.: Business rules driven approach for elicitation of IS requirements. In: WMSCI 2005: Proceedings of the 9th World Multiconference on Systemics, Cybernetics and Informatics, Florida, USA, July 10-13, vol. 4, pp. 276–281. International Institute of Informatics and Systemics, Orlando (2005)
17. Porter, M.E.: Competitive Strategy: Creating and Sustaining Superior Performance. The Free Press, New York (1985)

Market-Driven Software Project through Agility: Requirements Engineering Perspective

Deepti Mishra and Alok Mishra

Department of Computer Engineering, Atilim University,
Incek, 06836, Ankara, Turkey
deepti@atilim.edu.tr, alok@atilim.edu.tr

Abstract. Time-to-market and insufficient initial requirements are two major challenges that make requirement engineering for market-driven software projects different from bespoke software projects. These challenges can be resolved by using agile methods for market-driven software development as agile methods put emphasis on a dynamic approach for requirement engineering which work closely with an iterative release cycle. In this study, dynamic requirement engineering approach of Agile methods was used for the successful implementation of market-driven software (Supply Chain Management) project.

Keywords: requirement engineering, market-driven software, Agile methods.

1 Introduction

Requirement engineering for market-driven software development is different from customer specific software development. In Market-driven projects, requirement gathering is difficult as there is no distinct and defined set of users [1]. There are potential users, an imagined group of people who may fit into the profile of an intended product user [2], who can help in gathering some requirements. Often, requirements are invented by developers [3], based on strategic business objectives, domain knowledge and a product vision. All requirements cannot be known in advance before construction begins. In fact, there is a constant flow of requirements from various sources and these requirements needs to prioritized and managed during software development life cycle. This is similar to the requirement engineering phase of Agile methods. In Agile approach, development of requirements specifications is conceived as an incremental process, in which the stakeholders successively add requirements until getting to the specifications of the desired system [4].

When developing software for a market place rather than bespoke software for a specific customer, short time-to-market is very important [5][6]. It is important so as not to lose the market share to competitors and in case of probable delays, only high priority requirements are implemented in the current release. Low priority requirements are excluded from the current release and implemented in subsequent releases. Therefore, market-driven software products are often developed in several consecutive releases. This is in sync with the philosophy of Agile Methods which says software should be developed in an incremental and iterative way with high

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 103–112, 2009.

priority requirements to be included in initial releases and working software is seen as sign of progress.

Active participation of the user during development is one of the very important principles of Agile methods. Similarly, in order to succeed and capture the market with their market-driven software, organizations must have some means to get customer feedback early in the development process and thus minimize the risk of wasting valuable development efforts because of ambiguous and incomplete specifications.

We developed market-driven software (supply chain management software) with the motivation of Agile methods and the requirement engineering phase was done in an iterative way. We gathered requirements by conducting several sessions of interviews and workshops. Also, feedback from initial release helped in refining old requirements and gathering new ones. In this paper, we have discussed how the application of agile methods for market-driven software development resulted in the successful implementation of supply chain management software.

2 Literature Survey

Karlsson et al. [2] investigated current practices and challenges for market-driven requirement engineering in Swedish software development organizations in order to increase the understanding of the area of market-driven requirement engineering. They found many problems, some of them unique for market-driven projects and not applicable for customer-specific software projects. So, requirement engineering methods to develop customer-specific software may not be enough to support requirement engineering for market-driven software projects. This is also supported by Potts [3] that requirement engineering models and methods to develop bespoke software are not suitable for market-driven software development. He suggested some alternatives that would address those shortcomings for research and consulting communities.

In market-driven projects, eliciting requirements is difficult as there are no defined users but some potential customers. These requirements should be prioritized also as time-to-market is a key to succeed and capture the market and all gathered requirements may not be implemented within a specified time limit. Only high priority requirements are implemented in the current release and low priority requirements are left to be included in subsequent releases. Therefore, successful requirement engineering process for market-driven software development projects must have the ability to extract requirements from different sources. These different sources may give conflicting requirements so there must be some way to resolve conflicts among requirements and then prioritize and manage them. Lueke [7] presented a Structured Brainstorming and Evaluative Survey Technique (SBEST) for discovering, systematically gathering, prioritizing and implementing marketplace wants and needs while paying attention to any competition. SBEST yields the attributes of the ideal solution from the market's point of view. However, the process does not yield specifications for development of a product or service. Yeh [8] presented a market-driven requirement management process (REQUEST) that transforms systematically the many "voices of Customers" through various stages to a set of plan candidates by means of analysis,

validation, and prioritization. It tracks and relates original requirements to plan items and vice versa. Tuunanen and Rossi [9] developed a new RE method based on Critical Success Chain (CSC) method that includes top-down approach of planning and participation of information systems customers to get rich information. They extended CSC with customer segmentation and lead user concepts from marketing. They also constructed a support environment within Metaedit+ Meta CASE tool to present and manage requirements. Regnell [10] presented an industrial case study where a distributed prioritization process is proposed, observed and evaluated. A major objective of the distributed prioritization is to gather and highlight the differences and similarities in the requirement priorities of the different market segments. Various charts are proposed to present visually the disagreement between stakeholders and differences in satisfaction with a certain priority decision. These charts are intended to be used as decision support when determining what to implement in the coming release of a software package.

Sawyer et al. [5] pointed out that time-to-market is the overriding constraint for market-driven software development projects. When the project falls behind schedule, the preferred solution is to concentrate on meeting most critical requirements releasing the product on time. Other features can be added in later releases. Also, new requirements will also emerge when real users start using the software. So, requirement engineering process for market-driven software development has to be dynamic which must work closely with an iterative release cycle. They synthesized a number of good practices for requirement engineering for packaged software.

Also, there is a need to handle congestion in the requirement engineering process for market-driven software development which may occur when short time-to-market is combined with rapid arrival of new requirements from many different sources. Eliminating duplicity of requirements helps in dealing with congestion. Dag et al. [11] presented empirical evaluations of the benefits of automated similarity analysis of textual requirements, where existing information retrieval techniques are used to statistically measure requirements similarity. Host et al. [12] modeled a specific requirement management process (REPEAT) using discrete event simulation and the parameters of the model were estimated based on interviews with people from the specific organization where the process is used. Their aim was to investigate if simulation can help in exploring bottlenecks and overload situations in requirement engineering processes and to find changes to the process that may remove bottlenecks.

To compete in the market, the product should contain features or functions that do not exist in other similar software so developers tend to put more effort into inventing and implementing new functional features that are expected to improve the product but these new functions are useful only if users can use them easily. So, adding new functions is not only important but these newly added functions must be usable also. This is supported by Dag et al. [13] that although developers rely heavily on the number and the existence of new features, usability is recognized as a competitive advantage on its own. They presented results and experiences of an industrial case study that employs two known usability evaluation methods (a questionnaire and a heuristic evaluation) at a market-driven software development company inexperienced in usability.

The requirement engineering phase of our project didn't rely on any particular method mentioned above. We tried to learn lessons from all the above mentioned studies and used the parts which were applicable to our project. We initially

conducted a series of interviews with some prospective customers to gather initial information. This information was analyzed by a requirement engineering team (consisting of one business expert and team leaders of various software development teams) in a brainstorming meeting. Then this team conducted several workshops to refine, prioritize old requirements and also gather new ones. All the information collected here was also analyzed by the requirement engineering team in another meeting which helped in grouping the requirements and also developing high level architectural design of the proposed software system. Workshops and later brainstorming meetings were also part of Structured Brainstorming and Evaluative Survey Technique (SBEST) proposed by Lueke [7]. Also, the requirement engineering process for this market-driven software development was dynamic that worked closely with an iterative release cycle as recommended by Sawyer et. al.[5]. We did not stop gathering requirements when construction began. It continued even when construction was going on. We used simple natural language to store requirements in a repository. As the software was developed in several releases, we got the feedback from customers after every release. This feedback helped in adding new requirements, refining old ones and also gave an idea about the usability which is a very important attribute of any market driven software to have a competitive advantage in the market as supported by Dag et. al. [13].

3 Case Study

Supply Chain management was developed for the market. There were many potential customers. There were two perspectives of this project. One of them was engineering because of the business area and the solutions types. The other one is software that will provide those solutions as a high qualified, reliable, accurate and efficiently working software product. The engineering part was based on the operational-research area in the industrial engineering domain. All the optimization problems and solutions are defined in that topic theoretically. Their results have to be verified by using appropriate tools before they can be transferred as software solutions. Some of the characteristics of the project were:

- Large scale project
- Project complexity was high
- Acquaintance with the domain was less
- Insufficient requirement specification initially
- Quick release was important to have an edge in the market
- No defined set of customers (There were some prospective customers)
- There were multiple development teams and each team size was small. These teams concurrently developed different parts of the software.

The requirement engineering phase for supply chain management software was conducted in several iterations as shown in figure 1. According to Jiang et. al. [15], if the project size is large and project complexity is high, it is better to use a systematic technique to elicit, analyze, document, verify and validate requirements. Initially, we conducted interviews to gather requirements from many potential customers. We tried to observe their problems. Customers talked about their expectations from different

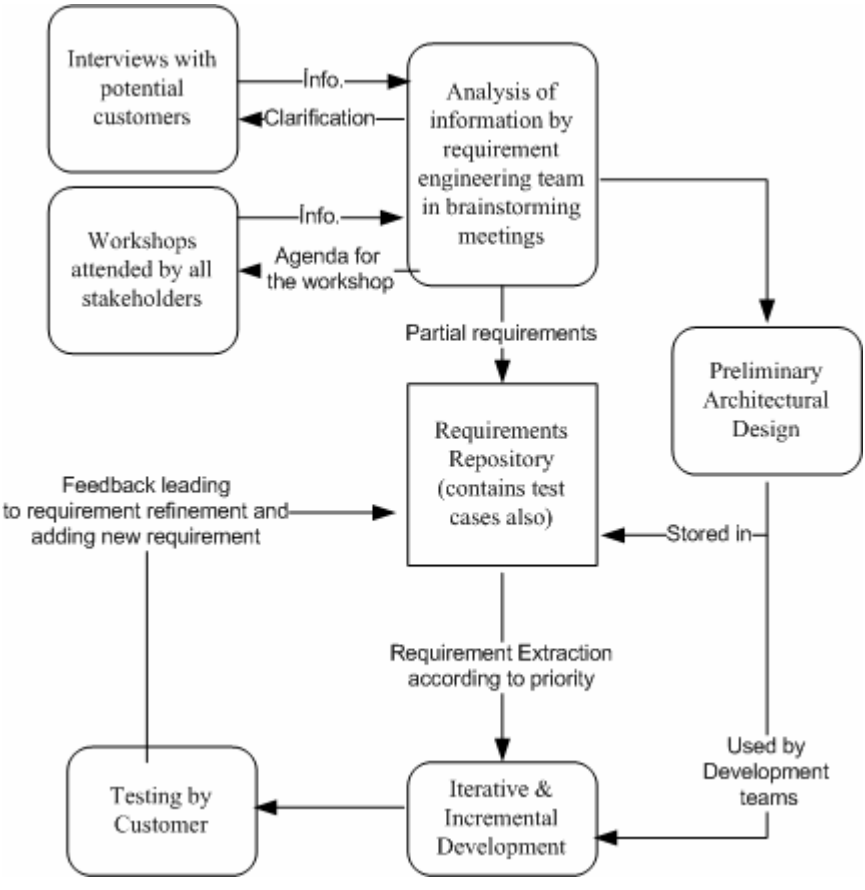


Fig. 1. Requirement Engineering Process for market-driven software development [14]

areas of the domain. This information was analyzed by a requirement engineering team during brainstorming meetings within the company and this analysis helped in filling the right location in the overall system working scenario. A requirement engineering team was consisting of team leaders of various small development teams and one business expert of supply chain management domain. One of the development team members, who had knowledge of this particular domain, played the role of business expert. In those meetings, business expert within the company pointed out the related problems and their solution that is defined in the concept of the working domain. In those early phases, there was an emphasis on defining a vision and scope, and identifying functions and features at a high level (such as just the names of use cases and features).

Later on workshops were conducted to refine, prioritize and resolve conflicts among requirements. These workshops also helped in determining cross-functional implications that are unknown to individual stakeholders and often missed or incompletely defined during stakeholder interviews. These workshops were held in a

series of sessions where each session often only lasted two to four hours, and were often attended by a majority of stakeholders.

These multiple sessions of interviews and workshops helped in obtaining initial requirements (partial requirements) which were used to initiate the development process but the requirement engineering process didn't stop here. Interviews and workshops for gathering requirements about parts of the software which were not clear are carried out in parallel with the development of different parts whose requirements were clear. Software requirements were stored in a repository along with their priority. At any point in time we gathered "could do", "should do" and "must do" requirements [16]. These requirements were stored in a centralized repository where they can be viewed, prioritized, and "mined" for iteration features. For each iteration, during the development, according to defined functionality of the iteration, the requirements were selected from that repository according to their priority, their use-cases and working scenarios were prepared by the domain expert and it was supplied to the development team.

Requirements were accessible to all team members, available to be enhanced and revised over time, and remain reasonably current to directly drive testing as they were implemented. There were other critical non-functional requirements also such as performance, portability, usability, reliability because of the constraints of the business domain. Some of these non-functional requirements were recognized by customers. Others like portability were recognized by the RE team.

Also, the requirement engineering team did the preliminary architectural design during the requirement engineering phase using the initial requirements. This is supported by Mead [17] that architecture modelling and trade studies are important activities that should take place during requirements engineering activities, not after the requirements have been defined. Software architecture must be considered during requirements engineering to ensure that the requirements are valid, consistent, complete, feasible etc.[17]. There were many development teams working concurrently on different parts of software. To avoid any confusion between these teams and also to have a common picture of what they were developing, RE team designed the core architecture of the system. This was the structure that specifies the whole system as major components, modules; collaborations, interactions and interfaces between them and the responsibility of each module. All the defined modules and components were drawn as black boxes and the interactions between them were represented by arrows. Development of each module was assigned to different teams as parallel tasks. The responsibilities and collaborations were defined clearly for each module (i.e. Controller, IO manager) and sites (DBMS, GIS, Application Server). This structure was also allowed to change as a result of customer's feedback from future iterations. Since it was a basis structure and there was collective ownership on that part by the team members, it was important to document that structure in order to be accessed easily. The diagrams, responsibilities, functionalities and scenarios were documented by the requirement engineering team. Object-Oriented design techniques were used for architectural design so as to increase applicability of the iterative and incremental development process because Object-oriented design provides modularity, minimum coupling

and maximum cohesion, thus a flexible structure. Ferrett and Offutt [18] also found in their study that object-oriented programs are more modular than procedural programs. Another benefit of using Object-oriented techniques was to define the tasks in parallel, since all modules provide encapsulation and a loosely coupled structure, each could be developed independently as a sub-product and then could be integrated easily because of well-defined interfaces. Once the core was built, team leaders who were part of the requirement engineering team along with a Business expert and actively participating in the requirement engineering process, returned to their respective teams and the development was done in parallel with multiple teams by using short iterations. Each team leader had a clearer picture and common vision due to the architectural design and could better convey and maintain that for the rest of the project. Further, each team leader acts as a liaison to the other teams. Also, after spending some close time with the other team leaders, there was improved communication between these teams. These team leaders played a dual role during the whole project. They took an active part during the requirement engineering process and also they were leading different development teams working in parallel on different parts of the software.

As this project was market-driven, it was not possible to get all the requirements by only conducting interviews and workshops with some potential customers. Requirements were not stable also. The rate of change in requirements was high so a more flexible approach, like prototyping, needs to be used to gather additional requirements and refine the old ones. Projects with higher requirements volatility require a more flexible approach [15]. Also, quick release of software was important to have an edge in the highly competitive market so we started developing software with the initial set of requirements by using an iterative and evolutionary approach. These iterations had short timeframes. These evolutionary prototypes of software were used to get feedback from customers. This feedback helped in extracting new requirements and further refinement of previous requirements.

4 Lessons Learned

Some lessons learned

a) The involvement of some prospective customers is important for the success of the project. Although we usually got the cooperation of customers, sometimes it posed some challenges too. Sometimes their ideas were entirely different from one another. Furthermore during a very crucial moment, they could not be present.
b) The presence of a business expert is also important for the success of the project. In this project, a development team member, who had worked in the past in a similar kind of project, played the role of business expert. Since this person was not entirely from this particular domain, we could not rely solely on his knowledge and that is why we saught the involvement of some prospective customers. But this business expert played a very crucial role in resolving the conflicts among requirements

and also prioritizing these requirements. Also, when the customers could not be present, this member filled that space which we think is quite significant for market driven project.

c) Initial interviews helped in deciding the scope of the problem. Also, they helped in describing the high level description of the requirements.
d) A brainstorming meeting among the requirement engineering team played the role of filter before workshops. Issues which could be resolved without the help of all stakeholers were solved here and therefore saved lots of time. Also they helped in setting the agenda for the workshops.
e) The role of workshop is very important. They not only helped in refining, prioritizing and resolving conflicts among requirements but also gelled all stakeholders. This instilled the feeling of a common goal among stakeholders and motivated them to work cohesively towards achieving it.
f) Architectural design, which was done by the requirement engineering team, helped in making a full and clearer picture of the entire system among all different development teams working on different parts of the system.

5 Conclusion

It has been established that Requirement Engineering for market-driven software projects is different than customer-specific software projects. Time-to-market and insufficient initial requirements are two major challenges in market-driven software projects. We handled these challenges by using agile methods for the market-driven software development. Agile methods advocate that software should be released in increments with higher priority requirements implemented in earlier releases and low priority requirements can be excluded to be implemented in a later release. Also, feedback from these releases help in refining old requirements and adding new ones. Agile methods put emphasis on the dynamic requirement engineering phase which work closely with an iterative release cycle. This process works well for market-driven software projects because it solves two major challenges written above.

Agile methods support gathering requirements in an iterative way as it is impossible to know all the requirements before the development begins. Having a complete set of requirements before construction begins is not a necessity if we use agile methods. In fact, with agility, the requirement engineering phase for market-driven software development can be made dynamic enough to gather and manage requirements from different sources during different timelines of the project.

The product can be released on time (as early as possible) using agility so as to have an edge in the market as time-to-market is another important challenge. This can be achieved by developing software in different releases with high priority requirements implemented in the first release. In future, this requirement engineering process can be applied to similar kind of projects and thus can be validated after comparing the results from those projects.

References

1. Sawyer, P.: Packaged Software: Challenges for RE. In: Proceedings of the Sixth Int. Workshop on Requirements Engineering: Foundations of Software Quality, Stockholm, Sweden, pp. 137–142 (2000)
2. Karlsson, L., Dahlstedt, S.G., Regnell, B., Natt och Dag, J., Persson, A.: Requirements engineering challenges in market-driven software development - An interview study with practitioners. Information and Software Technology 49(6), 588–604 (2007)
3. Potts, C.: Invented Requirements and Imagined Customers: Requirements Engineering for Off-the-Shelf Software. In: Proceedings of the second IEEE International Symposium on Requirements Engineering, pp. 128–130. IEEE Computer Society Press, New York (1995)
4. Lopez-Nores, M., Pazos-Arias, J., Garcia-Duque, J., Barragans-Martinez, B.: An agile approach to support Incremental Development of Requirements Specifications. In: Proceeding of the 2006 Australian Software Engineering Conference (ASWEC 2006) (2006)
5. Sawyer, P., Sommerville, I., Kotonya, G.: Improving Market-Driven RE Processes. In: Proceedings of the International Conference on Product-Focused Software Process Improvement (Profes 1999), Oulu, Finland, June 1999, pp. 222–236 (1999)
6. Novorita, R., Grube, G.: Benefits of Structured Requirements Methods for Market-Based Enterprises. In: Proceedings of International Council of Systems Engineering, sixth Annual International Symposium on systems Engineering: Practice and Tools (INCOSE 1996), Boston, USA (July 1996)
7. Lueke, E.: Gathering and implementing market-driven requirements. In: Proceedings of IEEE International Professional Communication Conference on Smooth sailing to the Future. IPCC 1995, September 27-29, pp. 122–126 (1995)
8. Yeh, A.C.: Requirements engineering support technique (REQUEST): a market driven requirements management process. In: Proceedings of the Second Symposium on assessment of quality software Development Tools, May 27-29, pp. 211–223. IEEE Computer Society Press, Los Alamitos (1992)
9. Tuunanen, T., Rossi, M.: Market driven requirements elicitation via critical success chains. In: Proceedings of 11th IEEE International Requirements Engineering Conference, September 8-12, pp. 367–368 (2003)
10. Regnell, B., Höst, M., Natt och Dag, J., Beremark, P., Hjelm, T.: An industrial case study on distributed prioritisation in market-driven requirements engineering for packaged software. Requirements Engineering 6, 51–62 (2001)
11. Natt och Dag, J., Regnell, B., Carlshamre, P., Andersson, M., Karlsson, J.: A feasibility study of automated natural language requirements analysis in market-driven development. Requirements Engineering 7, 20–33 (2002)
12. Höst, M., Regnell, B., Natt och Dag, J., Nedstam, J., Nyberg, C.: Exploring bottlenecks in market-driven requirements management processes with discrete event simulations. Journal of Systems and Software 59, 323–332 (2001)
13. Natt och Dag, J., Regnell, B., Madsen, O.S., Aurum, A.: An industrial case study of usability engineering in market-driven packaged software development. In: Smith, M.J., Salvendy, G., Harris, D., Koubek, R.J. (eds.) Proceedings of HCI International. Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality, vol. 1, pp. 425–429. Erlbaum, Mahwah (2001)
14. Mishra, D., Mishra, A., Yazici, A.: Successful Requirement Elicitation by Combining Requirement Engineering Techniques. In: IEEE ICADIWT 2008 conference, VSB-Technical University of Ostrava, Czech Republic, August 4-6, pp. 258–263 (2008)

15. Jiang, L., Eberlein, A., Far, B.F.: Combining Requirements Engineering Techniques – Theory and Case study. In: Proceeding of the 12th IEEE International Conference and Workshops on the Engineering of Computer-Based Systems (ECBS 2005) (2005)
16. Mishra, D., Mishra, A.: Achieving Success in Supply Chain Management Software by Agility. In: Münch, J., Abrahamsson, P. (eds.) PROFES 2007. LNCS, vol. 4589, pp. 237–246. Springer, Heidelberg (2007)
17. Mead, N.R., Shekaran, C., Garlan, D., Jackson, M., Potts, C., Reubenstein, H.B.: The role of software architecture in requirements engineering. In: Proceeding of the First International Conference on Requirements Engineering, April 18-22, pp. 239–245 (1994)
18. Ferrett L.K., Offutt, J.: An Empirical Comparison of Modularity of Procedural and Object-oriented Software. In: Proceedings of the Eighth IEEE International Conference on Engineering of Complex Computer Systems (ICECCS 2002), pp. 173–182 (2002)

On the Adaptation of Foreign Language Speech Recognition Engines for Lithuanian Speech Recognition

Vytautas Rudzionis[1], Rytis Maskeliunas[2], Algimantas Rudzionis[2], and Kastytis Ratkevicius[2]

[1] Vilnius University Kaunas faculty, Kaunas, Lithuania
[2] Kaunas University of Technology, Kaunas, Lithuania
vytautas.rudzionis@vukhf.lt

Abstract. This paper presents some of our activities trying to adapt the foreign language based speech recognition engines for the recognition of the Lithuanian speech commands. In recent years several quiet successful speech recognition engines became available for the most popular languages (such as English, French, Spanish, German, etc.). The speakers of a less widely used languages (such as Lithuanian) have several choices: to develop own speech recognition engines or to try adapting speech recognition models developed and trained for the foreign languages to the task of recognition of their native spoken language. First approach is expensive in time, financial and human resources sense. Second approach can lead to faster implementation of Lithuanian speech recognition modules into some practical tasks but proper adaptation and optimization procedures should be found and investigated.

Keywords: speech recognition, speech engine, transcriptions, phonetic.

1 Introduction

Speech recognition is long awaited mode for human computer interaction that needs to be introduced into many modern devices. In recent years pressure to develop faster voice based dialogue interfaces rises due to more and more widespread use of mobile devices such as mobile phones (many of them has capabilities significantly overcoming capabilities of traditional phones), tablet PCs, pocket PCs, etc. The characteristic property of similar devices is their portability which means reduced screen and keyboard. Small keyboard and screen introduces significant inconveniences for user to use them efficiently. Voice based dialogue interface often may become efficient substitute for traditional GUI based interface or in some cases may become single possible solution for mobile users.

Importance of voice based interfaces is stressed additionally by the fact that many business consulting companies predict that speech will remain dominant component in total data flow over wired and wireless networks in coming years.

From the advent of speech recognition research and the appearance of first commercial applications the main efforts were devoted to the recognition of widely used languages, particularly English language. The reason of such behavior is very clear – widely used languages have bigger market potential for practical applications. So

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 113–118, 2009.

looking at the general trend in the development of commercial speech recognition applications and tools for the development of speech recognition, such sequence could be observed: first version of speech recognition engine oriented to the recognition of English (and particularly US English) is released, then that system is supplemented with the engines for the other widely used languages (most often Spanish, French, German and several others) and sometimes but not necessarily with recognition modules of some other relatively widely used languages (in example Dutch, Italian, Turkish, Polish, etc.). Many other less widely used languages remains out of the scope of interest for the major speech recognition solution providers.

In such situation businesses and state institutions, in countries were such less popular languages are used as a main source of spoken language communication, faces a challenge of development of own speech recognition tools. Two major ways for solution are as follows:

- to develop own speech recognition engine from scratch;
- to adapt the foreign language based engine for the recognition of your native language.

The first approach has potentially higher capabilities to exploit peculiarities of selected language and hence to achieve higher recognition accuracy. But the drawback of such approach is that the providers of major speech technologies avoid the implementation of such languages in their products – high costs in the general sense of this word.

The second approach has the potential to achieve some practically acceptable results faster than developing entirely new speech recognition engine. Another advantage of this approach is potential to achieve faster compatibility with the existing technological platforms. Such advantage is often important for business customers, since they need to follow various technical specifications in order to guarantee consistent functioning of the enterprise. But this approach also requires careful investigation of the ways of adapting and optimizing adaptation algorithms.

This paper will present some of our activities trying to adapt speech recognition engine oriented to the recognition of spoken English for the recognition of several Lithuanian voice commands.

2 Expert – Driven and Data-Driven Approaches

The topic of the multilingual and crosslingual speech recognition is very important because of the reasons mentioned above. The importance of this research topic is expressed especially in Europe as this region is a multi-cultural society with many languages used in parallel. As the costs of generating a non-existing speech database and training acoustic models from that database can be very high, one possibility is to use crosslingual speech recognition. The idea behind this is to transfer the existing source acoustic models from source language to the target language without using speech corpora in that language and without full retraining of the speech recognition system.

Similarity measures used to transfer the source acoustic models to a target language can be divided into two major groups [4]: expert-driven methods and data-driven crosslingual speech recognition approaches.

In expert-driven methods mapping from one language to another is performed using human knowledge and is typically based on some acoustic-phonetic characteristics. One of the most frequently used methods is the use of so called IPA scheme. Expert knowledge of all included languages is needed. Such approach could be very difficult if many different languages were included in the system or must be used for the optimization. It was also observed that some subjective expert influence from the mapping can be expected.

IPA scheme [3] we will describe in more detail. For each phoneme in target language an equivalent phoneme in the source language was searched for. As an equivalent phoneme with the same IPA symbol is selected often. The ratio of the equivalent phonemes depends on the acoustic similarity of languages and on the number of phonemes in the all involved languages. In the case when IPA equivalent is non-existent in the target language the most similar phoneme according to IPA scheme is looked for. The search for the most similar candidate can be performed in horizontal or vertical direction through the IPA scheme. The main advantage of described scheme is that it can be applied without any speech material in the target language. Disadvantage of such approach is that expert knowledge should be obtained somehow and this knowledge also has subjective influence introduced by the expert. Data-driven crosslingual speech recognition approaches are based on data-driven similarity measures. In these methods the similarity measure is applied during mapping. Similarity measure itself is obtained from some data applying some algorithm. Data-driven approach with the phoneme confusion matrix will be described below in more details. The idea behind this method is that similar phonemes are confused during speech recognition by a phoneme recognizer. The basic characteristic of such recognizer is that it recognizes phoneme sequences instead of words from a vocabulary. For generating crosslingual confusion matrix, acoustic models of one of the source languages were applied on speech utterances of the target language. The recognized sequence of the source phonemes was then aligned to the reference sequence of the target phonemes. The output of this alignment was the crosslingual phoneme confusion matrix M. At this stage for each target phoneme f_{trg} the best corresponding source phoneme f_{src} should be looked for. As similarity measure, the number of phoneme confusions $c(f_{trg}, f_{src})$ is often selected.

So for each target phoneme f_{trg} the source phoneme f_{src} with the highest number of confusions c is selected in this schema. If two or more source phonemes has the same highest number of confusions it was proposed to leave decision for the expert which one of source phonemes should represent target phoneme f_{trg}. The same procedure could be applied if no confusions between source and target phonemes were observed.

The advantage of described data-driven approach based on a confusion matrix is that it is fully data-driven method and theoretically no subjective expert knowledge required (in practice expert knowledge is necessary to solve situation when same or similar confusions were observed).

3 Multiple Transcriptions Based Recognition

This paper will deal with the task of adapting Microsoft speech recognizer for Lithuanian speech recognition using two different vocabularies for 100 Lithuanian names

(first names and family names): it is expected that this vocabulary is less complicated since names are longer and task is related with the choice of possibilities;

Selection of vocabulary was determined by the practical potential of applications that could be developed from this vocabulary. The main characteristic of used approach is that multiple transcriptions were used for single command. The number of transcriptions used per word or command wasn't constant and was the case of some rough optimization (optimization was done by single speaker and developer trying to find some optimal performance level for that person and later those transcriptions were used for other speakers as well). And one more difference from that study was bigger number of speakers and utterances used in the experiments.

For Lithuanian first and family names database 33 different speakers uttered each name and surname once. The total number of combinations first name+family name was equal to 100 in this case (3300 utterances total).

4 Lithuanian Proper Names Recognition

Microsoft Speech Recognition engines were used as a basis for adaptation. Two speech databases were applied in these experiments: initial corpora and corrected corpora. Corrected database was freed from various inadequacies and mistakes that were present in the initial database. Between inadequacies and mistakes were pronunciation errors. Most of such pronunciation errors were situations when speaker used other phoneme than the phoneme present in the original family name, but still getting grammatically correct and often really existing name (in example, speaker said *Dalinkevicius* instead of *Danilkevicius*). Other errors were related with such problems as stammer near family name or some technical spoilage such as a truncated part of the word (most often at the end of the name).

Experiments were performed for male and female speakers separately and also for all speakers together. Table 1 show the results obtained in those experiments. Last row in the table presents results obtained using corrected speech corpora.

One of the most interesting observations was that recognition accuracy for initial and corrected databases was almost equal which show, that quite robust recognition system for small pronunciation errors may be developed (situation rather probable in applications, where we need to recognize people names). Detailed analysis showed that one speaker made even 19 errors when reading names and confusing at least one phoneme in the first name or family name, but all utterances were recognized correctly.

Table 1. The recognition accuracy of 100 Lithuanian names using the adapted transcriptions for the Microsoft Speech Recognition engine

Speakers	Correct, %	Insertions, %	Indeterminacies, %	Omissions, %
16 females	89.8	6.5	3.5	0.2
17 males	92.5	3.9	3.4	0.2
33 both genders	91.2	5.2	3.3	0.4
corrected	91.4	5.6	2.9	0.1

Another group of experiments using 100 Lithuanian names was carried out using clean and noisy or clipped speech. The aim was to evaluate robustness of recognizer for speech signal distortions. Only 5 speakers participated in this experiment and several SNR levels were selected and clipping coefficients were used. Table 2 shows results obtained in those experiments.

Table 2. Recognition accuracy of 100 Lithuanian names for different quality of speech signals (5 speakers)

Distortion	Correct, %	Insertions, %	Indeterminacies, %	Omissions, %
Clean	96.6	1.2	2.2	0.2
SNR 40dB	96.2	1.6	2.2	0.0
SNR 30 dB	91.6	1.4	7.0	0.0
SNR 20 dB	43.8	10.0	29.2	17.0
Clipped 0.3	93.0	4.8	2.2	0.0
Clipped 0.1	82.0	13.6	4.0	0.4

Looking at the table we see that the performance of the recognizer began to deteriorate significantly when the SNR level dropped below 30 dB and was in principle unacceptable at the SNR 20 dB. So the performance can't be treated as robust one looking from the SNR variations point of view. Using clipping coefficient 0.3 recognizer performance' dropped relatively insignificantly while clipping coefficient 0.1 resulted in much bigger loss in accuracy.

Table 3. Five names that result in the largest number of recognition errors in the 100 Lithuanian names recognition experiment

Indeterminacy errors		Insertion errors	
name	number of errors	name	number of errors
Gudas_audrius	17	Biaigo_sandra	16
Baublys_algis	12	Dolgij_andrej	16
Biaigo_sandra	6	Grigonis_audrius	11
Balcius_ernestas	6	Baublys_algis	10
Gailiunas_rytis	6	Giedra_nerijus	8

We've performed some recognition errors analysis trying to find the ways to improve the recognizer performance and find the ways to optimize adaptation procedure. There were 290 substitution and insertion errors (120 substitution and 170 insertion) in our first group of experiments and it was natural to expect that not all names will produce the equal number of errors. Here we've need to state that as an insertion error we treated situation, when recognition system produced some phonetic unit at the output of the system, which resulted in the name that wasn't present in the list of names (typical recognition system output in this situation was "*I don't understand you*"). Table 3 shows the 5 names that produced the largest number of errors in these experiments.

Looking to those results we see that 5 most confusing names produced almost 40% of all substitution errors and slightly more than 35% of all insertion errors. So the "concentration" of errors is big and more attention to the names that resulted in larger amounts of errors is necessary.

Detailed view to the most confusing names in these experiments showed that most of those names don't have difficult phonetic structure. The bigger number of errors obtained by the name *Gailiunas_rytis* may be explained by the presence of the name *Gailiunas_vytautas* in the same list. But most of the errors can't be explained straightforwardly. For example, name *Gudas* often was confused with the name *Butkus.*

5 Conclusions

This paper presented some of our activities trying to adapt the foreign language based speech recognition engines for the recognition of the Lithuanian speech commands. The speakers of less popular languages (such as Lithuanian) have several choices: to develop own speech recognition engines or to try adapting the speech recognition models developed and trained for the foreign languages to the task of recognition of their native spoken language. First approach is expensive in time, financial and human resources. Second approach can lead to a faster implementation of the Lithuanian speech recognition into some practical tasks, but the proper adaptation and optimization procedures should be found and investigated.

For 100 Lithuanian names recognition accuracy of more than 90% was achieved. These results show that the implementation of longer commands and transcription generation methodic proposed in [2] study were confirmed.

References

1. Lindberg, B., et al.: Noise Robust Multilingual Reference recognizer Based on Speech Dat. In: Proc. of ICSLP 2000, Beijing, vol. 3, pp. 370–373 (2000)
2. Kasparaitis, P.: Lithuanian Speech Recognition Using the English Recognizer. INFORMATICA 19(4), 505–516 (2008)
3. Zgank, A., et al.: The COST278 MASPER initiative – crosslingual speech recognition with large telephone databases. In: Proc. of 4th International Conference on Language Resources and Evaluation LREC 2004, Lisbon, pp. 2107–2110 (2004)
4. Zgank, A.: Data driven method for the transfer of source multilingual acoustic models to a new language. Ph.D. thesis, University of Maribor (2003)
5. Phoneme Table for English (United States), `http://msdn.microsoft.com/enus/library/bb813894.aspx` (retrieved December 19, 2008)

Analysis of the Share Price Bubbles in the Baltic Countries

Marius Dubnikovas[2], Vera Moskaliova[1], and Stasys Girdzijauskas[1]

[1] Vilnius University, Muitines str. 8, LT-44280, Kaunas, Lithuania
{stasys.girdzijauskas,vera.moskaliova}@vukhf.lt
[2] FBC "Jusu tarpininkas", A.Mickeviciaus str. 29, LT-44245, Kaunas, Lithuania
marius@jt.lt

Abstract. The last two years where marked with the formation of a number of financial bubbles and their bursts in various markets of capital, real estate and raw materials. The years 2007-2008 became a large scale trial both for professional and home economy that unavoidably was related with finance. Both bubble formation and burst determined the damage of the clients' global confidence in financial institutions and confounded their expectations. It caused corrections in the economic growth prediction since a number of countries found themselves on the verge of financial crisis. Three Baltic states that were earlier called the tigers of the European growth had encountered the described situations with the special difficulty – the indices of consumer and business confidence decreased by ten years; the growth of gross internal product decreased to minimum, or even acquired the features of recession, and the capital markets lost about one third of capitalization. The paper aims at the analysis of the alteration of the share market in the three Baltic countries – Lithuania, Latvia and Estonia – during the last nine years with regard to price bubble formation and bubble bursts. The research was carried out with the use of the comparative analysis of the scientific literature, the method of mathematical analysis and generalization. The analysis revealing how the three capital markets reached the level of capital saturation and when it manifested itself was performed with the computation program "Loglet Lab2".

Keywords: share price bubble, logistic growth model, Baltic stock exchange, index.

1 Introduction

Sometimes finance markets, similarly to other constituent parts of the economic structure, demonstrate an 'unnaturally' rapid growth of prices. It is considered that such upturns which last for several months or even years do inevitably end in a sudden drop to their 'true level'. The mentioned phenomenon is called the price bubble or crisis, and occurred in the Baltic countries as well. The dictionaries of financial terms describe the price bubble theory as "a theory under which security prices sometimes move wildly above their true values, or the price falls sharply until the 'bubble bursts'. It is also possible for a bubble to deflate gradually" [1].

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 119–129, 2009.

Each financial bubble burst results in heavy economic outcomes. Therefore this financial phenomenon deserves a special concern of researchers. In other words, the analysis of the errors of the past might help to avoid them in the future.

The paper aims at the analysis of the indices of the Baltic stock exchanges by focusing on the functions of the limited growth (or logistic functions) that describe the process of the capital accumulation (or growth).

The specificity of the logistic function lies in its limited growth aspect. To say more, it undergoes alteration exclusively within a described interval: from zero to a particular (maximum) rate. The logistic growth is a characteristic feature not only with respect to capital but, actually, to any population whose rate of growth is proportional to its size.

On the whole, the logistic models have been widely applied for the investigation of the biological systems. In the field of economic inquiry, they have been seldom applied – only single attempts at the analysis of the economic systems have been discovered by the authors of the paper [3], [4]. The main drawback of such models is that they do not offer the growth function expressed in compound interest. In Lithuania, the exploration of the mentioned problem started in 2002 [5].

By performing the analysis of the index alteration of the Baltic and employing information technologies it will be determined whether the explored logistic functions might be applied for the analysis of the practical data. The paper aims at determining whether the capital markets of the Baltic countries were faced with the price bubble burst situation and whether they reached the limited capital saturation that most probably provoked the sudden financial slumps.

The objective of the paper is to find out whether the capital markets of the Baltic countries have reached the level of the limited capital saturation by employing the logistic functions.

The object of exploration is the history of the indices of the Baltic countries in 2000-2009.

Methods of research: analytical, also embracing the methods of comparative and mathematical analysis and econometric calculations.

2 Logistic Modeling of Investments

Excluding rare exceptions, the contemporary economic research and specifically, in the field of investment postulates that economic growth is unlimited. While actually, sooner or later each growth gets exhausted. It is observed in the course of nature. The models analyzing the growth of populations in biology were worked out more than a hundred years ago. The cyclic development of economy in various regions and states confirms that the economic growth is limited as well. A recently created and developed theory of the logistic management of capital at Vilnius University, Lithuania, fits well for the description of the limits of the economic growth and the causes of the economic bubble formation, and might provide the researchers and patricians with the proper tools – the logistic growth models – to formalize these processes [10], [11].

Usually, the analysis of the growth of capital means that there is a particular **Investment Capacity** (or range) of the limited size which the given capital might

occupy. As a rule, the invested capital fills only a part of investment capacity. This part of capacity will be defined as **Investment Coverage**. A residual free part of capacity is intended for the capital growth and will be called **Resources of Growth**. Occasionally, investment capacity can be equal to capacity of the whole economic structure [3]. Consider:

Investment Capacity = Investment Coverage + Resources of Growth

The relation between variables in the logistic model is schematically shown in Figure 1.

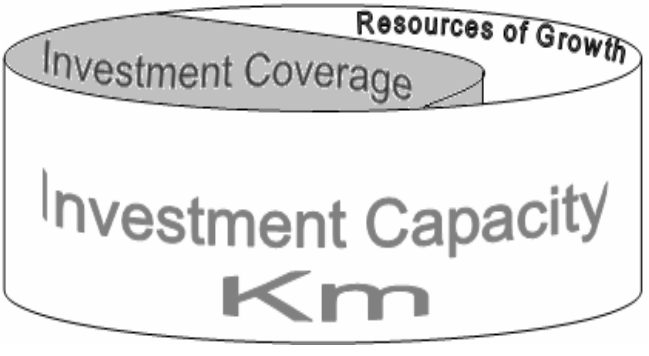


Fig. 1. The relation between variables of the logistic model

Investment capacity is limited, and with an increase of investment coverage the growth resources are diminishing. Therefore investment capacity limits the growth of investments. When investments are approaching the capacity limits, the economic bubbles begin to burst [9]. The bubble is formed when Investment Coverage increases in the fixed Investment Capacity and thus Resources of Growth decrease. In this situation, efficiency of investments, or logistic internal rate of return increases very sharply. Such behavior of the system causes the formation of the so called *bubble effect*.

It has been proved that the bubble can provoke crises (i.e. increase inflation, etc.) in the whole economy, into which certain capital is integrated. Hence it is necessary to emphasize that it is not inflation that causes the formation of the bubble, but vice versa – the forming bubble initiates and increases the processes of inflation.

The worked out logistic theory of capital management shows how to avoid the phenomena of overheated economy, or how to mitigate its fatal consequences. For this purpose, it is necessary to enlarge the capacity of capital. Seeking to escape the price bubbles, the investment capacity should be extended through globalization and by entering the new markets (an extensive mode), or through an implementation of technological innovations (an intensive mode). It is obvious that the second mode is more perspective.

Most frequently, in the cases when various financial problems occur in relation to payments or cash rate at the given moment of time, or when it is urgent to model the capital price, investments or any other cash flows, the present or future value of

capital is calculated. As a rule, such calculations are based on the so called formula of compound interest [2]. Consider:

$$K = K_0 \cdot r^t \tag{1}$$

here: K_0 is the present capital rate; K expresses the future capital rate, or the capital rate at the t moment of time; r describes the coefficient of accumulation rate; ($r = 1 + i$ here: i is interest rate); t is the duration of accumulation expressed in time units that are fixed in interest rate. Sometimes Equation (1) is called an exponential function of capital accumulation.

Traditionally, Equation (1) is used to calculate the growth of capital (population, products, etc.). However, such calculation may be performed only until the capital growth is not restricted by external factors [6]. Capital cannot increase at an equal rate endlessly, the more so if the system is completely or partially closed. When growing within such a system, it exhausts the limited resources in its environment. In other words, it enters into self-competition, which diminishes its growth – the system gets 'satiated'.

It is assumed that in the given environment, capital may increase up to a certain limit (in the given environment, only a particular amount of capital not larger than the determined one may be invested). The maximum rate of growth is K_m. Then the interval of the capital alteration or capital growth (relatively, it may be considered as an areal or space of growth) is as follows:

$$K_0 \leq K \leq K_m .$$

The growth of capital will be described by the logistic function of growth [5]. Consider:

$$K = \frac{K_m \cdot K_0 \cdot r^t}{K_m + K_0 \left(r^t - 1\right)} \tag{2}$$

here: K_0 is the present capital value; r defines the accumulation rate coefficient and t is time expressed in the same units as the time estimated in the interest rate of growth (in most cases, it points to the whole periods of the interest rate re-calculation).

It should be noted that, if the maximum value of the product K_m increases and approaches infinity $(K_m \to \infty)$, i.e. if for Equation (2) the limit $\lim_{x \to \infty} K$ will be calculated, then, as it might have been expected, Formula 2 will turn into an ordinary rule of compound interest (1). Then the formula of compound interest (1) will make a separate case of the logistic accumulation function (2), where the maximum capital rate K_m is extremely high.

On the basis of the studies concerning the logistic growth models the explanations of the formation of the stock market bubbles will be extended. Usually, in the analysis of capital price, investments, or other money flows the present value or future value of capital is calculated. The logistic present value may be expressed by the following equation [3], [7], [8], [9]. Consider:

$$K_0 = \frac{K_m \cdot K}{K + (K_m - K) \cdot r^t} \tag{3}$$

Here: K_0 is the present rate of capital; K describes the rate of capital at the time moment t; r denotes the rate of growth accumulation with the interest rate I; t is the time of accumulation in time units fixed in the interest rate. Actually, the described expression is the formula of logistic discount.

The logistic function of accumulation (2) is differentiated with respect to the coefficient of the accumulation rate r. The dependence of the logistic capital accumulation rate is found out from the coefficient of the accumulation rate r:

$$\frac{dK}{dr} = \frac{K_0 \cdot K_m \cdot r^{t-1} \cdot t \cdot (K_m - K_0)}{\left(K_m + K_0 \cdot \left(r^t - 1\right)\right)^2} \tag{4}$$

Figure 2 shows the dependence of the capital accumulation rate on the interest rate, when the rates of marginal capital vary, the value of the initial capital is equal to 1 (K_0 = 1) and duration of accumulation is equal to ten (t =10).

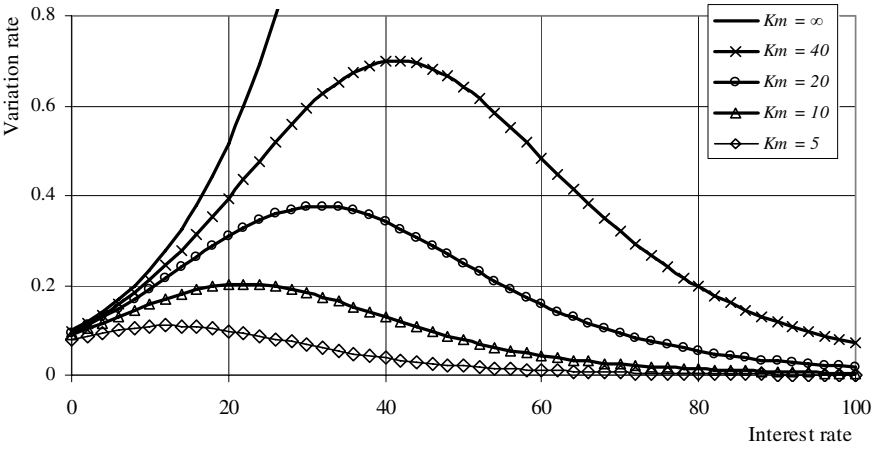


Fig. 2. Dependence of the capital accumulation rate on the interest rate when the rates of marginal capital differ; t = 10 and K_0 = 1

The logistic model demonstrates the economic growth under constraints. The pressure of constraints starts after having reached the peak of the diagram representing the growth rate (called the marginal growth rate) and then is going down which shows the slowing down rate of the economic growth, finally approaching an economic crisis. In fact, only the rapid progress in research and technologies may help to solve this problem allowing the entrance into a new stage of the economic development based on the implementation of the new transparent, efficient and resource-saving technologies.

3 The Situation in the Baltic Countries

The capital markets of Lithuania, Latvia and Estonia have a fifteen-year history. However, during this short period of time they experienced several vital conversions – the Baltic countries entered the NATO and the European Union; in 2005 they joined the OMX, i. e. the North Europe Security Trade System.

Despite an expanded market accessibility to investors these three stock exchanges have remained rather limited and enclosed. First and foremost, small market capitalizations (on an European scale) determined that there market liquidity was insufficient to receive large capital; secondly, the frozen list of the quoted companies did not allow to expand the markets.

An assumption has been made that the capital markets of the Baltic countries were limited enough to reach the level of limited capital saturation, which later on caused the price bubble bursts and initiated the market crises.

In 2008, the indices of the three Baltic countries approximately lost about 2/3 of their rate and thus considerably exceeded the world scale of crisis when the average drop of 70 most popular indices reached 40%. Due to the data of the stock exchanges, the main index OMXV (of Vilnius Stock Exchange) decreased to 65.14%; the main index OMXR (of Riga Stock Exchange) dropped down to 54.43% and the main index OMXT (of Tallinn Stock Exchange) dropped down to 62.98%.

The differences having been considered, each Baltic market was examined individually. The computation program LogLet was used to determine whether the rate of potential capital (i. e. the point of capital saturation) was reached and where it was likely to be fixed.

3.1 Vilnius Stock Exchange

The data of the Vilnius Stock Exchange index OMXV that covered the period of 2000–2009 has been thoroughly examined in Figure 3. The index break served as a signal of the formation of the bubble whose burst later on caused the crisis. The economic logistic analysis shows that when the market exhausts its growth possibilities, i. e. approaches the rate of potential capital (the limit of growth), an accelerated increase of invested capital return takes place. It evokes illusory optimism among the investors, thus encouraging people to get interested in this alternative way of capital management.

Such moves were observed both in 2005 and in 2007. In the first half of 2006 an obvious equity price correction was carried out on the basis of the apprehension for the future financial situation when the rapidly growing prices for raw materials and especially for crude oil initiated the growth of inflation.

Meanwhile, in the second half of 2007 the information about the price bubble burst in the real estate market was received. In fact, an obvious chain reaction in the price bubble bursts took place within the economic structure.

The graph shows that a few stages of the increase of the investors' optimism occurred in Vilnius Stock Exchange which ended in the price bubble burst. It should be stressed that the growth of index above the line of limited capital saturation indicates the price bubble formation.

Among the reasons causing bubble formation one of the most important was the limitation of the market amount. When the trade list is not expanding with the additional quoted companies, the potential (i. e. limited) capital rate is reached. The very fact that no new companies appear on the market determines the limited choice open for the investors as well as the limited amount of the potential investments. It means that the new investors are forced to 'buy' the former investors as well and thus against their will they do increase the distance between the share market prices of the

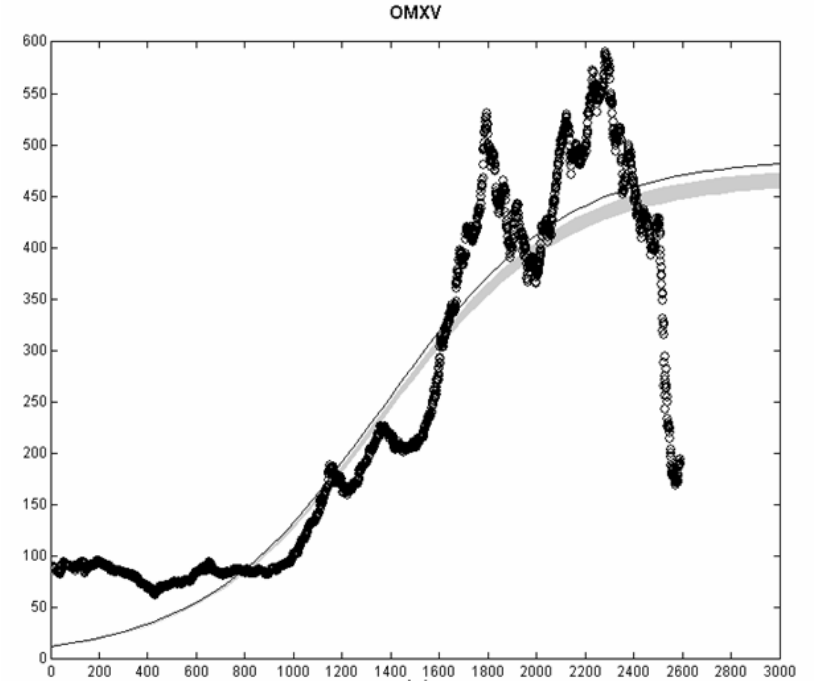


Fig. 3. Dynamics of the Vilnius Stock Exchange index and the curve of capital saturation

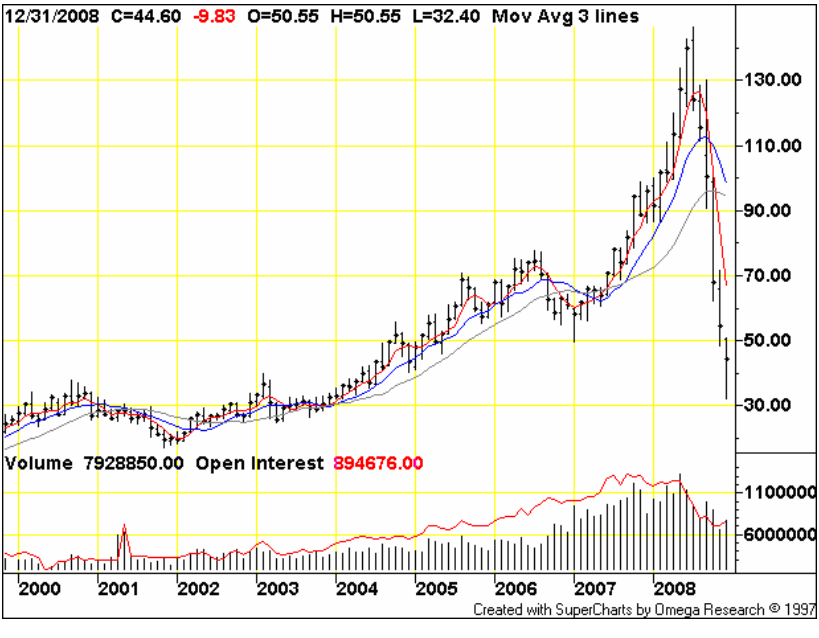


Fig. 4. Dynamics of crude oil prices

quoted companies and their real rate. The second reason is the debt capital increase within the market. It shows that, on the one hand, with the growing popularity of investment the market did not increase its share amount; on the other hand, the market was gradually filled with a rising amount of capital, a considerable part of which was made of debt capital. As the data of the Lithuanian Securities Commission embracing the three quarters of the year 2008 shows, every fourth deal in Vilnius Stock Exchange was financed with debt capital.

It should be also noted that the market growth observed in 2005 was stimulated by several essential changes in the Baltic (and hence the Lithuanian) capital markets. In that very year, the Baltic countries became the members of the European Union and the members of the OMX (i. e. the stock exchange net uniting all the stock exchanges around the Baltic Sea).

With their membership in the European Union and OMX, the Baltic markets attracted more attention that might be explained in the following way: the markets became open for the investment funds which could promote investment exclusively in the region around the Baltic Sea. It had certainly determined the growth of capital, since the funds had to invest at least a small amount of their capital in order not to leave a vacant field in the sample of their possible investments. Actually, it shows a direct link with the above discussed theory – the amount of invested capital grew rather rapidly when it was invested into the non-expanding market. The sample remained the same, but the amount of investments increased. Consequently, the markets reached saturation and even crossed the limits of the potential possibilities for capital growth.

The OMX membership formed a new sample of investors whose investment was done only within the OMX Net. Hence the Baltic countries still increased the round of their potential investors without increasing the depth of investments. Therefore the investment liquidity started growing on the base of its initial sample.

The Lithuanian brokerage company *Jusu tarpininkas (Your Broker)* agreed to provide data on the alteration of the number of clients who participated in the market and were actively involved in the Inernet access. Consider Figure 5:

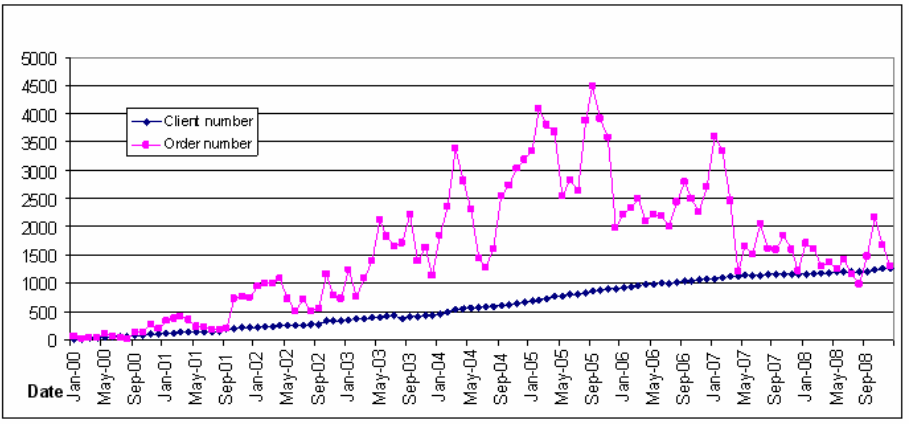


Fig. 5. Dynamics of client and order number in FBC *Jusu tarpininkas (Your Broker)*

Figure 5 also confirms that in 2005 the high dynamics of investment service was observed. It provoked a growing amount of additional capital in the market. It should be stressed that the growth of the number of clients was followed by a similar level of growth of the number of orders, which testifies to the growing activity of the investors and their growing market concern.

3.2 Riga Stock Exchange

Similar financial moves may be observed in Riga Stock Exchange that demonstrates similar tendencies. The graph below (Fig. 6) shows that there were two similar attempts at breaking away from the curve of capital saturation that took place at the same time as in the Lithuanian capital market. Both attempts ended in the index downgrade corrections. The OMXR index reduction in 2007-2008 indicates an obvious bubble burst. Consider:

It is interesting to note that the level of capital saturation of the Riga Stock Exchange index with the current population of the quoted companies was formed at point 681.

3.3 Tallinn Stock Exchange

The index of Tallinn Stock Exchange indicates similar tendencies (Fig. 7) as well which allow to consider that the indices of the three Baltic countries demonstrate a strong correlation and thus support the popular attitude claiming that these three markets are often regarded as one larger market.

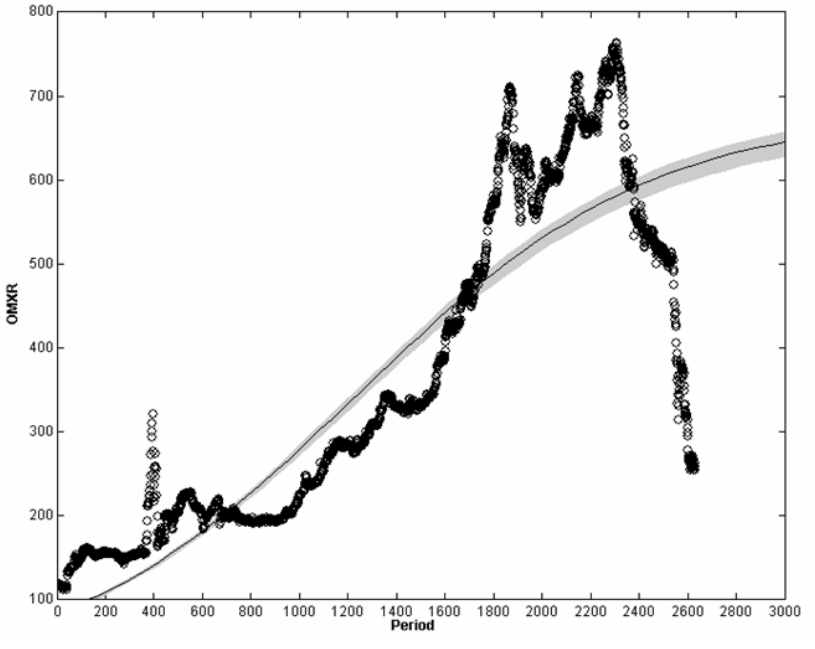


Fig. 6. Riga Stock Exchange index OMXR dynamics and its saturation curve

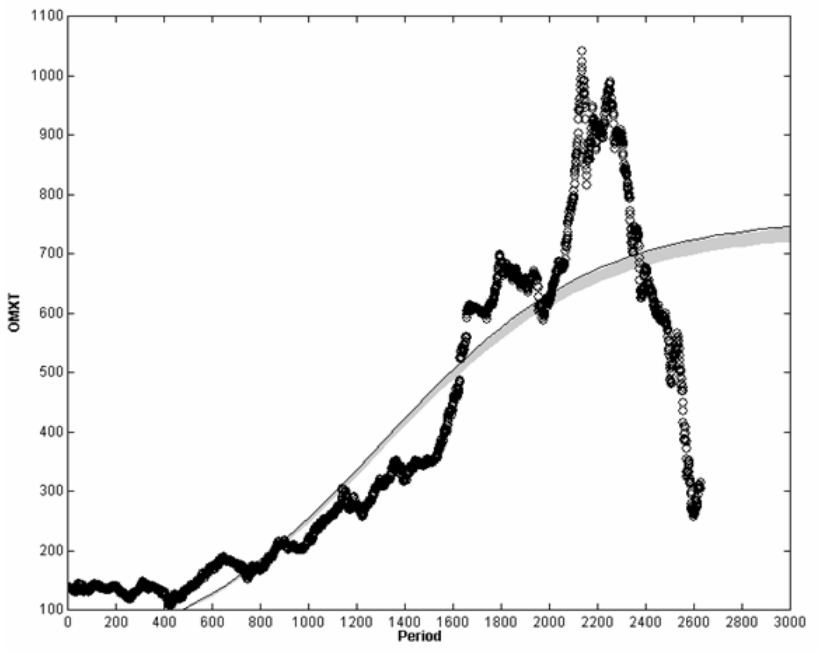


Fig. 7. Tallinn Stock Exchange index dynamics and its saturation curve

It should be stressed that Tallinn Stock Exchange is distinguished among the three Baltic countries in that the float of index in 2005–2006 was inconsiderable and thus did not indicate a marked bubble formation. Nevertheless, in 2007 a very strong break away from the fundamental rates was observed thereby much more obviously demonstrating the bubble formation.

4 Conclusions

1. The logistic economic analysis shows that in the stock exchanges of the Baltic countries the price bubble formation occurred and resulted in the bubble bursts. The capital markets collapsed under the pressure of their own weight. Rather self-contained capital markets demonstrated that such bubble formation was caused by the flow of additional money resources. All the three Baltic capital markets show a serious proof that market saturation (i. e. approaching limited capital) determines bubble inflation.
2. All the three Baltic countries demonstrated similar tendencies and this confirms expectations to consider them as a single larger market.
3. To avoid such undesirable situations in the future the market expansion is urgent. To escape bubble inflation it is necessary to introduce the new investment objects, or with the introduction of investment instruments, to push capital behind the limits of the market. Only thus market saturation will be escaped.

4. The financial crisis in the Baltic stock exchanges was caused by the burst of the share price bubble. The economic logistic analysis demonstrates that two conditions are necessary for the bubble formation: the fundamental and the psychological one. The first condition is related with the exhaustion of growth resources; the second one reflects the psychological inclination to earn much money. Therefore the process of bubble formation undergoes two stages: the first, or fundamental stage occurs when due to an exhaustion of growth resources the market starts increasing its capital return (i. e. gives a signal to the market participants about the growing returnability), and the second, or psychological stage occurs when the participants experience an ardent desire to invest profitably and earn much money. The first condition guarantees the birth of the bubble; the second one determines its size.
5. The Baltic stock exchange indices suffered a stronger collapse than in other countries, which confirms the inertia of the small markets and an adequate degree of risk.
6. To understand the lessons of the past financial events further inquiry into the theory of logistic capital should be conducted. The methods of logistic analysis should be employed to avoid financial crises, or at least to soften their consequences.

References

1. The Free Dictionary, http://financial-dictionary.thefreedictionary.com
2. Bodie, Z., Kane, A., Marcus, A.J.: Essentials of Investments. McGraw-Hill, New York (2001)
3. Shone, R.: An Introduction to Economic Dynamics. Cambridge University Press, Cambridge (2001)
4. Sterman, J.: Business Dynamics: Systems of Thinking and Modeling for a Complex World, p. 944. McGraw Hill, New York (2000)
5. Girdzijauskas, S.: Logistiniai (ribiniai) kaupimo modeliai (Logistic (Limited) Models of Accumulation. Information Sciences 23, 95–102 (2002)
6. Merkevicius, E., Garsva, G., Girdzijauskas, S.: A Hybrid SOM-Altman Model for Bankruptcy Prediction. In: Alexandrov, V.N., van Albada, G.D., Sloot, P.M.A., Dongarra, J. (eds.) ICCS 2006. LNCS, vol. 3994, pp. 364–371. Springer, Heidelberg (2006)
7. Girdzijauskas, S.: The Logistic Theory of Capital Management: Deterministic Methods; Monograph No.1. Transformations in Business & Economics, 2008 7(2) (2008)
8. Girdzijauskas, S., Cepinskis, J., Jurkonyte, E.: Modern Accounting Method in Insurance Tariffs – Novelty on the Insurance Market; Technological and Economic Development of Ecnomy, vol. XIII(3), pp. 179–183. Technika, Vilnius (2007)
9. Girdzijauskas, S., Pikturna, A., Ivanauskas, F., Merkevicius, E., Moskaliova, V.: Investigation of the Elasticity of the Price Bubble Functions: Continuous Optimization and Knowledge-based Technologies. In: 20th EURO Mini Conference (EurOPT 2008), Neringa, Lithuania, May 20-23, pp. 131–136. Technika, Vilnius (2008)
10. Girdzijauskas, S., Streimikienė, D.: Logistic Growth Models for the Analysis of Stock Markets' Bubbles. In: The 2008 International Conference of Financial Engineering. Lecture Notes in Engineering and Computer Science, pp. 1166–1170 (2008)
11. Streimikienė, D., Girdzijauskas, S.: Logistic Growth Models for the Analysis of Sustainable Growth. Transformations in Business and Economics 7(3)(15), 218–235 (2008)

Data Quality Issues and Duel Purpose Lexicon Construction for Mining Emotions

Rajib Verma

PSE—EHESS

Abstract. Mining emotions over the Internet has seen limited use even though it has important research implications in fields such as applied econometrics, the interdisciplinary study of happiness and well-being, and for various applications in customer relationship management, finance, marketing, human resources, and managerial science.

A key ingredient to making progress in these areas is the development of an emotion specific lexicon, one that can capture intensity and select relevant sentiment laden texts from online sources. An approach to doing this is developed, issues relating to data quality are pointed out, and methods to overcome them are explained.

Justifications for constructing the lexicon are given using state of the art empirical results and research. Then a 10 step algorithm that populates a lexicon using a hybrid procedure (thesaurus-corpus based) is developed. It captures sentiment no matter how it is expressed, and balances issues of speed, cost, and data quality.

Keywords: Word list, domain dependent, emotive, intensive, thesaurus, corpus, data quality.

1 Introduction

A greater understanding of individual and collective feelings is now possible because of the increased use of the Internet by individuals and organizations. Given that online communication and publication is virtually instantaneous, new media provides an opportunity to gather data that is much more current than it is in traditional sources. With the increase in computing power, and the advances in information retrieval and NLP technologies, a new world of inexpensive and undiscovered data is waiting to be used.

Mining complements both qualitative and quantitative methods. As such, it is apt for applications in a wide variety of disciplines in commerce and economics. These include human resources, econometrics, marketing, customer relationship management (CRM), finance, and managerial science. While this type of research has not been done extensively for research in business, econometrics, and well-being, there is reason to believe that it will be a promising area because similar work has been done in finance, politics, and other fields with positive results.

The potential is enormous. Due to the speed and immediacy with which content appears online, this approach provides a means of assessing the emotional states of a

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 130–138, 2009.

large and uncoupled population in real time. Ultimately, this will lead to the collection of large and diverse sets of repeated samples; which will allow researchers to do things like plot the level of well-being, observe its direction, monitor the changes in its magnitude, and assess the impact of events like policy changes on happiness with speed and precision not currently possible. This is in sharp contrast to traditional surveys, which take lots of time, are expensive, make cross national comparisons difficult, are limited in their availability, and make answering questions outside of their scope difficult.

As more researchers contribute to this line of inquiry, our understanding of basic human characteristics and their relationship to economic and other outcomes will improve. However, to achieve this goal, a key step is the development of a lexicon that is capable of differentiating texts based on their emotional content, and that can capture their intensity. Its purpose is to act as a data filter. By adjusting its contents, more relevant pages can be picked out of the World Wide Web or other new media sources; these can then be analyzed qualitatively or statistically.

In this sense, it is just one part of a complete analytical framework. After mining the Internet, the idea is to move forward by translating the heap into useable data, analyzing it, and converting it into useful information. This is then used as the basis for supporting or rejecting existing theories, and making policy recommendations. The sentiment extraction process is described in [10], in that context, the lexicon can be seen as an input into the various stages that begin with the crawler, continue with the sentiment processing phases (extraction, classification, and quantification), and end with data analysis.

1.1 Summary of the Paper

The thrust of the paper begins in Section 1.2 where facts about reference lists and classification are pointed out. It highlights methods for reducing ambiguity, and improving the classification accuracy of texts. Based on this it argues that the lexicon needs to be composed of high frequency items that are specific to the research question. This will reduce the errors involved in text selection, and result in better quality data selection.

Since the ultimate aim is to use the collected texts for rigorous scientific work, care needs to be taken to ensure that data quality is maintained. With this in mind, section 2 discusses a few of the issues relating to mining online data and how the difficulties can be assuaged. After all, only by collecting useful and good quality data can meaningful analysis be done.

The third section shows how to build a lexicon that can both separate data based on emotive content and that can capture the intensity of the emotions in the text. This is done by using a metric from information theory (specific mutual information) to determine how useful a word is. It provides a quantitative way of selecting highly informative words; these are kept and used to select appropriate passages. Doing this prevents low polarity terms from populating the emotive-lexicon and hence allows for the execution of more refined searches. In essence, the algorithm starts with a set of application-specific seed words and then uses a combination of existing references and the mined corpus to iteratively expand the lexicon, thus combining the two main strands in the literature. This is done jointly by expert and automated means to

balance speed and cost while ensuring data quality; and consequently, the robustness of the subsequent analyses. One of the advantages of this algorithm is that it captures sentiments no matter how they are expressed, which minimizes information loss.

1.2 Background

An important insight is that not all words are useful for classification. In fact, [2] shows that 21.3% of the terms found in different references can actually diminish the quality of results since classification-accuracy ranges from 52.6 percent when using lists with more ambiguous words, to a high of 87.5 percent when more definite sentiment laden words are used. Moreover, it finds that the boundary between sentiment and neutral words accounts for 93 percent of the misclassifications. This implies that large and random lexicons may not be representative of the population of categories. Thus, the lexicon's elements need to be carefully selected based on their ability to classify the text's emotional content. By doing this we can reduce the misclassifications substantially and improve the relevance of mined pages.

With this in mind, an important issue is the assessment of lexical elements. Ideally the metrics used should be easy to understand, computationally light, and they should select the best items for the word lists. Frequency based metrics have been shown to meet these criteria. Even though simple word counts have some limitations, [6] find that they make the best compromise for creating lexicons for sentiment extraction. This is supported by [3] since true terms tend to have high frequency and because the metric is computationally very light. So it is not surprising that statistical measures fare worse when applied to low frequency terms [5].

Intuitively, one would expect the classification of texts into emotive and non-emotive categories to improve as the precision of the lexicon increases; which is accomplished by populating it with relatively unambiguous elements. Indeed, this is what happens when measuring ambiguity with a Net Overlap Metric (NOM). A particular lexical item does not always correctly classify the sentiment and it does not always recognize that a text has emotional content [2]; these errors are more common for some items than for others. Those that make classification errors more frequently will tend to classify truly positive sentiments as both negative and positive (similarly for negative sentiments), so taking the difference between these frequency counts (i.e., calculating the NOM) will yield values comparatively closer to zero. Those elements that fail to select texts will also have frequency counts close to zero, and taking their difference will again produce neutral NOM values. For these reasons, NOM scores near zero are suggestive of imprecision.

Empirically, less ambiguity relates to higher classification accuracy, which corresponds to the intuition above. A category consisting of items that reflect neutral sentiments, and items that have been classified as both positive and negative an equal number of times, has a NOM value of zero; it gives a classification-accuracy of only 20%. A category that only contains items that have been classified as positive and negative an equal number of times has a NOM score of zero too, but in this case it classifies better, yielding an accuracy of 57%. Overall, the tendency is for accuracy to increase with higher absolute NOM scores (exceeding 90 percent for absolute values above seven) [2]. The pattern is that by eliminating items unclear in meaning, ambiguity decreases and sentiment is classified (as either positive or negative) more precisely.

Based on these observations and results, the approach adopted here is to populate an emotive-intensive lexicon with specialized high frequency words, phrases, and symbols that are relevant to the research question. Since there is a correspondence between counts and accuracy, this method allows the use of frequency based metrics which are computationally light and easy to understand. Naturally, the elements will tend to have higher NOM scores, and as a consequence, there will be fewer misclassifications since the elements will be unambiguous. This will yield higher quality data, which will contribute to more robust statistical or qualitative analyses at later stages; however, these are not addressed in this article.

2 Basic Data Quality Issues

Electronic communication tends to be rather noisy compared to print and other more traditional media. It consists of two broad components: the communication and the content. Poor spelling, bad grammar, and the use of foreign or slang words all add to the confusion and lead to enormous amounts of non-dictionary words in the text. In [9], pre-correction negotiation data contained 11 353 non-dictionary words (due to the noise factors mentioned above) and only 3 255 dictionary words. This suggests that when mining manually typed correspondence, blogs, and the like, it is critical that the lexicon or algorithms be robust to noise. In the procedure outlined below, this is accomplished in stages 2-7 by including a wide spectrum of word forms, symbols, new words, phrases, and possible errors in the lexicon. The consequence is that more pages will be collected for analysis, so less data will be left undiscovered.

Further, social media can often be off topic, ambiguous and composed of messages without any sentiment content (e.g., questions, factual responses, non-emotive or irrelevant opinion). As [1] shows, within eight hours of corporate press releases, on topic posts account for 50 to 75 percent of total postings. Of the on topic posts, between two and nine percent are questions, one to thirteen percent are factual, and between 40 and 65 percent are opinionated. This suggests that, at most, you can expect to find relevant sentiments in 49 percent of the messages, of course, this is an upper bound, and the useful information will be far less. Beyond the way emotions are communicated and the content of the text, the elements of the lexicon play an important role in data quality.

Selection of the lexicon's contents should be based on the domain of study, the requirements for the upcoming analysis, and the method used to collect and analyse the data. For instance, if a search engine is being used to amass the pages, then one needs to bear in mind that they have limitations which can distort the page results. In particular, they do not have functionality for lemmatization, so words and languages with a rich morphology will tend to be underrepresented in the search results [7]. Since we are interested in measuring relative or total values during the econometric phase that follows text mining, the data may lead to skewed results; methods like collecting copious amounts of pages (law of large numbers) will be unable to resolve the statistical issues.

When building a lexicon, these factors should be considered because they will affect which pages are collected during the mining stage. The solution presented here is to include misspellings, inflections, short-forms, emoticons, and relevant foreign and slang words in the word list. This increases the potential information without biasing

the corpus, then, at a later stage the collection can be cleaned. Including foreign and slang words has the advantage that other language groups and subcultures are included. This leads to a more representative sample and is particularly important for studies at the national and international levels. Since it contains a large yet specific collection of items, the lexicon is flexible in that it can be used with a number of different tools and approaches to gathering data.

Using a selective wordlist will to a large extent sieve out irrelevant content because the page hits are based on a match between the lexical items and the contents of the page. Other ways of improving the initial data quality is to avoid strings that have no relevance to the domain, are functional, or polysemous (unless they refer to emotive or intensive semantics) [7]. Other wise unrelated pages or context bereft text (e.g., sale prices) will be uncovered.

This approach is meant for mining a large and changing corpus like the Internet. However, it can be used in fixed or small bodies of writing. In those cases, it is also possible to use other tools like spell checkers [9]. In this application, however, that is impractical. More generally, it is not advised since tampering with the raw data can lead to information loss, e.g., accidentally changing the spelling to a wrong word, which could cause the analyst to discard a useful text.

3 Lexicon Construction

When applying NLP to questions in the social sciences and business, researchers are interested in detecting both the polarity and the intensity of emotion. This creates a slight complication in the selection of texts and in their analysis. Since the evaluation is along 2 dimensions, up to 2 instruments are needed; in this case, items that correlate heavily with emotion (such as depressed) and those that signify intensity (e.g., very). However, peppering a single lexicon with both has problems.

Intensive terminology tends to have an ambiguous polarity, so it should not be used to select emotive passages since pages will be over-collected. Weeding them out not only wastes time and increases the financial burden, it also increases the economic cost by introducing misclassification errors and reducing the quality of the final dataset. Instead, two word-lists are constructed, one for each dimension. Using just the first one allows for more precision when selecting sentiment laden texts, whereas using a combination of the two allows for the measurement of intensity and more advanced analytics at later stages.

Although adjective phrases could be extracted from texts using tools like dependency parsers, they are not included in the lexicon because the application does not require it. More advanced analysis using the phrases is reserved for a later mining stage not in the scope of this paper. Similarly, complex syntactic forms are not used.

The focus here is to use this 10 step procedure to construct a bifurcated lexicon:

1. Select seed emotive or intensive words.
2. Use a thesaurus to add synonyms, antonyms, and slang words.
3. Use a multilingual dictionary to add emotive or intensive foreign words.
4. Add inflected forms, short forms, misspellings.
5. Add emoticons or non-standard text / symbols.

6. Count the number of positive and negative sentiments for each item in the lexicon.
7. Apply Specific Mutual Information (SMI) to separate emotive and intensive items in order to create 2 word-lists.
8. Use the current emotive lexicon to select emotion laden sample texts.
9. Add new item by searching the samples.
10. Repeat steps 2-9 using the new items until the marginal cost = marginal benefit.

The first stage requires an initial set of words that relate to emotion or its intensity. There are a number of ways of doing this such as manually searching through documents, using previously published word lists, or using reference dictionaries. However, randomization is always forbidden since most natural language words are not relevant for the domain and because their noise will be magnified by several orders of magnitude at stages 2-4. Although they are neutral in polarity, their effect is not, and they may lead to misclassification, as pointed out in section 2.

The advantage of this ten step procedure is that even with a comparatively small set of seeds, a large lexicon can be constructed because of the iterative nature used to populate the list. Other studies often rely on experts or existing references such as the General Inquirer or WordNet [8] [2]. This is the approach advised here since a large number of words with positive or negative connotations can be quickly and easily chosen. In total, [8]'s procedure yielded 9677 polarity words. However, while they are positive or negative in sense, they are usually not emotive. So a human expert is required to prune the list further. Although more work is required initially, there is less chance that relevant words will be left out. Further, the lexicon will be more specialized for the domain, and the amount of reiterations will decline.

The literature has approached lexicon construction from primarily 2 avenues: thesaurus and corpus based. This approach combines the two, in that, stages 2-3 use existing references and well known words, while steps 7-8 rely on the contents of a mined corpus. This hybrid approach builds on the strengths of the underlying methods while mediating some of their restrictions. In particular it quickly expands the lexicon through the use of synonyms and antonyms, and it allows the lexicon to hold new words or phrases not found in formal sources.

Stages 3-5 capture sentiments no matter how they are expressed. This is important given the sloppy casualness with which people communicate over the Net, and given the variety of word forms and symbols they use to express themselves. Unlike in traditional media, online-feelings are often communicated more directly using emoticons, whose lists are available and easily appended to the lexicon. Misspellings are found by visually scanning the data collected, and from the expert's judgment since words are often misspelled the same way (e.g., transposition errors). Inflections and short-forms are added in a similar way, although standardized references or grammatical rules can be used to somewhat automate the process and add a degree of independence.

Based on the assumption that emotive words tend to have a strong polarity in only one direction (positive or negative), stages 6 and 7 are used to define which items in the lexicon are emotive and which are intensive. For instance, the word "Happy" by itself is emotive, but its degree changes when used in conjunction with "very". In this case, only "Happy" would be included in the emotive list because it has a positive frequency count in step 6, while "very" tends to be more neutral (i.e., it is used in

non-emotive phrases (e.g. very hot weather), and to express opposite emotions (like "very Sad")). Instead, it would be included in the intensive word lexicon, even though in combination it gives information about the emotional state.

The degree of polarity is assessed statistically with an association measure: Specific Mutual Information (SMI), which is described in information theory [4]. Since its value ranges from 0 to -log(p(x))[1], it can be applied to discrete outcomes[2] in order to see how different they are. Intuitively, SMI reveals how much information y gives about x (e.g. how much one word gives about the other), but formally it is defined as[3]:

$$I(x;y) = \log \frac{f(x,y)}{f(x)f(y)} \ . \tag{3.1}$$

The numerator is the joint distribution, and the denominator consists of the product of marginal distributions for x and y respectively. Since we are using the frequency counts to differentiate emotive and intensive words, it specifically becomes:

$$I(x;y) = \log \frac{f(word, pos.\ sentences)}{f(word)f(pos.\ sentences)} \ . \tag{3.2}$$

In essence, it checks how often the word is positive compared to the random case. Similarly, we can determine which words have negative sentiment:

$$I(x;y) = \log \frac{f(word, neg.\ sentences)}{f(word)f(neg.\ sentences)} \ . \tag{3.3}$$

If an item occurs significantly more often in positive / negative sentences than independence would suggest (i.e., if the ratio of 3.2 or 3.3 is >>1), we conclude it is emotive. However, if both 3.2 and 3.3 are near unity the indication is that the item is not—it occurs in sentiment laden contexts as often as it would by chance. Consider the term "somewhat depressed". It is negative because the word "depressed" cannot be positive or neutral by itself. Whereas "somewhat" also makes sense in the opposite term "somewhat joyful"; consequently, it tends to have strong values in all the equations above. Thus, equal frequency in both negative and positive sentences (i.e., both 3.2 and 3.3 are >>1) also indicates non-emotiveness.

A simple difference (Equation 3.2 – equation 3.3) encapsulates these three possibilities and gives an easy measure of semantic orientation (SO). Those words with a positive or negative SO are put into the emotive lexicon since we are interested only in the emotion and not the direction, while those with a neutral orientation are put into the intensive lexicon ($SO \approx 0$). Both types of information are needed to do quantitative analysis on the type and degree of emotion expressed online, so both instruments need to be kept apart.

---

[1] 0 when two discrete outcomes x and y are independent, and -log(p(x)) when they are perfectly associated.

[2] Not to random variables, that is for Mutual Information.

[3] Since all logarithms are equivalent up to a constant, the base of a log does not affect the analysis, so it is left as base 10. However, it is common to use base 2 or the natural log.

The 8th step uses the current lexicon to discover web pages that are likely to have emotional content. Once the pages are downloaded, a domain expert scans them in the 9th stage to find words not on the list. The process can be aided by automatically deleting words from the web / text files that are already in the lexicon. This part of the algorithm allows the lexicon to keep up with changing language, and to detect special items as in steps 4 and 5.

The last stage reiterates the process using the new additions until the marginal value is sufficiently diminished. Economically, this is the point where the value of an additional item is less than the cost to discover it. Finding that point is beyond the scope of this article, but its essence can be operationalized by using intuitive heuristics like "stop once x-pages are covered without a new item being found" or by setting a cutoff value for the lexicon size, number of new page returns, etc.

The procedures described require a large amount of manpower, but this is by design. The purpose of the lexicon is to generate mined data with which researchers can conduct precise scientific analysis. In order to make a representative corpus and a high quality dataset, the lexicon needs to be domain specific and complete. Humans do this best; especially when the texts are complex, the content is truly new, and when there are several subjects, features, or sentiments. So using experts leads to higher quality data and better results.

Once the lexicon is constructed, there is very little need to change it because large amounts of relevant new symbols or words rarely crop up. This creates economies of scale. So the time invested upfront in creating a thorough and precise lexicon pays for itself through its reusability, and the ability to distribute it freely. This means that no one else needs to repeat the work, which saves resources globally.

4 Conclusion

The requirements for doing socio-emotional research online differ from other text mining applications. Key to the successful acquisition of data is the construction a lexicon that can separate new media texts based on emotive content, and that can contribute to greater analysis at later stages by allowing for the measurement of intensity. By separating the items into emotive and intensive parts, the ten step procedure outlined above allows the construction of a rich lexicon that meets both of these requirements. It also integrates the thesaurus and corpus based approaches typically used in the literature, and allows researchers to use data mining techniques to study emotions and their intensity.

The ability to do both qualitative and quantitative analysis opens the door to work in diverse areas like customer relationship management, human resources, applied statistics, behavioural economics, happiness & well-being, marketing, business information systems, and management. Since online emotions mining has rarely been done in these fields, [10] provides some examples of applications and develops a process for doing this type of research. For instance, mining emotions could be used to study how influence & trust, social contagion, and the interconnection of new media sites (e.g., blogs, web pages, and social networks) affects happiness at the micro level. Work in political science indicates that influence via trust can drastically affect sentiment indexes, suggesting that perhaps a similar effect could be seen with online emotions.

More technical research directions also show promise. Determining exactly which methods improve the accuracy of frequency based lexicons, developing better algorithms (for searching, classification, cleaning, reducing, and analysis), and better quantifying noise are also worth pursuing. Other possibilities include making geographic comparisons and doing descriptive statistics. These areas have a lot of potential; but harnessing it requires constructing a bifurcated lexicon as described above, and using the workflow described in [10]. In effect, the lexicon can be seen as an input into the crawler, which gathers the raw data, and as an input into the later sentiment processing & analysis phases.

The paper argues that there are two noise components inherent in the data: communication and content. As such, a combination of a manual and automated process is advocated since it leads to higher quality datasets. This is in part due to the reduction of errors that are inherent in text mining (the third noise component). Clean and valid data is critical for doing robust scientific analysis later on, so using a human expert is paramount. By including a wide spectrum of relevant content in the lexicon, and avoiding strings that do not provide emotive information (e.g., those that have low domain relevance, are functional, are polysemous, or have low net frequency counts), the algorithm captures sentiment no matter how it is expressed while limiting unrelated hits. This leads to the collection of a more representative sample and an increase in overall data quality.

References

1. Admati, A.R., Pfleiderer, P.: Noisytalk.com: Broadcasting opinions in a noisy environment, WP1970R, Stanford University (2000)
2. Andreevskaia, A., Bergler, S.: Mining Wordnet for Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses. In: Proceedings of 5th International Conference on Language Resources and Evaluation (LREC) (2006)
3. Evert, S., Krenn, B.: Methods for the qualitative evaluation of lexical association measures. In: Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, Toulouse, France (2001)
4. Fano, R.: Transmission of Information: A Statistical Theory of Communications. American Journal of Physics 29(11) (1961)
5. Krenn, B.: Empirical Implications on Lexical Association Measures. In: Proceedings of The Ninth EURALEX International Congress, Stuttgart, Germany (2000)
6. Pazienza, M.T., Pennacchiotti, M., Zanzotto, F.M.: Terminology Extraction: An Analysis of Linguistic & Statistical Approaches. In: Knowledge Mining: Proceedings of the NEMIS 2004 Final Conference (2005)
7. Sharoff, S.: Creating general-purpose corpora using automated search engine queries. In: WaCky! Working papers on the Web as Corpus (2006)
8. Sista, S., Srinivasan, S.: Polarized Lexicon for Review Classification. In: Proceedings of the International Conference on Machine Learning, Models, Technologies & Applications (2004)
9. Sokolova, M., Szpakowicz, S., Nastase, V.: Automatically Building a Lexicon from Raw Noisy Data in a Closed Domain. INTERNEG working papers, INR 01/04 (2004)
10. Verma, R.: Extraction and Classification of Emotions for Business Research. In: Communications in Computer and Information Science (CCIS), vol. 31, pp. 46–53. Springer, Heidelberg (2009)

Enterprise X.0 and ECONOM Workshops Chairs' Message

Malgorzata Mochol[1], Tobias Bürger[2], Markus Luczak-Rösch[1], Elena Simperl[2], Lyndon JB Nixon[3], Agata Filipowska[4], and Christoph Tempich[5]

[1] Freie Universität Berlin, Netzbasierte Informationssysteme (NBI), Berlin, Germany
{mochol,luczak}@inf.fu-berlin.de
[2] Semantic Technology Institute (STI) Innsbruck, University of Innsbruck, Innsbruck, Austria
{elena.simperl,tobias.buerger}@sti2.at
[3] Semantic Technology Institute (STI) International, Vienna, Austria
lyndon.nixon@sti2.org
[4] Poznan University of Economics, Poznan, Poland
A.Filipowska@kie.ae.poznan.pl
[5] Detecon International GmbH, Germany
Christoph.Tempich@detecon.com

After the first generation Web which started with manually created HTML pages, the second generation made the step to machine generated and often active HTML pages. Since these first two generations were meant for direct human processing, the third generation Web, the Semantic Web and Web 2.0 provide machine processable information and social collaboration, respectively. Over the last decades, the WWW has rapidly evolved into a vast repository containing huge amounts of decentralized information on all matters of interest. It is now evolving from the medium intended for human utilization into *a medium for collaborative knowledge generation and intelligent knowledge exchange achieving the time-to-market demand in a competitive environment.* This is why CIOs are starting to acknowledge the technical value of knowledge-based technologies for enterprizes: In the last years, early adopters have been increasingly using the technologies in various application settings ranging from content management to enterprize integration platforms. Core technological building blocks and development platforms are meanwhile available from established vendors. Despite this promising position, it is still difficult to argue in favor of knowledge-based technologies in front of the CFOs because of the lack of convincing measurable benefits or proven-and-tested methods to determine and quantify these.

The ***Enterprise X.0****: From Web 2.0 in Enterprises towards a Corporate Web X.0* and the *Economics of Knowledge-based Technologies (**ECONOM**)* workshops concerned the economic aspects and the future development of knowledge-based systems in the corporate context. The workshops targeted visionaries (researchers and practitioners) who are not only working on Web-based information systems using Web 2.0 and Semantic Web in the business context, but are looking forward to exploiting the next wave of Web developments: Corporate Web X.0. In this context we believed that the time was ripe for the next visionary view and question: What is the next, logical step after the use of the

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 139–140, 2009.

Web 1.0, Web 2.0, and semantic technologies in business settings? Furthermore, considering such a development, it is important to assess the potential business value, more precisely the costs and benefits of the modern applications, and qualitative and quantitative methods therefor. The workshops addressed research looking into the aforementioned aspects for both knowledge-based systems and knowledge structures (i.e., ontologies, taxonomies, folksonomies) supporting these systems. The main goal of the workshops was to provide *a communication platform for researchers & practitioners to discuss ideas and results and to identify new challenges in the areas of economics and knowledge-based technologies in business context*. The audience got an overview *how new trends in and after the Web 2.0 and Semantic Web era can influence corporate processes and where benefits for the business world can be found*. The submitted contributions for both workshops published in these proceedings reflected current research in the aforementioned areas: the topics ranged from value assessment of knowledge-based technologies and information systems, business aspects of ontology engineering, and data integration supported by knowledge-based technologies. The contributions were from both academia and industry. Especially contributions from the latter gave interesting insights into real-world industrial problems. All talks discussed aspects of how knowledge-based technologies can provide added value or how their value can be assessed: Irene Celino in her invited talk on *"Business opportunities of Linked Data"* showed and explained the added value of data integration and Linked Data. The paper *"Framework for Value Prediction of Knowledge-Based Applications"* by Imtiaz et al. presented a generic framework to assess the value of applications in general and exemplified how to apply it to knowledge-based applications. Strasunskas and Tomasgard's paper *"In Quest of ICT Value through Integrated Operations: Assessment of Organisational Technological Capabilities"*[1] introduced an approach for ICT valuation and its application to valuate integrated operations. van Teeseling and Heller showed how to derive, build, and use business patterns in ontology design by analyzing characteristics problems and translating them into patterns to design ontologies. Their paper *"Business Patterns in Ontology Design"* further described how to use the patterns to quickly develop knowledge-based applications and to gain value by that. The paper *"E-Business in the construction sector: a service oriented approach"* by Sanchez et al. delivered insights into the ENVISION platform which provides added value for SME's, materialized by services to facilitate e-tendering or e-procurement. Finally, Nekvasil and Svatek, in the paper *"Towards Models for Judging the Maturity of Enterprises for Semantics"*, presented a framework and critical success factors for gaining value out of knowledge-based applications.

The workshop chairs would like to thank the PC members for their support in the reviewing process, the organizers of the BIS 2009 conference and especially Dominik Flejter for a kind assistance throughout the organizational process, as well as our speakers and participants for the interesting and stimulating talk during our workshops.

[1] This paper won the best paper award of the 12th International Conference on Business Information Systems.

From Research to Business: The Web of Linked Data

Irene Celino[1], Emanuele Della Valle[1,2], and Dario Cerizza[1]

[1] CEFRIEL – ICT Institute, Politecnico of Milano,
Via Fucini 2, 20133 Milano, Italy
name.surname@cefriel.it
[2] Dipartimento di Elettronica e Informazione, Politecnico di Milano,
Piazza Leonardo da Vinci 32, 20133 Milano, Italy
emanuele.dellavalle@polimi.it

Abstract. The last decade of research in the Web field gave a great importance to the studies about the Semantic Web. The idea of a Web of Data is now becoming more and more popular also outside of the pure scientific community. The idea of *linked data* is thus gaining ground and demonstrating its advantages and its opportunities in the business world. Still a lot of research is there to come.

In this paper, we discuss the need for linked data technologies, we illustrate two case studies from European research projects and we examine the opportunities from the business and the technological point of view.

Keywords: linked data, Semantic Web, Web of Data, Service Web, Reasoning Web, Urban Computing.

1 Introduction

In the area of information management, the market is constantly asked for more and better solutions to solve the *problem of integration*. Why is industry so keen in finding new answers to information integration? Today organizations and enterprises have to face at least three different challenges:

- they have a problem of *scale*: they must manage very large amounts of data, which grow and evolve continuously;
- they have a problem of *data heterogeneity*: the data they produce and consume every day belong to numerous and different typologies (documents, media, email, Web results, contacts, etc.);
- they have a problem of *system heterogeneity*: those data are managed by numerous and different information systems (DB, legacy systems, ERP, etc.).

But why is integration so significant? Because integration always gives an added value: in getting a global high-level view over different and independent systems; in sharing knowledge between groups and partners; in unleashing business opportunities which are enabled only by unifying and combining efforts; in answering the questions of decisions makers, as it happens in the Business Intelligence area. Integration – using an effective maths metaphor – is when $1 + 1 > 2$.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 141–152, 2009.

Therefore, solving the integration problem seems to be the first point in everybody innovation agenda. In order to understand the possible ways out, it is interesting to have a look at those who were able to ride and exploit the integration challenge to their own advantage. Some lessons learned come from the so called Web 2.0.

As James Surowiecki effectively explained in his famous bestseller "Wisdom of Crowds"[1], a strong driver of the integration in the Web 2.0 is constituted by the *participation politics*: the collaboration of people makes a large task smaller, turns a big problem into a manageable one.

Moving from the attitudinal and social aspects to the technological point of view, the Web 2.0 revealed the great success of *mash-ups*. A mash-up is an application made up of the light integration of artifacts provided by third parties (like API or REST services). Beside being a way to implement Web applications, mash-ups are also a new *integration paradigm* to software development, in which loose agreements – instead of tight and complex interface specifications – are needed to get to a useful result. Last but not least, the Web proved to be indispensable for data and information *publication and access*: we are more and more accustomed to store our information on the Web and we also access it more and more frequently on the Web (e.g. by retrieving it via search engines).

From all those considerations, we can easily come to a question: is it possible to reach *integration on the Web*? Can we use the Web as a *platform* for integration? How can we leverage the Web *prosumers* (producers and consumers) to get to data integration? The scientific community shows us the road to the Web integration, i.e. moving from the current Web of Documents – made by and for people – to the Web of Data – where machines can play a crucial role in knowledge management, e.g. by advancing from a pure information retrieval (from a request to some documents that could contain useful information) to a smarter *data retrieval* (from a question to its answer).

In order to reach this objective of the Web of Data, it is clear that we are all invited to take part in this Web evolution, by exposing our data on the Web. This appears to be quite easy and natural for individuals, since it already happens with the so called "user-generated contents", which are more and more frequently annotated with metadata like the tags that can be used by machines for their processing. On the other hand, enterprises and organizations must find a straightforward way to expose data on the Web. This operation can be realized with two different approaches:

- by the *conversion* of the data source, which is translated into a suitable format for its publication on the Web; however, this solution is not always feasible, e.g. when the data source is frequently updated or its scale is so large that the translation process takes too much (in terms of time or costs) with respect to the use of the data;
- by the *wrapping* of the data source, i.e. by inserting a virtualization layer on top of the source which translates the queries and their responses from and to a Web-compliant format; several tools exist now, coming from both the scientific community and the industry [2,3,4,5,6,7,8,9].

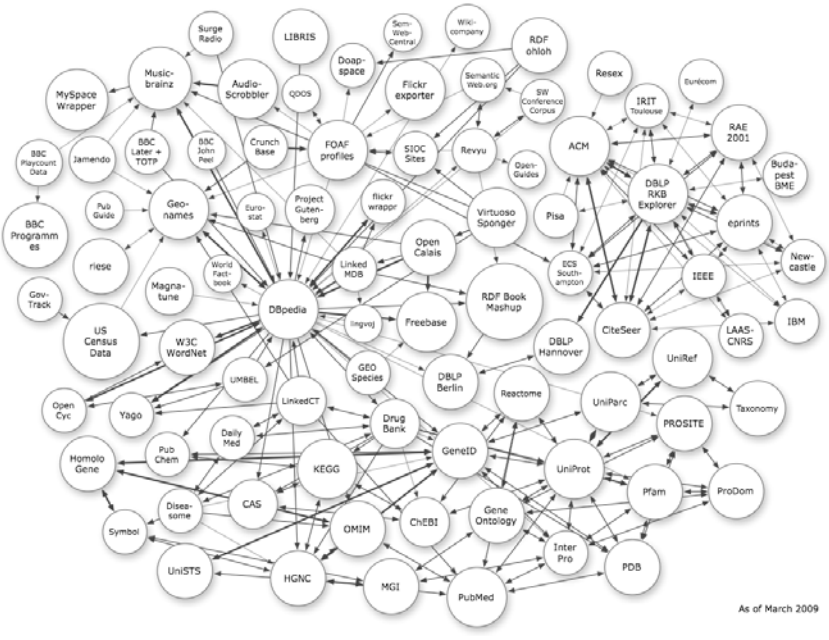


Fig. 1. The LOD cloud, as of March 2009 (source: [11])

In both cases, by publishing the data on the Web or by providing an access point to them, we are concretizing the idea of the Linked Data, as described by Tim Berners-Lee [10]. The community already recognized the value of this idea of data linked, connected to other pieces of data, aimed at forming what is sometimes indicated as Giant Global Graph (GGG); a specific initiative [11] was started to collect available data sources and to link them together as much as possible, so that the contained information can be seamlessly navigated regardless to the sources' boundaries. The so called Linked Open Dataset (LOD), depicted in Figure 1, now encompasses numerous sources of different kind and topic and reached the size of more than 4.5 billion triples.

The LOD "cloud" and its size makes immediately think about the problem of managing that scale of data: is current cloud computing technology up to the task of processing and handling the Web of Data? The scientific and industrial community must find a solution to this problem, by joining the efforts about scalable systems and Web technologies, since it is less and less a computational or storage issue but the challenge lays in data and knowledge management.

Finally, the trend behind the current popularity of linked data and the increasing availability of tools and techniques to deal with them do not mean that the research agenda of Semantic Web technology is over. Instead, several interesting challenges are still to come; among them we list the following ones:

– *Automatic linked data creation and linkage*: the automatic generation of linked data and smart mechanisms to identify "contact points" between different data sources and to seamlessly link them;

- *Distributed querying*: querying distributed data over different Web sources regardless the "physical position" of data and getting aggregated results;
- *Distributed reasoning*: applying inference techniques to distributed data, preserving consistency and correctness of the reasoning.

In the following we present two running research projects which try to address the aforementioned challenges and we foresee the future of the linked data, both from a business and a technological point of view.

2 Production of Linked Data: Service-Finder

The first question to answer when talking about the linked data is: how can we produce them? The easy generation and maintenance of machine-readable data published and accessible on the Web is the first step to take. In this chapter, we provide an example of how to derive linked data from the current Web.

2.1 Concept and Architecture of Service-Finder

The Service-Finder project [12] is addressing the problem of utilizing the Web Service technology for a wider audience by realizing a portal[1] for Web Service discovery by making Web Services available to potential consumers similarly to how current search engines do for content pages.

An essential, but mostly unaddressed problem in the area of Service Oriented Architectures (SOA) is the creation of such semantic descriptions of Web Services. Service-Finder aims to offer automatic creation of service descriptions for a different range of Services (all the publicly available services) and to enable service consumers, not just service providers, to enrich the semantic service descriptions, following a typical contribution-based approach in a Web 2.0 fashion.

The Service-Finder project delivered a Search Engine that enables users to find up-to-date information on available Web Services. By employing automated crawling and analysis techniques, the Service-Finder approach is able to scale with the increasing number of services and does not rely on a central editorial team. Consequently, Service-Finder can adapt quickly to changes in terms of the available services. The Search Engine leverages available information exposed by current technologies and extends this information with semantic annotations to allow for a more accurate retrieval. Thus, Service-Finder approaches the discovery problem by developing novel means of obtaining the underlying semantic models for discovery, by analyzing available Web content and leveraging direct and indirect user feedback on the extracted data.

The Service-Finder Portal is built on top of some core technologies and expertises brought by the project partners that act as the strategic *ingredients* for the project. Figure 2 shows the main six ingredients:

- *Web Services* are the basic paradigm and technology to implement a Service Oriented Architecture;

[1] The Service-Finder Portal is available at `http://demo.service-finder.eu/`

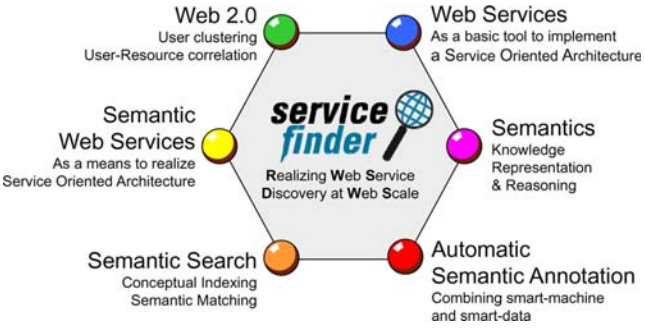


Fig. 2. Service-Finder Core Technologies

- *Semantics* provides the methodologies and tools to represent knowledge and to reason over it;
- *Automatic Semantic Annotation* is a way to enrich the gathered Web Service descriptions with semantic annotations;
- *Semantic Search* improves recall and precision by indexing Web Services related data at the conceptual level and by enabling semantic matching between Web Service descriptions;
- *Semantic Web Services* extend Web Services descriptions to easily realize Service Oriented Architecture;
- *Web 2.0* as a paradigm to involve users in the process of improving their experience with the portal.

Service-Finder combines those ingredients into one coherent architecture, depicted in Figure 3 and based on five internal components:

- The *Service Crawler* (SC) obtains the available services and related information by crawling HTML and PDF documents from the Web.
- The *Automatic Annotator* (AA) receives the crawled data and enriches it with annotations according to the Service-Finder ontology and the Service Category ontology.
- The *Conceptual Indexer and Matcher* (CIM) receives and integrates all the information into a coherent semantic model based on the ontologies and provides reasoning and querying capabilities.
- The *Service-Finder Portal Interface* (SFP) provides the user interface for searching and browsing the data managed by the CIM. It also enables users to contribute information in a Web 2.0 fashion by providing tags, categorizations, ratings, comments and wiki contributions.
- The *Cluster Engine* (CE) analyzes the users' behavior in interacting with the SFP in order to provide them with recommendations and statistics.

2.2 From Service-Finder to the Web of Data

Even if the intended result of the Service-Finder project is the realization of a Web portal for searching for Web Services, a sort of "collateral effect" directly

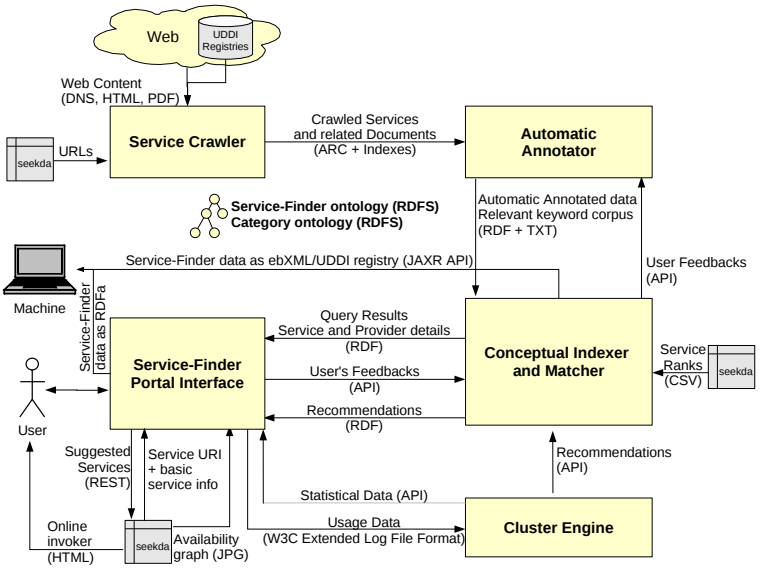


Fig. 3. The logical architecture of the Service-Finder Portal

contributes to the Web of Data. In fact, Service-Finder takes information about services from the Web, translates it into structured information describing services with regards to two domain-specific ontologies, and finally gives this information back to the community that can further enrich it.

Is this linked data? Not completely, since it is not yet in a "linked" and "Webby" format. But also this objective is close, since the Service-Finder project will provide linked data for machines in the following ways (cf. also Figure 3):

- the Service-Finder Portal will soon include *RDFa annotations* [13] in all service pages, so that GRDDL processors [14] will be able to extract service descriptions from the portal pages;
- Service-Finder will also provide *free services* (under the form of API or REST services) to query the knowledge base and get the service information directly via code; in this way, new applications will be allowed to access and exploit the "content" of the Service-Finder system;
- finally, it is possible to envisage a publication or release of a complete "*dump*" of the *Service-Finder knowledge base*; in this way, the content of the Service-Finder system could be easily published on the Web as linked data.

More information and news about Service-Finder are available at the project Web site: `http://www.service-finder.eu`.

3 Management of Linked Data: LarKC

Once the Web of Data grows and more and more data sources are turned into linked data published on the Web, the problem arises of how to manage those

data and how to exploit, in an intelligent and scalable way, the knowledge available on the Web. LarKC[2] is a running research project aimed at realizing a platform for reasoning at Web scale. In this chapter, we present a possible use of linked data in one of the project use case scenarios.

3.1 Urban Computing in the LarKC Project

Urban settings range from our own cars, while we drive them in town, to public spaces such as streets and squares including semipublic ones like cafés and tourist attractions. Urban lifestyles are even broader and include people living, working, visiting and having fun in those settings. Not surprisingly, people constantly enter and leave urban spaces, occupying them with highly variable densities and even changing their usage patterns between day and night [15].

Some years ago, due to the lack of data, solving Urban Computing problems with ICT looked like a Sci-Fi idea. Nowadays, a large amount of the required information can be made available on the Internet at almost no cost: computerized systems contain maps with the commercial activities and meeting places (e.g., Google Earth), events scheduled in the city and their locations, positions and speed information of public transportation vehicles and of mobile phone users, parking availabilities in specific parking areas, and so on.

However, current ICT technologies are not up to the challenge of solving Urban Computing problems: this requires the combination of a huge amount of static knowledge about the city (i.e., urbanistic, social and cultural knowledge) with an even larger set of dynamic data (originating in real time from heterogeneous and noisy data sources) and reasoning above the resulting time-varying knowledge. A new generation of reasoners is clearly needed. This is the purpose of the Urban Computing use case in the LarKC project.

Taking into consideration the peculiarities of urban environments, the LarKC project derived requirements [16] to be addressed by the reasoning community:

- *Coping with Heterogeneity*: data heterogeneity is a common problem for semantic technologies; we distinguish between the following heterogeneity cases:
  - *Representational Heterogeneity*, which means that semantic data are represented by using different specification languages.
  - *Reasoning Heterogeneity*, which means that the systems allow for multiple paradigms of reasoners, like temporal, spatial or causal reasoning; moreover, sometimes precise and consistent inference is needed, but in other cases approximate reasoning or imperfect estimations can be better.
  - *Default Heterogeneity*, which means that systems support for various specification defaults of semantic data; for example, closed world assumption vs. open world assumption, or unique name assumption vs. non-unique name assumption.

[2] http://www.larkc.eu

– *Coping with Scale*: although we encounter large scale data which are not manageable, this does not necessarily mean that all of the data must be dealt with simultaneously.
– *Coping with time-dependency*: knowledge and data can change over the time; for instance, in Urban Computing names of streets, landmarks, kind of events, etc. change very slowly, whereas the number of cars that go through a traffic detector in five minutes changes very quickly. This means that the system must have the notion of "observation period", defined as the period when the system is subject to querying.
– *Coping with Noisy, Uncertain and Inconsistent Data*: data about a urban environment can often be noisy (when a part of data is useless or semantically meaningless), inconsistent (when parts of data are in logical contradiction with each another, or are semantically impossible) or uncertain (when the data semantics is partial or incomplete).

This set of requirements clearly shows that linking data and publishing them on the Web is just the first step: a smarter and scalable processing solution is needed.

3.2 A Urban Linked Data Mash-Up in LarKC

In Figure 4, the first Urban Computing application developed on top of the LarKC platform is represented. This is not a fully-fledged Urban Computing system, but it is the first prototype running over the LarKC platform.

The scenario describes a user which is in a city (e.g. Milano) and wants to organize his Saturday night; for this reason, he wants to know what interesting places he can visit (e.g., if he is a tourist, he would like to know what monuments are open at night and are easy reachable from his place), or he would like to

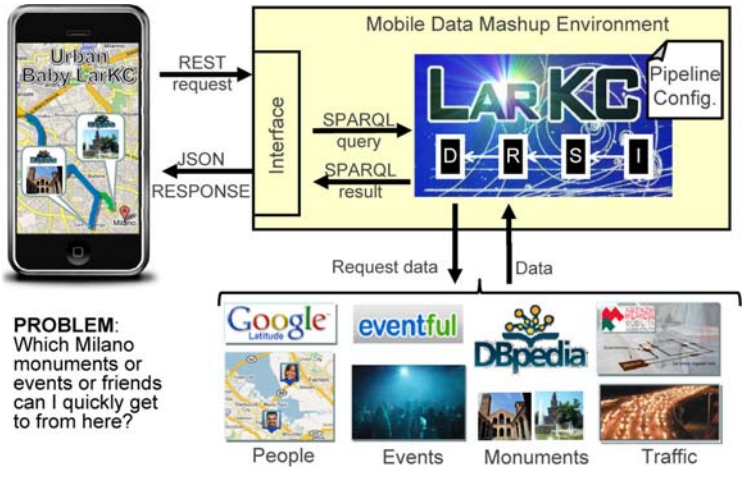


Fig. 4. A graphical representation of the scenario

attend some cultural or music event, or he wishes to meet some friends that happen to be in the same city. Moreover, since traffic in large urban environments is often a mess, he would like to know the most desirable path to his destination and maybe also where to park his car. And, last but not least, he would like to use a single application that fulfill all his needs, without having to manually make sense of the results of different services and applications.

The hypothetic user can turn to a demonstrative application – named Urban Baby LarKC to stress its experimental state – which, by using the LarKC platform, is able to access multiple data sources on the Web and to retrieve the most suitable information to give an answer to the user request. LarKC, in fact, allows for a smarter way to identify relevant data sources, to select a meaningful and useful subset of data and to apply the proper reasoning strategy.

In the scenario case, LarKC identifies the interesting data sources (e.g. an encyclopedia to get information about relevant places, a social networking site to find interesting events or the position of friends, municipality sources to get the real-time situation of traffic, and so on); then LarKC selects the relevant subset of data (e.g., it filters the encyclopedia to get information only about Milano, or events happening on a specific date, or streets only in the area surrounding the current position of the user); finally, LarKC reasons and processes the data to find the most suitable path(s) to suggest to the user.

It is clear that, by increasing the number and typology of data sources, the design of a single application that solves the complete problem becomes more difficult, and scalability, heterogeneity, real-time and noise-tolerance problems become harder to address. LarKC is designed as an extensible platform; it is designed to offer support for building "pipelines" (composite work-flows to realize applications) which invoke several "plug-ins" (reusable pieces of software dedicated to implement a specific function); it allows for data distribution, computation parallelization and so on. For those reasons, LarKC is a suitable architecture to be employed to develop Urban Computing applications that leverage the large amount of linked data that describe the various facets of a urban environment.

4 Conclusions and Outlook

The advent of the Web of Data is capable of radically changing the way we look at our data and their management. The possibilities and opportunities enabled by the linked data approach are emerging and various examples can be found not only in the research field, but also in the industry.

A notable case is *Freebase*[3]. Freebase is "an open, shared database of the world's information"; it is a collaborative effort to build an on-line collection of structured data harvested from many sources, including individual contributions. Freebase releases its content under a Creative Commons "attribution" license, and also offers to programmers an API, an RDF endpoint and a database dump. In this way, Freebase allows people (and machines) to access common information more effectively.

[3] http://www.freebase.com

While at the beginning it was an initiative completely disconnected from the Semantic Web community, the recent release of its RDF version[4] made Freebase a relevant part of the Web of Data.

Another interesting example is constituted by *OpenCalais*[5]. The renowned company Thomsom-Reuters released a free service called Calais, which can be used to derive linked data from documents. Unstructured text is processed by Calais, which extracts entities, facts and events; those "metadata" are returned enriched with "keys" that provide access to the Calais Linked Data cloud; in turn, this cloud provides information and other Linked Data pointers to a range of open and partner Linked Data assets, among which Wikipedia, Metaweb and Cnet, and possibly in the future also the data of Thomson-Reuter itself.

The fact that the service is free of charge and that it is in line with the linked data principles makes OpenCalais a good example of how enterprises can leverage Semantic Web technologies and both profit from them and contributing to the community vantage.

The aforementioned examples demonstrate that the linked data vision is far from being a pure academic question. On the contrary, it can represent a valid means to get to a new generation of improved solutions for the information management and for the problem of integration. Our opinion is that the Web of Data should be better explored both from the business and the technological point of view.

From a business perspective, it is incontrovertible that, being storage a commodity, organizations today are used to produce lots of data; this implies that they more and more frequently experience the problem of managing and making sense of all their data. As a consequence, they often ask for Business Intelligence solutions or turn to similar or related technologies to "understand" their data wealth and take informed decisions. However, it also happens that, when strategic decisions are needed, the data within the organization are not enough and they should be integrated or enhanced with external knowledge.

Clearly, this is a case where linked data technologies can play an important role. One obstacle that can hinder the application of linked data approaches is the claim for privacy and security of data; even if there is undoubtedly the need for better and robust solutions to preserve data confidentiality, ownership and protection, enterprises often use the security "excuse" in order to prevent others to access their data. In this way, however, they block innovation and new business opportunities. As Tim Berners-Lee effectively stated in his speech at TED 2009 [17], organizations should "stop hugging their data" and unleash the full power of linked data.

It is worth noting that, also from a pure scientific and technological point of view, new challenges wait for linked data to find a solution. For example, taking into consideration the case cited above, linked data approaches can mine Business Intelligence techniques' basic assumptions: data and data sources can dynamically change, not only because of data streaming, but also because of the

[4] http://rdf.freebase.com/
[5] http://opencalais.com

unreliability and uncertainty of archives available remotely on the Web (the so called 404 problem); moreover, the Web is intrinsically inconsistent, since you can find everything and the opposite or the negation of everything; the information on the Web can be partial, because of implicit or common knowledge or because of the relations to unavailable sources; finally, on the Web more information than expected or than needed can be found, thus the need for scalable systems or sampling/filtering approaches.

How long is the way to a large scale adoption of linked data in business environments? It is difficult to give a precise answer, but our opinion is that the goal is not far from being accomplished. Best practices and success cases are now more helpful and required than fully-fledged solutions to convince business decision makers to invest in linked data technology; accurate business plans and concrete cost/risk assessments are also needed. But the experience of the Web itself, born in small research labs and gradually spread all over the world, is an outstanding sign of the success ready and waiting for linked data.

Acknowledgments

This research has been partially supported by the *Service-Finder* (FP7-IST-215876) and the LarKC (FP7-IST-215535) EU co-funded projects. The speech entitled "From research to business: the Web of linked data" was given in Poznan on April 29th, 2009 at the joint Enterprise X.0 and Econom Workshops, co-located with the 12th Business Information Systems conference (BIS 2009). The slides are available on the Web at `http://www.slideshare.net/iricelino`.

References

1. Surowiecki, J.: The Wisdom of Crowds. Anchor Books, New York (2005)
2. Bizer, C., Cyganiak, R.: D2RQ: Lessons Learned. In: W3C Workshop on RDF Access to Relational Databases (2007)
3. OpenLink Software: Exposing SQL Data as RDF (2007)
4. Barrasa, J., Corcho, O., Gómez-Pérez, A.: R_2O, an Extensible and Semantically Based Database-to-ontology Mapping Language. In: Second International Workshop on Semantic Web and Databases (2004)
5. Cullot, N., Ghawi, R., Yétongnon, K.: DB2OWL: A Tool for Automatic Database-to-Ontology Mapping. Université de Bourgogne (2007)
6. Seleng, M., Laclavik, M., Balogh, Z., Hluchý, L.: RDB2Onto: Approach for creating semantic metadata from relational database data. In: Informatics 2007, the ninth international conference on informatics (2007)
7. de Laborda, C.P., Conrad, S.: Relational.OWL - A Data and Schema Representation Format Based on OWL. In: Second Asia-Pacific Conference on Conceptual Modelling (APCCM 2005) (2005)
8. SquirrelRDF (2007), `http://jena.sourceforge.net/SquirrelRDF/`
9. Corno, W., Corcoglioniti, F., Celino, I., Della Valle, E.: Exposing heterogeneous data sources as SPARQL endpoints through an object-oriented abstraction. In: Domingue, J., Anutariya, C. (eds.) ASWC 2008. LNCS, vol. 5367, pp. 434–448. Springer, Heidelberg (2008)

10. Berners Lee, T.: Linked Data (2006), http://www.w3.org/DesignIssues/LinkedData.html
11. Linking Open Data Initiative (2008), http://www.linkeddata.org
12. Della Valle, E., Cerizza, D., Celino, I., Turati, A., Lausen, H., Steinmetz, N., Erdmann, M., Schoch, W., Funk, A.: Realizing Service-Finder – Web Service Discovery at Web Scale. In: Proceedings of the 2nd European Semantic Technology Conference (ESTC 2008) (2008)
13. Adida, B., Birbeck, M., McCarron, S., Pemberton, S.: RDFa in XHTML: Syntax and Processing – A collection of attributes and processing rules for extending XHTML to support RDF, W3C Recommendation (2008), http://www.w3.org/TR/rdfa-syntax/
14. Connolly, D.: Gleaning Resource Descriptions from Dialects of Languages (GRDDL) W3C Recommendation (2007), http://www.w3.org/TR/grddl/
15. Reades, J., Calabrese, F., Sevtsuk, A., Ratti, C.: Cellular census: Explorations in urban data collection. IEEE Pervasive Computing 6(3), 30–38 (2007)
16. Della Valle, E., Celino, I., Dell'Aglio, D., Kim, K., Huang, Z., Tresp, V., Hauptmann, W., Huang, Y., Grothmann, R.: Urban computing: a challenging problem for semantic technologies. In: 2nd International Workshop on New Forms of Reasoning for the Semantic Web (NEFORS 2008) co-located with the 3rd Asian Semantic Web Conference (ASWC 2008) (2008)
17. Berners Lee, T.: The next Web of open, linked data, Speech at TED 2009 (2009), http://www.ted.com/index.php/talks/tim_berners_lee_on_the_next_web.html

Framework for Value Prediction of Knowledge-Based Applications

Ali Imtiaz[1], Tobias Bürger[2], Igor O. Popov[2], and Elena Simperl[2]

[1] Research Institute for Operations Management (FIR) at Aachen University of Technology, Germany
[2] Semantic Technology Institute (STI) Innsbruck, University of Innsbruck, 6020 Innsbruck, Austria

Abstract. Knowledge-based applications are characterized by their use of machine-understandable formalizations of expert knowledge. Complex knowledge structures, and the features which exploit them, can have a significant effect on the effort needed to develop such applications. Means to estimate this effort are, however, lacking. Furthermore, precise benefits of such applications, which are directly attributed to specific functionalities, remain unknown.

In this paper we propose a preliminary *Framework for Value Prediction* whose intention is to study and to effectively predict the development effort as well as benefits of knowledge-based applications. The framework consists of five pillars which act as a road map to propose well-defined models. We furthermore discuss our initial experiences with using the framework to adapt existing software cost and benefit estimation models.

Keywords: *Framework for Value Prediction*, cost estimation, benefit estimation, knowledge-based technologies.

1 Introduction

Knowledge-based applications are maturing and a considerable number of systems and core technological components has left the research labs towards the industry in the last years. These applications differ from classical software applications in that they use some form of formal descriptions of the data on which they operate. This allows various tasks (e.g. data integration, reasoning, or search) to be performed depending on the type, domain or particular use of the application. The added dimension of knowledge representation and associated functionalities clearly sets knowledge-based applications as a distinct class of software, required to be studied in their own right. In order to encourage their wide industrial uptake, methods to assess their potential economic benefit and to predict the total costs of their development and deployment are a must. To tackle this emerging challenge, our work focuses on devising a preliminary predictive *Framework for Value Prediction* customized for knowledge-based applications.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 153–158, 2009.

2 The *Framework for Value Prediction*

The general expectation for a framework for value prediction is to identify relevant value drivers (both for cost and benefit) that are measurable in financial terms. Additionally, the predictive framework is expected to define a modular process in which models and methods are applied to best predict the value of the effort to develop, implement, integrate and eventually evaluate benefit factors. The framework is generic and thus customizable to each working domain. A high-level view of the levels of our preliminary *Framework for Value Prediction* is shown in Figure 1. The envisioned framework constitutes of two levels: The first one is the *individual component level*. At this level, the parameters relevant to a domain are assessed and their individual financial values are predicted. The second one is the *organizational level*. On this level an overall integration strategy is devised.

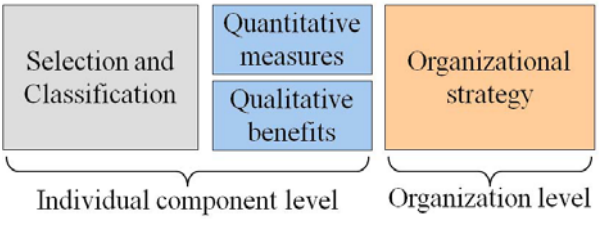


Fig. 1. Levels of the Predictive Framework

As illustrated in Figure 2, the *Framework for Value Prediction* is structured in three functional parts at both the *individual component level* and the *overall organizational level*. The *individual component level* consists of two parts: the

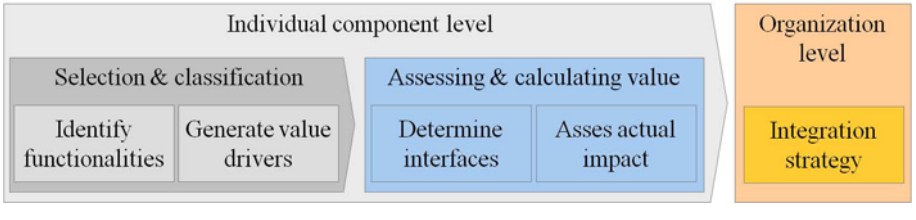


Fig. 2. Parts of the Predictive Framework

analysis part covering the classification of the application functionalities with domain specific value drivers and the *calculation part*, specifying the qualitative and quantitative values of the functionalities. The values determined for the individual components are included in the overall organization level to determine their organizational value.

Each part of the framework is represented as a *pillar* which defines methods or models and corresponding outcomes. In defining the framework, clear segregation

is made between a *model* and a *method*. Within the framework a *method* is described as a systematic procedure of accomplishing something (e.g. catalogue development as a method to capture requirements). A *model* is described as a set of rules used to generate a specific outcome. Unlike a method, a model has clearly structured inputs and refined values as output. For example the requirements could be generated through a catalogue, which is a method. Qualitative values could be assigned to each variable in the requirements using a model. A graphical overview of each pillar of the framework is presented in Figure 3.

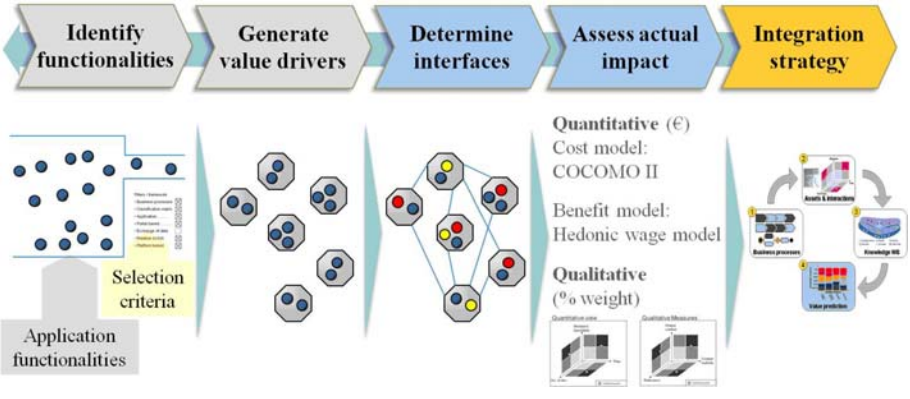


Fig. 3. Five Pillar Predictive Framework

Pillar-1: Identify Functionalities (Methods Implied). The aim of the first pillar is to drive the domain specific selection criteria based on the functionalities of the application or one module of the application. The most suitable methods to perform the selection are expert interviews, application classification catalogue and requirement engineering.

Pillar-2: Generate Domain Specific Cost and Benefit Drivers (Identification and Relevance). The second pillar of the framework analyzes the domain and generates the relevant value drivers. The number of the drivers is directly related to the accuracy of the predicted value of the costs and benefits. Therefore, depending on the complexity of the domain and the resources for evaluation or validation of the drivers, their number should be between 5 and 15. The expected outcome is the dictionary of all relevant drivers including their detailed description, relevance for the domain and organizational placement. The methods considered best for identification of the drivers are: bottom-up analysis, top-down analysis, analogy estimation and expert judgment. The validation of the drivers can be based on domain-expert interviews or ethnographic studies.

Pillar-3: Determine Benefit Interfaces (Based on Pillar 1 and 2 Develop Interrelations). The outcome of this pillar are matrix interfaces between value drivers, task level processes and higher level processes (e.g. business

processes). The generation of the interfaces can be achieved by employing analogy estimations and expert judgment estimations to identify and then validate possible relations between the outcomes of the first and the second pillar.

Pillar-4: Impact of Benefit Drivers (Identify Models both for Quantitative and Qualitative Measures). In this pillar the value drivers that can be quantified and measured are taken up from the preceding pillar. Each value driver is then run through the relevant cost and benefit models to assign their costs, savings or weighted benefits respectively. Before the drivers can be selected and their cost or benefits are estimated, the selection of the right models is critical.

Pillar-5: Driving Integration Strategy. The previous pillars provide the value of individual parameters. In this pillar an overall evaluation at the organizational level is performed, taking into account the results from the preceding pillars. Therefore, the fifth pillar of the framework provides a set of methods to measure organizational profitability and helps the company to analyze the proper balance of possible attraction and retention. The analysis is performed for organizational needs of the overall integration through quantifiable measures. The goal is to define the scope of the application area in relation to the existing organizational structure and then to calculate the overall expected monitory impact [1].

3 Using the Framework for Cost and Benefit Estimation of Knowledge-Based Applications

Our initial observations of the process of creating knowledge-based applications pointed out, that their development is divided into two major subtasks which are handled seperately: one for creating the underlying knowledge structures and the second one for developing the application itself.

The *Framework for Value Prediction* is used to identify and adapt models to be used to estimate the value of a particular type of application. For cost estimation of knowledge-based applications, we adapted existing parametric cost estimation models from the area of software engineering, such as COCOMO II [2], to reflect aspects that are unique to this specific type of application. The preliminary predictive framework is used to support the selection of a set of cost as well as benefit factors that associate the impact on effort with the functionalities found in knowledge-based applications as explained in the following: At first, functionalities were identified to derive factors that might impact overall costs for the first and the second pillar of the framework. For the development costs, we considered a top-down strategy to identify these factors: first general functionalities were selected that need to be measured within an application. Subsequently these factors are broken down into more specific, measurable ones. This results in a catalogue which includes a minimal set of functionalities that are inclusive for every knowledge-based application. The catalogue includes components on

the user interface, application logic, storage, or reasoning level. Subsequently we proceeded in proposing a set of cost drivers which are domain dependent and based on which software cost estimation models can be adapted. Generic factors which account for the general software development environment such as personnel and project factors (e.g. personnel experience, tool support, multi-site development etc.) are also included at this stage but typically remain the same. Determining interfaces for the cost estimation models should be done at the cost driver level. Determining the interfaces between these drivers requires an expert evaluation on the relationships between them. The input of the experts can then be used to refine the set of cost factors and to account for the relations in the cost model itself. Integrating domain specific cost drivers into the parametric software cost models is accompanied by an analysis of further cost factors that may require changing or are considered to be unnecessary. Finally, this preliminary model requires an evaluation on the structure of the model w.r.t. to the proposed cost drivers as well as an initial set of quantitative values proposed by experts for its statistic refinement.

For the benefit part of the value prediction, both quantitative and qualitative measures were considered. In brief, the above mentioned parametric models are considered for the quantitative factors for both cost and benefit. For the qualitative part of the benefits, other factors are identified (c.f. [3]). Impact-based weighted values attained from domain experts are then assigned to the factors. Based on that, the qualitative factors can be considered for driving the organizational integration strategy in the fifth pillar. The qualitative factors may then be quantified based on their overall organizational impact.

4 Related Work

The framework presented in this paper is derived from a composition of different frameworks, models and methods, which are taken from an array of different application domains. There are three fundamental frameworks that have a significant impact and therefore standout from others: Aachner's House of Value Creation [4], RFID business case calculation [5] and WIGG'S knowledge management framework [6]. These fundamental frameworks are prominent in their respective domains, but lack the flexibility and customizability required for the domain of knowledge-based applications. Therefore the research focused on defining a modular framework only taking into account relevant parts of the mentioned frameworks. Another aspect, that played a key role in defining the framework, is the classification of value of knowledge-based application components, which is having both intrinsic and extrinsic value at the individual component level and organizational level of the framework [7,8,9,10]. This classification is derived from the ongoing research effort at Iowa State University in the field of consumer behavior [7,11]. As the predictive framework for value estimation is still evolving, it requires additional research and controlled deployment before it can be considered for greater adoption.

5 Conclusions and Outlook

We have laid out a preliminary framework and used it to derive models to assess cost and benefit for a general class of applications. Our framework provides an integrated and unified approach to derive preliminary cost and benefit models. This ensures that both are compatible w.r.t. to identified functionalities and that they can be used to effectively assess costs and benefits in quantifiable terms. Our future work will go beyond expert evaluations and will include an evaluation of a selection of quantitative models based on historical data.

Acknowledgements. The research leading to this paper was partially supported by the European Commission under contract FP7-215040 "ACTIVE".

References

1. Imtiaz, A., Giernalczyk, A., Davies, J., Kings, N.J., Thurlow, I.: Cost, Benefit Engineering for Collaborative Knowledge Creation within Knowledge Workspaces. In: Collaboration and the Knowledge Economy. IOS Press, Amsterdam (2008)
2. Boehm, B.W., Abts, C., Clark, B., Devnani-Chulani, S.: Cocomo II model definition manual (1997)
3. Bürger, T., Simperl, E.: Measuring the Benefits of Ontologies. In: Proc. of Ontology Content and Evaluation in Enterprise, (OntoContent 2008) (2008)
4. Quadt, A., Laing, P., Forzi, T., Bleck, S.: Development of approaches and mobile business solutions for field service. In: Proc. of the 8th Int. Workshop on Mobile Multimedia Communications (2003)
5. Rhensius, T., Dünnebacke, D.: An integrated approach for the planning and evaluation of Auto-ID applications. In: Proc. of the European Workshop on RFID Systems and Technologies, pp. 1–6 (2008)
6. Wigg, K.M.: Knowledge management: an introduction and perspective. Journal of Knowledge Management 1(1), 6–14 (1997)
7. Teas, R.K., Agarwal, S.: The Effects of Extrinsic Product Cues on Consumers' Perceptions of Quality, Sacrifice, and Value. Journal of the Academy of Marketing Science 28(2), 278–290 (2000)
8. Knight, S., Burn, J.: Developing a Framework for Assessing Information Quality on the World Wide Web. Informing Science Journal 8 (2005)
9. Denne, M.J.: Demonstrating IT Payoff 01 (2007), http://www.cio.com
10. Denne, M.J.: Chargeback Demonstrates IT Value in the enterprise 01 (2007), http://www.cio.com
11. Cherubin, S., Terribile, K.: Business Analysts: A Key to Companies' Success 08 (2008), http://www.cio.com

In Quest of ICT Value through Integrated Operations: Assessment of Organisational – Technological Capabilities

Darijus Strasunskas and Asgeir Tomasgard

Dept. of Industrial Economics and Technology Management, Norwegian University of Science and Technology (NTNU), NO-7491 Trondheim, Norway
{darijus.strasunskas,asgeir.tomasgard}@iot.ntnu.no

Abstract. Knowledge based systems improve information interoperability, integration, and knowledge management. Consequently, there is envisioned a set of the associated business benefits. However, knowledge technology as any other information technology is barely an enabler for productivity and resulting benefits, whereas the real drivers are process optimisation and organisational changes. The paper proposes a valuation method to qualitatively assess organisational and technological capabilities. The business value of system implementation is calculated by accounting for uncertainties explicitly defined in implementation scenarios. The valuation method is aligned to a generic process of system and organisational change implementation.

Keywords: qualitative, quantitative, evaluation, business value, ICT value, integrated operations, influence diagram.

1 Introduction

Valuation of information and communication technology (ICT) investments still is a challenging endeavour. Typically, the business value of ICT is defined as the ability of ICT to enhance the company's business performance [31]. This broad definition is decomposed into three main categories that should be evaluated [26]: support of the strategic and operational goals (value is created indirectly); contribution to positive or reduction of negative cash flows (value is created directly); and technological and organisational risks. Furthermore, there are several studies investigating the connection between ICT adoption and productivity. There are evidences showing a positive return on ICT investments using a production function approach [5]. However, ICT investments alone do not suffice to create value, other assets as organisational structure [6], works processes need to be correspondingly changed [13].

Moreover, ICT governance is often viewed as a cost centre [40] and is evaluated through total cost [27] of ownership without analysis of the impact on profitability. The difficulty of valuating intangibles causes that most studies on the ICT business value focus on a structural qualitative frameworks to plot where value is created, without any attempt to derive a monetary value. Decision to invest in a new technology as knowledge based systems (KBS) typically would require a monetary valuation.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 159–170, 2009.

KBS are seen as differentiating, giving competitive advantage that need to be proven. Conventional ICT tools have already become commodities and have proven their utility [44] and are typically adapted without detailed analysis of payoff. Main benefits of KBS are integration of information and reasoning over the encoded knowledge. Consequently, the value is potentially created by improved efficacy of information and knowledge management resulting in better decisions that are made faster through Integrated Operations [1]. Furthermore, a significant part of the value comes from the fact that semantic technologies have the potential to be leveraged in the development of future applications, i.e. having invested in building of an ontology it will provide a good starting point for information integration and service creation in the future. This way, flexibility of infrastructure creates business value [15]. At the same time it brings along many risks as the technology is based on new standards (SPARQL, OWL/RDFS), new tools, and lack of documentation, e.g., risk fitting existing infrastructure, risk becoming obsolete because of changing technology, risk not being accepted, etc. Therefore, flexibility to defer the investment in new technology is an important value source [10] in KBS as technology is not yet mature, as well as organisations are at different levels of maturity with respect to knowledge management [20] and interoperability [17]. Such risks need to be accounted for during valuation of semantic technologies. Value sources of such systems can be classified into four categories [40]: efficiency, effectiveness, flexibility and innovation, each of them requiring different valuation methods. Whereas quality aspect is very important, it is barely a determinant for efficiency and effectiveness.

Furthermore, tasks as annotation of data resources, ontology evolution and maintenance, etc. create new responsibilities and roles within organisation directly affecting the payoff of KBS. Semantic technologies bring a cognitive shift to a company, both to end-user and even to a greater extent to one responsible for maintenance of such solutions. While information technology is the catalyst, the main drivers of productivity are the organisational changes that complement ICT investment [4]. Therefore, current level of organisational and human capital has crucial impact on successful adoption [34, 37, 38] and, consequently, on value of the solution [46]. However, the existing state-of-the-art methods for valuation of KBS are limited to estimating ontology building costs [41], evaluating quality of ontology and its impact on the application performance (e.g., [7, 43]).

The objective of this paper is to elaborate on the prospective valuation method addressing the overall ICT value in integrated operations. The proposed method targets a holistic evaluation of people, process, technology and organisation dimensions. Capabilities for technology adoption are mapped to implementation scenarios with an account for uncertainties during implementation and adoption.

The rest of the paper is organised as follows. We continue the paper with a brief account on the Integrated Operations – context and motivation for our method – where most important aspects (people, process, technology and organisation) are integrated by the means of knowledge technologies. In Section 3 we discuss the proposed valuation method. In Section 4 we review the related work in the area of technology valuation. Finally, in Section 5 we conclude the paper and lay down the future work.

2 Integrated Operations

Integrated Operations (IO) is a vision of future operations in the petroleum industry. IO is defined as "collaboration across disciplines, companies, and organisational and geographical boundaries, made possible by real-time data and new work processes, in order to reach safer and better decisions—faster" [1]. There are even more company-specific modifications of the IO definition, however, most of them treat IO as the integration of people and technology, change of processes and organisation, in order to make and execute better decisions quicker, i.e. IO are enabled by the use of integrated real time data, collaborative technologies, and multi-discipline workflows.

IO spans a range of layers in the ICT infrastructure. A value (at least a fundament for value) is created in each of the layers (i.e., end-devices, communication infrastructure, collaboration infrastructure, and visualisation); however a better economical effect is achieved integrating with a higher level. For instance, data gathering might be seen as a cost generating activity with no direct value in itself (similarly as semantic data annotation), but when data are fed into model-based simulations, model precision can be significantly improved resulting in improved oil recovery. Therefore, each constituent counts separately and, even to the greater extent, in combination with other technologies. The enabling technologies are as follows: *remote sensing* for data gathering; *intelligent drilling & completion* to optimise well productivity, maximise reservoir penetration; *automation* of remote monitoring and control, prediction and production optimisation technologies; *data integration & aggregation* to integrate data from various devices; *communication & collaboration technology* to enable cooperation of geographically dispersed multidisciplinary working groups and reduce decision time; *simulation & visualisation* to assist in optimisation of daily operations.

Many satellite projects have been launched in order to achieve the objectives of IO, e.g. converting the ISO 15926 standard to OWL [35], the AKSIO project [28], the APRIO project [3], the Integrated Information Platform (IIP) project [18], Integrated Information Framework (IIF) by IBM [21]. An exemplary IO implementation is a condition and performance monitoring (CPM) using technical condition index (TCI). CPM is a continuous real-time monitoring of pressure, temperature, and vibration, etc. Technical condition indexes (TCIs) are used for proactive monitoring and maintenance of equipment. TCI is defined in as the degree of degradation relative to the design condition [29]. It is measured in a range, where the maximum point describes the design condition and the minimum describes the state of total degradation. TCI uses integrated data from process control systems, CPM systems, and inspection reports. It is an effective tool for operations monitoring and decision-making as a preventive maintenance [29]. After introduction of TCI for CPM, the work processes are changed by integrating original equipment manufacturer's (OEM) condition monitoring centres with operators' onshore support centres and offshore control rooms. Consequently, people (CPM enables onshore monitoring centres), technology (TCI, ISO 15926 implementation), process (integrated CPM with vendors) and organisation (cross disciplinary data exchange and cooperation) must be changed in order to take full advantage of the technology.

However, changes brought by IO are not restricted to the cooperation among companies, but demand updated and new competencies of employees [16]: *cognitive competence* (knowing why and thinking right); *functional competence* (knowing how

and acting right); *social competence* (social skill and collaborating right); *meta competence* (analytical focus and reflexive knowledge). For instance, interdisciplinary work processes will develop interdisciplinary knowledge and new cognitive knowledge among disciplines. New analytical and simulation tools will increase demand for functional skills and ICT-literacy. Collaborative and complex work processes will increase demand for social and meta competencies. Therefore it is vital to evaluate intellectual capital of the company [2, 38].

3 The Prospective Valuation Method

In order to maximise the benefits, technological advancements should be incorporated in overall organisational infrastructure and related to overall business goals. Work processes and working environment need to be changed with respect to new communication infrastructure and opportunities opened up by improved technology. Technology shall not be analysed separately from human affairs [30]. Therefore, development of the prospective valuation method has the objectives as follows. The method should document value of the knowledge-based system implementation and be apt to assess strategic and operational impacts, taking into account the involved flexibilities and uncertainties. Deployment of KBS without the right incentive systems, training, work process changes, or corporate culture may be worse that ineffective [4], therefore assessment of those aspects is an important task of any valuation method.

3.1 Qualitative Analysis

Consequently, we consider four key aspects that are crucial in KBS implementation. These aspects are as follows.

1. *People (stakeholders).* This dimension analyses responsibility for the process and participation in the processes (i.e. who owns the process and who is involved, what are their roles). The dimension as well concerns organisational and cultural aspects of the sharing and usage of knowledge through collaboration, analysing whether people involved are committed to improve the process and work together. The key issues: development of skills and intellectual capital, collaboration in virtual teams with internal and/or external experts, internalizing and sharing of multidisciplinary knowledge, use of collaboration and knowledge based tools.
2. *Process*. In this dimension methods and techniques for managing the flow of data and information, decisions making and execution are analysed. The dimension concerns tools and systems used for every specific task, leveraging internal, external and vendor expertise within a particular organisational context for collaborative and real time decision-making.
3. *Technology*. This dimension gives an account on tools and infrastructure that are used in daily operations and assist in providing access to and exchanging of information and knowledge. The main purpose of technology is to provide real-time information for on-time decision-making and ensure the optimal execution of the process. This dimension concerns integration and interoperability of various tools for, e.g., condition monitoring and remote diagnostics, sensors and automation.

Further, it assesses quality and scalability of ICT architecture and data management, efficiency of collaboration environment.

4. *Organisation.* This dimension charts and analyses organisational structure, business models (contract strategies), personnel strategy, HSE management. Further concerns along this dimension are legal matters, organisational culture, change management and managerial involvement. This dimension provides settings and context for interaction of above discussed elements, namely people, process and technology.

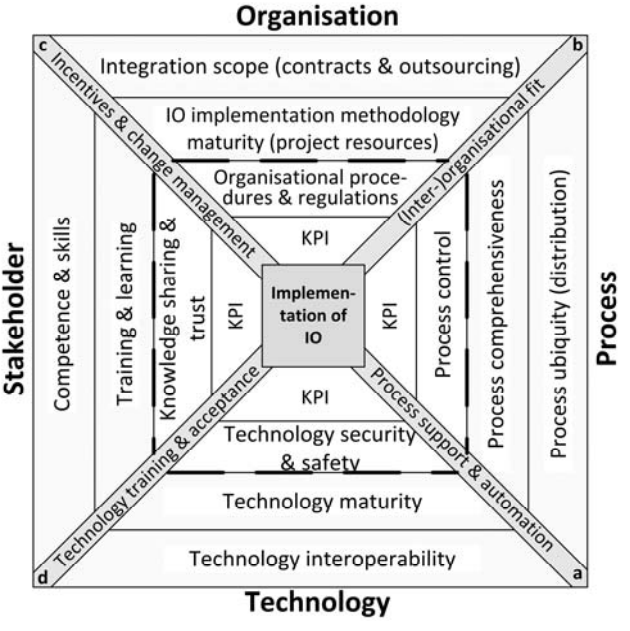


Fig. 1. A pyramid depicting our analytical framework

Thus, for qualitative analysis of IO (or any KBS) implementation capabilities we use the above discussed analytical dimensions to assess the outset situation (see Fig. 1). Each dimension is composed of four layers (two bottom layers are designated for strategic viewpoint, while the upper layers for assessment of operations level) defining the main criteria to be assessed as follows. The bottom layer characterises fundamental properties of the dimension and defines complexity of implementation, i.e. *integration of organisations* (contracts and inter-organisational cooperation); level of *competence and skills*, *process distribution* (ubiquity) and *interoperability of technology* are assessed. The next layer concerns assessment of learning and adoption (fitness), there criteria as maturity of organisational *IO implementation methodology* (established routines and policies for administering IO implementation in practice, shared understanding of change management), staff's *training and learning* capabilities and procedures, *technology maturity* (including technology quality assessment), and *process comprehensiveness* are assessed. The third layer already deals with the

following trust and control criteria at the level of daily operations: *knowledge sharing and trust*, *technology safety and security*, *process control* and *organisational procedures and regulations*. Finally, the fourth layer analyses the main key performance indicators (*KPI*) that are later augmented by the potential effect from the implementation of IO. These analytical dimensions are used in the valuation method (see Fig. 2) for the qualitative description of the situation prior-IO, defining the capability of the IO implementation.

Further, the dimensions in the analytical framework are interconnected by four edges ("pillars") defining the relationship between the connected dimensions. The pillars also distinguish the main analytical scenarios of the IO implementation. *a)* Any project of the IO deployment should consider the level of new technology fitness to the current processes, their automation. *b)* The altered processes should be aligned with the organisational structure. *c)* A thorough change management should be performed including the package of incentives in order to motivate the stakeholders, as well as *d)* training in new technology and studies of technology acceptance should be performed. Those activities are constituents of the new technology implementation project and should be well planned and performed in order to successfully deploy new technology.

Each category is measured by a set of evaluation criteria (they are not detailed here because of space limit). Every evaluation criterion is assigned a score from a scale [1, …, 10], importance weight in a scale [1, …, 3] denoting how important is a particular criterion. The scores are normalised to fall into range [0, …, 1], then the scores are aggregated using AHP (Analytical Hierarchy Process [39]). The final score of a layer in a particular dimension is given in a range of [0, …, 1], where 1 would indicate a 'full capability.

3.2 The Proposed Method

The valuation method is displayed in Fig. 2 using UML Class diagram notation. There the pillars from the analytical framework (Fig. 1) are used as constituents in analysis of the IO implementation scenario. The analytical dimensions are used for qualitative analysis of the IO implementation capabilities. The qualitative analysis together with the parameters of external environment are used as an input for construction of influence diagrams (ID) (see next subsection) to compute the value of current operations. Whereas, the IO implementation scenario (on the right side of the model in Fig. 2), described by the analytical pillars from the analytical framework, quantifies the implementation plan and rate, and associated costs [22]. The implementation scenario is used to estimate the potential effect on KPIs and correspondingly update probabilities in ID. The updated ID is then used to calculate value of the post-IO operations. Finally, the value of the IO investments is computed as a difference between NPV of the current operations and NPV of the future situation.

3.3 Accounting for Uncertainty

Bayesian Network (BN) is a powerful technique for reasoning under uncertainty and representing knowledge [25]. BN provides a natural way to structure information about a problem domain. One advantage of the BN is that it not only captures the

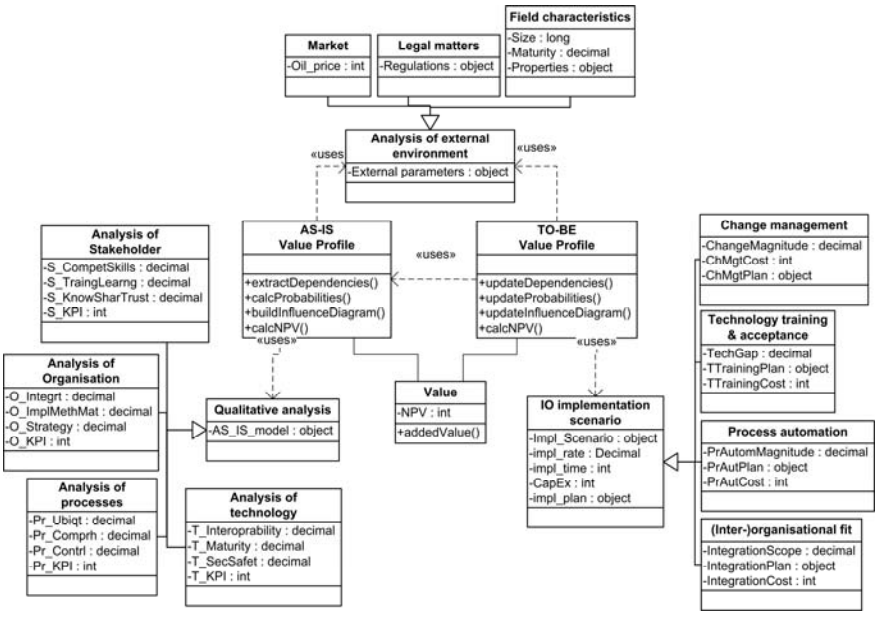


Fig. 2. The proposed valuation method

qualitative relationships among variables (denoted by nodes) but also quantifies the relationships. This is achieved by assigning a conditional probability to each node in the BN.

An influence diagram is a subset of the Bayesian network model and is often used as a compact representation of a decision tree, where a set of variables is observed between each decision. An influence diagram is represented by nodes and arcs. There are four types of nodes: decision, chance, deterministic and value; and two types of arcs: influence (denoting causal influence) and informational (use to represent flow of information that decision is based upon) [25]. An influence diagram is often used as a visual representation of the model. It is a useful technique allowing the model to be built in parts, and for the effects of various parts to be seen without getting in the details of the model [9].

3.4 Functional View of the Approach

The earlier presented framework for qualitative analysis and quantitative valuation method are used in line with organisational change planning and implementation of knowledge based system. Generalised main steps of organisation change management and valuation are depicted in Fig. 3. ICT implementation with organisational change consists of four generic steps. Where step *'As-Is analysis'* provides an input for our qualitative analytical framework (recall Section 3.1, Fig. 1). Further, a roadmap from step *'To-Be analysis'* and *'Qualitative assessment of As-Is'* are used for gap/capability analysis. The capability analysis is directly used to estimate implementation scenarios (recall Section 3.2, Fig. 2). Initial assessment of scenarios might be then used to refine

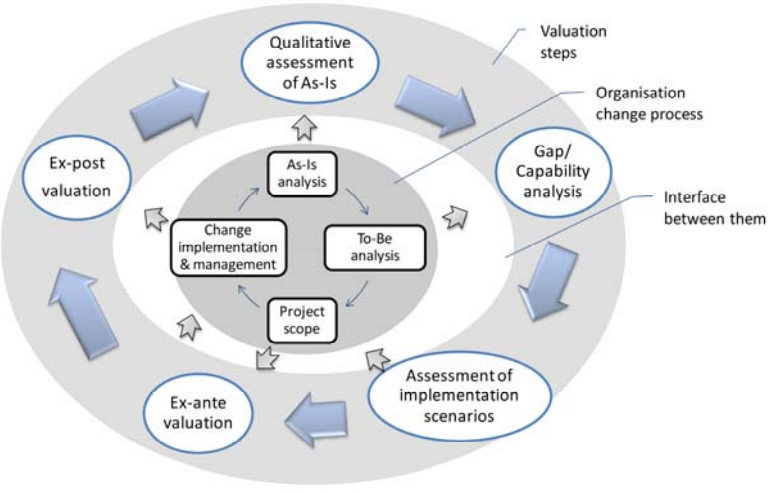


Fig. 3. Generic functional valuation steps

the project scope. The delimited project scope is then used to finalise the implementation scope and forecast the value. Finally, ICT implementation may be used to calculate the achieved value (*'ex-post valuation'*).

Influence diagram of evaluation model is constructed from the analytical output of step *'As-Is analysis'*. Uncertainties and corresponding probabilities are updated after assessment of implementation scenarios. The expected value is calculated as difference between the initial output of the influence diagram and the updated version. At last, the achieved value (ex-post) might be calculated from the revised influence diagram based on the actual implementation.

4 Related Work

ICT valuation typically includes both, qualitative and quantitative methods, although qualitative assessments dominate as a result of many intangible benefits resulting from ICT implementation. The existing body of knowledge on ICT evaluation can be classified into four categories based on the evaluation aspects as follows.

- *Evaluation of strategic value*. The approaches in this category deal with measuring alignment of ICT (e.g., operational activities) with the strategic goals and accounting for its strategic value. The Balanced Scorecard (BSC) [23] method is most broadly used for this purpose. The method allows organisations to measure financial outputs and factors that influenced such financial outputs, e.g., process performance, long term learning and skills development. The analysis is based on four main analytical perspectives: *financial*, *customer*, *internal process perspective*, and *learning and innovation*. The BSC method is applied in various evaluations, generic ICT valuation [42], or ERP system assessment [8].
- *Evaluation of impact on current state of practice.* Here approaches holds on evaluation of changes caused by deployment of a new system and typically are compared with

the performance of legacy systems, e.g. assessing an impact of ontology quality on application [43]. The decision value of an intelligent decision support system is compared with a decision support system without artificial intelligence method in [33].

- *Evaluation of socio-technical aspects.* These approaches focus on usability of technology and evaluate end-user acceptance of technology. There are many evaluation frameworks adopting the socio-technical evaluation perspective (e.g. by assessing perceived efficiency and effectiveness of the tools, intention to use, cf. Technology Acceptance Model [11], IS Success Model [12], User Information Satisfaction [24]). Moreover, the need to include decision maker quality when investigating ICT-performance relationship is emphasized in [36]. Whereas IS Success Model [12] includes analytical aspects as information quality, system quality and service quality. Similarly, Peffers & Saarinen [32] extend the above mentioned three categories by adding successful development and procurement, and successful use and operations. These aspects are important in valuation of KBS where domain experts should become technology experts for optimal adoption of new technology. The unified theory of acceptance and use of technology (UTAUT) [45] aims to explain user intentions to use an information system and subsequent usage behaviour. The theory is based on four key constructs (performance expectancy, effort expectancy, social influence, and facilitating conditions). There, facilitating conditions describe a "degree to which an individual believes that an organisation and technical infrastructure exists to support use of the system" [45]. In similar line, Hallikainen & Chen [19] propose a framework for information system evaluation assessing organisational norms and values, project contingencies, project resources, etc. They argue that technology adoption and success of usage really depend on current level of process/technology and human capital in the organisation.
- *Evaluation based on multi-criteria models and simulation.* The techno-business assessment (TBA) framework [47] is designated to analyse and evaluate technological service platforms and their business value. The analysis starts from qualitative analysis and move toward quantitative cash flow based valuation. Four domain perspectives are analysed: *user model* describes services from the end-user perspective; *business model* conveys a conceptual framework for the business logic; *system model* provides complete details of the services from a system point-of-view; and *technical model* exhibits technology and specifics of implementation. Whereas, information economics [31] assess value linking, value acceleration and job enrichment. Value linking concerns the costs and benefits of organisational changes that are results the new system, but that are not the immediate targets. Value acceleration accounts for the future effects of an investment. For instance, introduction of the ISO 15926 standard in OWL [35] (with XML-based standards for data interchange as PRODML, WISTML) allows faster information integration among companies in the gas and oil industry. Job enrichment in information economics method deals with individual and organisational learning and increased skills. Furthermore, the method assesses the risk of failure of implementing the ICT investment by qualitative evaluation of the business domain (e.g. organisational risk, competitive advantage) and the technological domain evaluation including strategic investment and risk assessment. Though being complex, expensive and cumbersome for small projects [14], the method emphasizes learning and knowledge management that are some of the concerns in the KBS deployment.

Our method is novel by explicitly accounting for four aspects of implementation scenarios that are directly dependable on the human, process and technological capital

of the organisation. Furthermore, the valuation process is aligned to organisational perspective of ICT implementation and change management. Usage of the influence diagram facilitates management and communication of the evaluation model, as well as it allows incorporating and accounting for uncertainty and flexibility related to the implementation.

5 Conclusions and Outlook

We have presented ICT valuation method in the context of Integrated Operations (IO) in the petroleum industry. The method endeavours to provide practical valuation insights on new technology implementation projects relying on the analysis of the organisational-technical capabilities and implementation scenarios. Four analytical dimensions, such as organisation, people, process and technology are analysed from the strategic and operational viewpoint. Then implementation scenarios are modelled with the purpose to simulate and analyse success of technology deployment and organisational change. The analytical implementation scenarios identify possible uncertainties and risks that help to better estimate the potential effects of IO. Finally, functional valuation steps are aligned to organisational perspective of ICT implementation and change management.

A certain future necessity is to conduct case studies in order to validate the method. Calibration of mappings between qualitative evaluation results and impact probabilities used in influence diagrams is a must too. The method also needs to be implemented. The usability and utility of the method a lot depends on its implementation.

Acknowledgement. This research work is financed by the Valuation Methodology for Integrated Operations project, financed by the iO Center (the Center for Integrated Operations in the Petroleum Industry, http://www.ntnu.no/iocenter).

References

1. Andersen, T.M., Vatland, S., Doyle, P.: Oil company of the future: Wireless, real-time data keys to growth for Statoil technology consortium. In: ISA InTech. (April 2008)
2. Binney, D., Guthrie, J., Boedker, C., Nagm, F.: A Framework for Identifying the Intangible Capital Value of ICT Investments. In: Proceedings of the PACIS 2007, pp. 284–298 (2007)
3. Blomskoeld, A., Klingenberg, F.: SemTask – Semantic Task Support in Integrated Operations. Master thesis, University of Oslo, Norway, p. 166 (2008)
4. Brynjolfsson, E., Hitt, L.M., Saunders, A.: Information Technology and Organizational Capital. In: Proceedings of the Workshop on Information Systems Economics (WISE 2007), Montreal, Quebec, Canada (2007)
5. Brynjolfsson, E., Hitt, L.M.: Computing productivity: firm-level evidence. The Review of Economics and Statistics 85(4), 793–808 (2003)
6. Brynjolfsson, E., Hitt, L.M., Yang, S.: Intangible Assets: Computers and Organizational Capital. Brookings Papers on Economic Activity 1, 138–199 (2002)
7. Buerger, T., Simperl, E.: Measuring the Benefits of Ontologies. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM-WS 2008. LNCS, vol. 5333, pp. 584–594. Springer, Heidelberg (2008)
8. Chand, D., Hachey, G., Hunton, J., Owhoso, V., Vasudevan, S.: A balanced scorecard based framework for assessing the strategic impacts of ERP systems. Computers in Industry 56(6), 558–572 (2005)

9. Crundwell, F.K.: Finance for Engineers - Evaluation and Funding of Capital Projects. Springer, Heidelberg (2008)
10. Dai, Q., Kauffman, R.J., March, S.T.: Valuing information technology infrastructures: a growth options approach. Information Technology and Management 8(1), 1–17 (2007)
11. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 13(3), 319–340 (1989)
12. DeLone, W.H., McLean, E.R.: The DeLone and McLean Model of Information Systems Success: A Ten Year Update. Journal of Management Information Systems 19(4), 9–30 (2003)
13. Ebert, C., De Man, J.: Effectively utilizing project, product and process knowledge. Information and Software Technology 50, 579–594 (2008)
14. Farbey, B., Finkelstein, A.: Evaluation in Software Engineering: ROI, but more than ROI. In: Proceedings of the 3rd Int'l Workshop on Economics-Driven Software Engineering Research (EDSER-3 2001) (2001)
15. Fink, L., Neumann, S.: Exploring the perceived business value of the flexibility enabled by information technology infrastructure. Information & Management 46, 90–99 (2009)
16. Fjaertoft, I.: Consequences of implementation of integrated operations on the Norwegian continental shelf. Presentation at the EURAM conference (2006), http://www.npd.no/NR/rdonlyres/299C3416-BF9E-4329-A61B-7C2ABB8E2954/11340/Presentasjon150506.pdf (last visited, 2009.03.02)
17. Gottschalk, P.: Maturity levels of interoperability in digital government. Government Information Quaterly 26, 75–81 (2009)
18. Gulla, J.A., Tomassen, S.L., Strasunskas, D.: Semantic Interoperability in the Norwegian Petroleum Industry. In: Proceedings of the 5th Int'l Conf. on Information Systems Technology and its Applications (ISTA 2006). LNI, vol. P-84, pp. 81–94 (2006)
19. Hallikainen, P., Chen, L.: A Holistic Framework on Information Systems Evaluation with a Case Analysis. The Electronic Journal Information Systems Evaluation 9(2), 57–64 (2005), http://www.ejise.com
20. Hsieh, P.J., Lin, B., Lin, C.: The construction and application of knowledge navigator model (KNM): An evaluation of knowledge management maturity. Expert Systems with Applications 36, 4087–4100 (2009)
21. IBM. Achieving integrated operations and unit efficiency with the IBM chemical and petroleum integrated information framework (2008)
22. Irani, Z., Ghoneim, A., Love, P.E.D.: Evaluating cost taxonomies for information systems management. European Journal of Operational Research 173, 1103-1122 (2006)
23. Kaplan, R.S., Norton, D.P.: The balanced scorecard: measures that drive performance. Harvard Business Review, 71–80 (January - February1992)
24. Kim, K.K.: User satisfaction: A synthesis of three different perspectives. Journal of Information Systems 4(1), 1–12 (1989)
25. Kjaerulff, U.B., Madsen, A.L.: Bayesian Networks and Influence Diagrams. In: A Guide to Construction and Analysis, p. 318. Springer, Heidelberg (2008)
26. Lech, P.: Proposal of a Compact IT Value Assessment Method. The Electronic Journal Information Systems Evaluation 10(1), 73–82 (2007), http://www.ejise.com
27. Love, P.E.D., Irani, Z., Ghoneim, A., Themistocleous, M.: An exploratory study of indirect ICT costs using the structured case method. International Journal of Information Management 26, 167–177 (2006)
28. Norheim, D., Fjellheim, R.: AKSIO - Active Knowledge management in the petroleum industry. In: Proceedings of ESWC 2006 Industry Forum (2006), http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-194/paper3.pdf

29. Nystad, B.H.: Technical Condition Indexes and Remaining Useful Life of Aggregated Systems. Ph.D Thesis, NTNU, Trondheim, Norway, p. 191(2008)
30. Orlikowski, W.J.: Sociomaterial practices: exploring technology at work. Organization Studies 28(9), 1435–1448 (2007)
31. Parker, M.M., Benson, R.J., Trainor, H.E.: Information economics: Linking business performance to information technology. Prentice Hall, Englewood Cliffs (1998)
32. Peffers, K., Saarinen, T.: Measuring the Business Value of IT Investments: Inferences from a Study of Senior Bank Executives. Journal of Organizational Computing and Electronic Commerce 12(1), 17–38 (2002)
33. Phillips-Wren, G., Mora, M., Forgionne, G.A., Gupta, J.N.D.: An integrative evaluation framework for intelligent decision support systems. European Journal of Operational Research 195(3), 642–652 (2009)
34. Plaza, M., Ngwenyama, O.K., Rohlf, K.: A comparative analysis of learning curves: Implications for new technology implementation management. European Journal of Operational Research (2009), doi:10.1016/j.ejor.2009.01.010
35. POSC Ceasar. ISO 15926 in OWL (2008), `https://trac.posccaesar.org/wiki/`ISO15926inOWL (last visited, 2009.03.09)
36. Raghunathan, S.: Impact of information quality and decision-maker quality on decision quality: a theoretical model and simulation analysis. Decision Support Systems 26, 275–286 (1999)
37. Rastogi, P.N.: Knowledge management and intellectual capital: the new virtuous reality of competitiveness. Human Systems Management 19(1), 39–48 (2000)
38. Rodriguez Montequin, V., Ortega Fernandez, F., Alvarez Cabal, V., Roqueni Gutierrez, N.: An integrated framework for intellectual capital measurement and knowledge management implementation in small and medium-sized enterprises. Journal of Information Science 32(6), 525–538 (2006)
39. Saaty, T.L.: Relative Measurement and its Generalization in Decision Making: Why Pairwise Comparisons are Central in Mathematics for the Measurement of Intangible Factors - The Analytic Hierarchy/Network Process. RACSAM 102(2), 251–318 (2008)
40. Silvius, A.J.G.: Does ROI Matter? Insights into the true Business Value of IT. The Electronic Journal Information System Evaluation 9(2), 93–104 (2006), `http://www.ejise.com`
41. Simperl, E., Tempich, C., Sure, Y.: A Cost Estimation Model for Ontology Engineering. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 625–639. Springer, Heidelberg (2006)
42. Stewart, R.A.: A framework for the life cycle management of information technology projects: Project IT. International Journal of Project Management 26, 203–212 (2008)
43. Strasunskas, D., Tomassen, S.L.: The role of ontology in enhancing semantic searches: the EvOQS framework and its initial validation. Int. J. Knowledge and Learning 4(4), 398–414 (2008)
44. Urwiler, R., Frolik, M.N.: The IT Value Hierarchy: Using Maslow's Hierarchy of Needs as a Metaphor for Gauging the Maturity Level of Information Technology Use within Competitive Organizations. Information Systems Management 25(1), 83–88 (2008)
45. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: Toward a unified view. MIS Quarterly 27(3), 425–478 (2003)
46. Vorakulpipat, C., Rezgui, Y.: Value creation: the future of knowledge management. The Knowledge Engineering Review 23(3), 283–294 (2008)
47. Zoric, J., Strasunskas, D.: Techno-business assessment of services and service platforms: Quantitative, scenario-based analysis. In: Proceedings of ICT-MobileSummit 2008, Stockholm, Sweden (2008)

e-Business in the Construction Sector: A Service Oriented Approach

Valentín Sánchez, Iñaki Angulo, and Sonia Bilbao

Robotiker-Tecnalia, Parque Tecnológico de Bizkaia, Edificio 202
E-48170 Zamudio Bizkaia, Spain
Tel.: +34 94 600 22 66, Fax: +34 94 600 22
{vsanchez,iangulo,sbilbao}@robotiker.es

Abstract. The e-NVISION project (www.e-nvision.org) aims to develop and validate an innovative e-Business platform for the SMEs allowing them: to model and adapt in their organizations particular business scenarios requested by their customers and suppliers; to integrate all their enterprise applications following a service-oriented architecture; and to incorporate legal, economical and social services offered by external organizations. This paper provides an overview of the results of the project regarding the definition, implementation and validation of external and integration services within the e-Business platform. The overview includes details on the construction sector e-Business scenarios, the service-oriented techniques used, open issues surrounding actual implementations and applications, and the lessons learned in the field.

Keywords: Semantic Web Services, SOA, e-Business, Business Processes, Construction.

1 Introduction

The e-NVISION project (www.e-nvision.org) aims to develop and validate an innovative e-Business platform for the SMEs allowing them: to model and adapt in their organizations particular business scenarios requested by their customers and suppliers; to integrate all their enterprise applications following a service-oriented architecture; and to incorporate legal, economical and social services offered by external organizations, with the overall goal of facilitating the participation of SMEs, especially those coming from the New Member States, in European e-Business scenarios. The e-Business platform is been validated in the Construction and Building Industry Sector taking as reference several cases executed in 4 different European countries.

From the technological and research points of view, e-NVISION is a blueprint for the future e-Business scenario plus a test case implementation or proof of concept in the construction sector. This blueprint includes a e-business orchestration platform covering internal and external processes, guidelines for interaction between partners to discover each other and collaborate, and a **semantic integration solution to interact with back-end and external services**.

From the point of view of the construction sector, e-NVISION provides a vertical Semantic e-Business Solution for the construction companies that will allow them to

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 171–182, 2009.

participate in e-Business transactions mainly with the other construction companies participating in the same project. This vertical solution is a customization of the general e-Business Platform defined in the Blueprint and it includes a construction related ontology, several construction business processes and **a set of semantic external and integration services**.

2 Construction Sector e-Business Scenarios

One of the main results of the e-NVISION project (http://www.e-nvision.org) is a vertical e-Business Solution for the Construction Sector. This solution provides the means to participate in the future e-Business scenarios of three core construction processes: e-Tendering, e-Procurement, and e-Site. [1][2][3]

The e-Tendering scenario tries to enhance SMEs participation in calls for tenders (e.g. as a group of SMEs or as a Virtual Enterprise) on equal footing compared to bigger tenderers, reducing the work needed to analyse paper propositions, in an open and transparent world-wide electronic market, with mechanisms to look for partners internationally, and supported by trust and quality external organizations.

The e-Site scenario will improve the companies' coordination in the construction time, reporting in an automatic way any change or incident at the construction site to the interested partners so that they can react as soon as possible.

The e-Procurement scenario looks for potential providers both internally and externally thanks to an effective and rational supplier selection model that allows discovering, evaluating and finally selecting the list of providers of a certain schedule of deliveries.

e-NVISION e-Business Model is based on three pillars: business messages, internal business processes and semantic services. These pillars are interconnected through a set of Construction Ontologies as shown in the following figure.

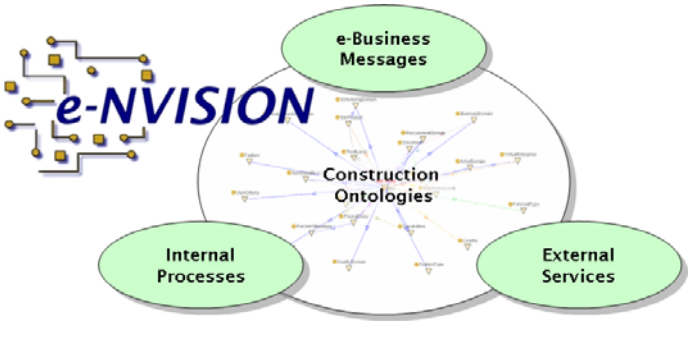


Fig. 1. e-NVISION e-Business Model Pillars

- **Business messages**, which are sent between the different actors that take part in a construction project, in order to:
  - Invite partners to participate in a tender,
  - Involve suppliers and subcontractors in the construction project,

  - Notify design changes, and
  - Notify site incidents
- **Internal business processes**, which define the tasks to be done internally in a construction company when an e-NVISION message is received and the way the business processes are integrated with the back-end systems of the company, and
- **External services**, which provide the information needed to govern the flow of the business processes. The construction company, especially if it is an SME, does not have enough resources to maintain this information internally.
- **Construction Ontology**, which defines the common concepts, entities and relationships.

3 Construction e-Business Services

3.1 Why Services?

In the execution of a business process, companies have to make decisions based on information about the process or about the partners. Questions like these have to be answered during the business process:

- Which are the best candidates for participating in a tender?
- Do we trust this new partner?
- Which are the best quotations received for a task?
- What suppliers are affected by this change in the design?

To answer these questions, the SME has to consult information previously processed and stored by the own SME, ***integration services***, or by external organizations, ***external services***.

It is necessary to dedicate a lot of time and human resources to obtain this kind of information, from previous experiences or from external sources of information, to process it and to register it in the back-office systems.

In the case of large companies, this is not a problem. They have the necessary resources and practically all the information is stored internally. Therefore, the services to access the information will be integration services. In other words, large firms are self-sufficient.

However, SMEs do not have these resources, and sometimes they are not able to participate in some business activities due to this fact. What e-NVISION proposes is to externalise these services to other organisations like Chambers of Commerce, Construction Associations or Clusters, Consultancy Firms, or Local Governments. They have the resources that SMEs do not have and can provide this information at a lower price. The result is that SMEs can obtain similar information to that managed by large companies. The external services form the third pillar on which e-NVISION is supported.

Nowadays, some companies or associations already provide this kind of services. One example is ASCOBI, the Construction Association of Bizkaia, which provides on its web page information about all the tender awards of the last five years. All ASCOBI members have access to this information that allows them to check, for example, the projects carried out by a specific company before deciding whether they are willing or not to participate together in a new tender call. Another example is the

Spanish Company CONSTRUDATA, which sends an e-mail to the companies subscribed to their services informing about the public tenders which are published in the Official Bulletins filtered according to their profile: type of project, estimated budget, locations, etc. By paying a prefixed fee, any SME can receive the information at the same time as large firms.

In the future, a constellation of semantic web services will be available for SMEs. The SME will select to which external services it will be subscribed. The following figure represents this service constellation in an upper level. SMEs can be subscribed to different services, and different organisations could provide similar services competing for SMEs subscription.

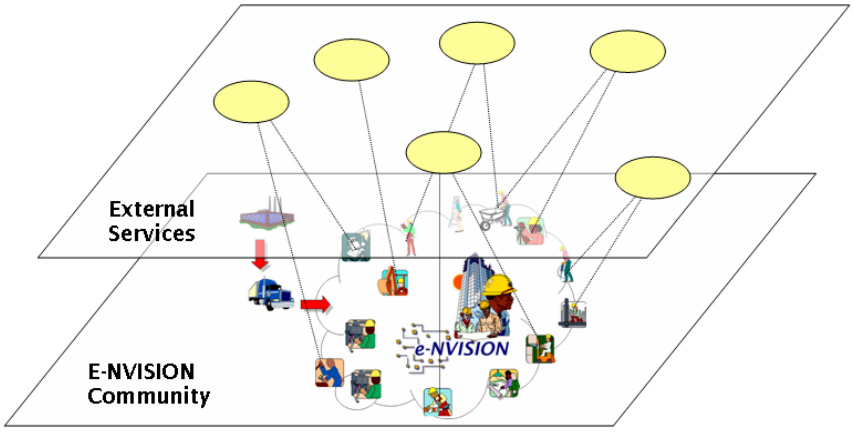


Fig. 2. External services

It is true that currently the number of available e-Business Web Services is very low, and that nearly all of them are accessible through web portals or e-mail. However it is also true that it is very easy to transform these services into web services that could be accessible from, for example, the internal BPEL processes executed inside the e-NVISION platform.

3.2 Services Definition Methodology

A bottom-up approach has been followed in the definition of the e-NVISION business models and the semantic services, starting from the current construction processes based on the know-how of the end users.

The first step was the definition of the future e-Business scenarios. Envisioning ideas, gathered from construction experts through brainstorming sessions, were the base in the definition e-NVISION e-Business scenarios via story telling incorporating "higher" level goals to the process definition, such as customer perceived value, whole life performance, legal, social, economic, trust and if possible, sustainability aspects. During this process we have guaranteed that the work developed was in the line proposed by other construction experts and sources, including e-Business Watch, ECTP and other European Projects related to ICT in Construction [4, 5].

The second step was the identification of those requirements needed for these scenarios from two points of view: on the one hand, the point of view of the SMEs involved in the e-NVISION project and on the other hand, the point of view of the external business environment including public bodies and construction clusters. These requirements have allowed us to define up to 17 external services to the SMEs offered by external Agents, and 12 internal services allowing for integration purposes with enterprise applications like ERP, CRM, e-mail, etc. [6,7,8]

Finally an external user validation was conducted to validate whether the envisioned scenarios were in the line of construction SMEs expectations. More than 30 construction companies, from 5 European countries, have participated in the evaluation [9]. The following chart shows the external services considered more important for the participants.

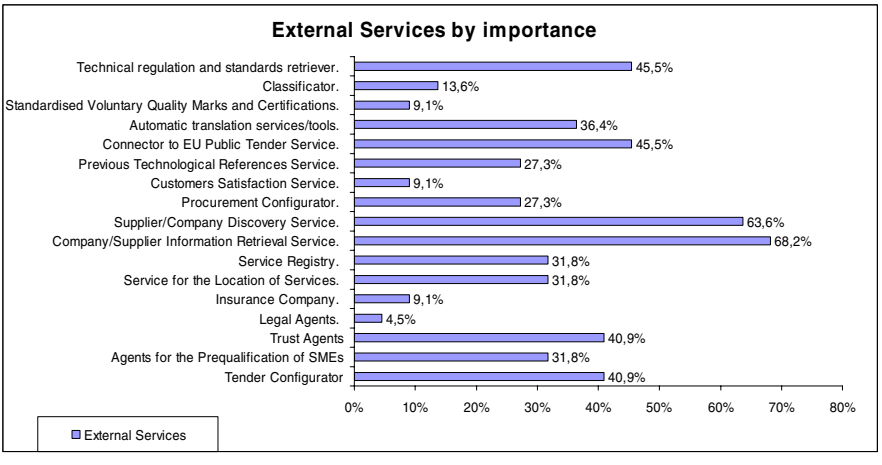


Fig. 3. External services by importance for construction companies

As a result of this validation the following external services were finally implemented:

1. **Company/Supplier Information Retrieval Service.** Since not all companies will be members of e-NVISION, this service covers two aspects: collecting and updating information from the SMEs that are already registered in the e-NVISION platform and capturing information from newly registered SMEs - new members of the e-NVISION platform.
2. **Supplier/Company Discovery Service.** This service gives the possibility of searching for suppliers that offer a certain product, material, service, machinery or equipment. Besides, this service can also provide detailed information about a company, i.e. it allows searching for a company given some kind of information such as its name, or its VAT number, or its organisation identifier.
3. **Connector to EU Public Tender Service.** This service allows automatic retrieval of the tenders published in the TED (Tenders Electronic Daily).

4. **Tender configurator.** This service offers several possible Virtual Enterprises or groups of SMEs with the skills and competences to participate in a tender and to carry out the tender services/works.
5. **Procurement Configurator.** It provides means to configure the list of potential suppliers that can provide a certain schedule of deliveries.

In the same way the following chart shows all the integration services identified and listed by importance for construction SMEs.

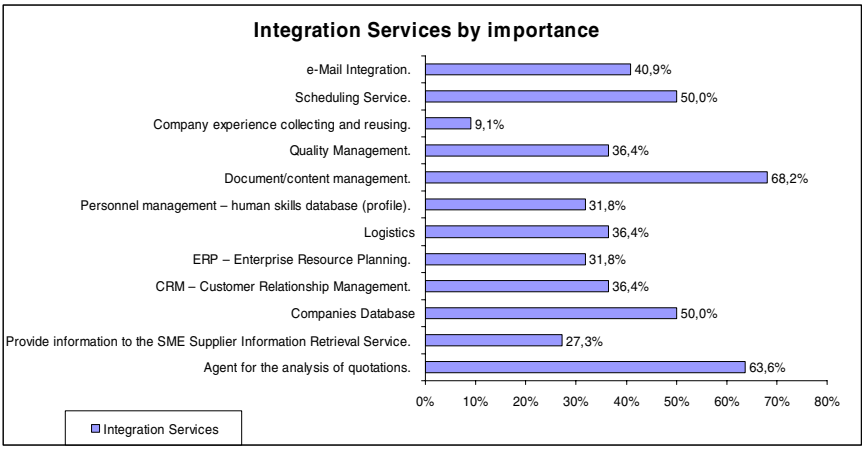


Fig. 4. Integration services by importance for construction companies

1. **Document/content management.** This service is in charge of notifying the interested companies when a new document is provided to the project or when a document changes.
2. **Agent for the analysis of quotations.** Nowadays, the PMC needs time to manually compare the different quotations that the suppliers send to him. This service will enable a first automatic analysis of quotations.
3. **Scheduling service.** When a task is delayed, the entire schedule has to be manually revised to see if other tasks are affected by the delay. This service analyses two schedules and for each of the modified tasks, provides the list of companies that should be notified about the modification.
4. **SME Retrieval Information Service.** Its objective is to avoid a company the fact of having to register or update its information online via a Web form. If the company has this Web service available, the company's information is always available and can be queried at any time.
5. **Companies database.** It provides a standardized way to get the information about business partners (suppliers, customers and so on) stored in the internal systems of the company.
6. **ERP integration.** It is in charge of retrieving the business information stored in the ERP in a standard way.
7. **e-Mail integration.** This service allows integration of the SMEs e-Mail internal solution (backend) with their external e-Business processes.

4 e-Business Platform Technical Implementation

e-NVISION business model is based on a complete set of messages that are interchanged between the different actors that take part in a construction project in order to invite partners to participate in a tender, involve suppliers and subcontractors in the construction project, or notify design changes and site incidents.

From the technological and research points of view, e-NVISION is a blueprint for the future e-Business scenario plus a test case implementation or proof of concept in the construction sector. This blueprint includes a e-business orchestration platform covering internal and external processes, guidelines for interaction between partners to discover each other and collaborate, and **a semantic integration solution to interact with back-end and external services**. The following figure shows the e-NVISION platform architecture:

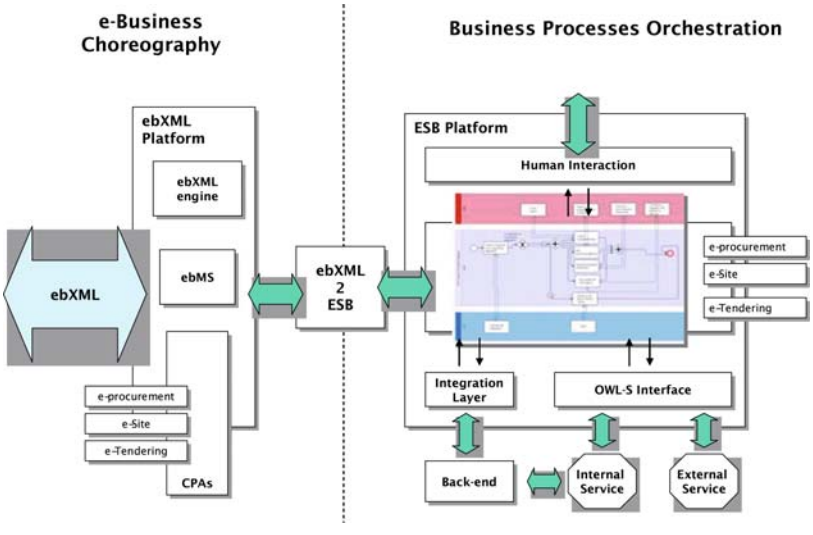


Fig. 5. e-NVISION Platform Architecture

When the e-NVISION platform receives one message the corresponding BPEL process is loaded into the BPEL engine of the e-NVISION platform which executes it.

The BPEL description of a business process can be composed by the following elements:

- Unit actions that are executed inside the BPEL process.
- Decision points which control the business flow according to some criteria. For example in the definition of the supplier list for a construction project "an external searching for supplier service has to be called only if the internal white list of the Main Contractor does not include any".
- Calls to **internal services** which provide information that have been previously processed by the SME and stored in back-office systems.
- Calls to **external services** which provide information that have been previously processes by external organizations to the SME and stored in external databases.

4.1 Services Implementation Approach: OWL-S Semantic Web Services

Web services provide data interoperability by means of XML, SOAP and WSDL standards. However, XML standards specify only syntactic interoperability, not the semantic meaning of messages. For example, WSDL can specify the operations available through a Web service and the structure of data sent and received but cannot specify semantic meaning of the data or semantic constraints on the data. This requires programmers to reach specific agreements on the interaction of web services and makes automatic web service composition difficult.

Moreover, with the proliferation of Web Services, it is becoming increasingly difficult for service requester to automatically find service providers that satisfy its requirements. Some of these difficulties are attributed to the use of XML to describe the interactions and the data in the Web service infrastructures. Although XML guarantees syntactic interoperability of data between applications, it fails to provide semantic operability between these applications. Hence two syntactically identical XML descriptions may have very different meaning, and two syntactically different XML descriptions may have the same meaning. The above restriction poses significant challenges for dynamically interacting with web services.

Semantic Web services solve these problems by providing another layer on top of the Web service infrastructure to supply semantic meaning for Web services. Among the possible options to implement Semantic Web Services, including OWL-S, WSMO, and SAWSDL, we selected OWL-S [10,11] mainly due the availability of tools like a Protégé based editor, several APIs to call services and semantic registry prototypes.

The Semantic Web services augment Web Service descriptions through Semantic Web annotations, in order to support greater automation in service discovery, selection and invocation, automated translation of message content between heterogeneous interoperating services, service composition and monitoring.

Discovery: A program must first be able to automatically find, or discover, an appropriate Web service. Neither the Web Service Description Language (WSDL) nor the Universal Discovery and Description language (UDDI) allows for software to determine what a Web service offers to the client. A Semantic Web service describes its properties and capabilities so that software can automatically determine its purpose. e-NVISON project has implemented an external service to access a OWL-S based semantic service registry and a semantic search service built on top of the semantic registry.

Invocation: Software must be able to automatically determine how to invoke or execute the service. For example, if executing the service is a multi-step procedure, the software needs to know how to interact with the service to complete the necessary sequence. A Semantic Web service provides a descriptive list of what an agent needs to do to be able to execute and fulfill the service. This includes defining the inputs and outputs of the service. e-NVISION e-Business platform is able to call OWL-S based services from a BPEL engine.

Composition: Software must be able to select and combine a number of Web services to complete a certain objective. The services have to interoperate with each other seamlessly so that the combined results are a valid solution. The e-NVISION e-Business platform does not have automatic service composition capabilities. Instead,

the platform implements predefined business processes (BPMN and BPEL) which are able to call semantic web services.

***Monitoring*:** Agent software needs to be able to verify and monitor the service properties while in operation. Monitoring is out of the scope of the e-NVISION project.

4.2 Semantic Web Service BPEL Integration

Nowadays, the Enterprise Service Bus (ESB) implementations available (commercial and open-source) do not include the modules and libraries needed to deal with ontologies and Semantic Web Services. Therefore, after analysing the communication between the Enterprise Service Bus (ESB) and the modules and libraries needed to call the Semantic Web Services, we have provided a simple solution in order to call OWL-S based semantic web services, using the OWL-S Application Programming Interface.

OWL-S API provides a Java API (Application Programming Interface) for programmatic access to read, execute and write OWL-S service descriptions. There are different implementations of this API available but for the development of the External and Integration Services, the OWL-S API implementation provided with Protégé OWL-S editor has been used. This API deals with OWL-S version 1.2.

The OWL-S API provides the means to read and write OWL-S files and to call the underlying Web Service defined using WSDL. It allows calling a Semantic Web Service and getting the results from a Java Program.

The next figure shows the architecture of this approach:

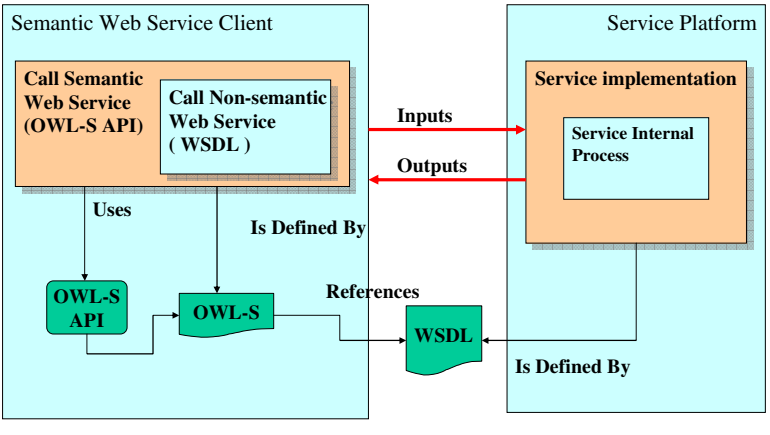


Fig. 6. Semantic Web Service Client tools

Since we use BPEL to implement the scenarios, a method to call the Semantic Web Services from BPEL is required. How to call OWL-S Semantic Services from BPEL is still an open issue. BPEL is a declarative language that offers a set of predefined activities.

The <invoke> BPEL activity is used to call a "standard Web Service" defined using WSDL. But BPEL has no means to call a Semantic Web Service. The best solution would be to add a new BPEL activity to deal with Semantic Web Services but

this would imply modifying the BPEL standard. Another solution is to use the capability of some BPEL engines to execute Java programmes. However, this solution depends on the BPEL engine, so that the result is not BPEL compliant. The simplest solution to that problem, although not the most elegant, is to define a "proxy web service". In this case, the BPEL engine calls a standard Web Service which, in turn, calls the "Semantic Web Service" using the OWL-S API.

5 Business Model for the e-NVISION External Services

External Services will be offered by external organisations to the Construction SMEs: Construction Associations, Cluster, Public Administrations, Chamber of Commerce, Consultancy Firm, etc. Depending on the type of organisation, the Business Model could differ

Construction Associations or Clusters are generally composed by Constructions Companies that join together in order to defend their interest against Public Administrations. In general the Business Model of the Associations is through a membership fee, which is paid annually by all associated members. This fee allows them to receive all the services provided by the Association. In principle, e-NVISION Services could be included into the annual membership fee.

In most cases Public administrations have an Innovation and Economic Promotion Department whose aim is to create permanent jobs and generate wealth in the territory. As part of the activities performed, they provide public services in the region. These services are free or of low cost for SMEs. Therefore, if public administrations want to promote e-Business in the construction sector, they could offer some of the external services in e-NVISION for free such as prequalification of SMEs, legal support, regulation support, connector to public tender service, etc.

Some of the purposes of a Chamber of Commerce are the promotion of trade in their own towns or cities, the collection of information and statistics which may be of use to their members, and the recording of a blacklist for members' reference. Therefore, Chambers of Commerce are suitable to provide some of the external services such as e.g. the supplier/company discovery service, previous technological references service or prequalification of SMEs. As in the case of Associations, the Business Model of the Chambers of Commerce is through a membership fee, which is paid annually by all associated members. This fee allows them to receive all the services provided by the Chamber of Commerce. In principle, e-NVISION Services could be included into the annual membership fee.

Finally Consultancy Firms gather business information which usually includes commercial, financial and marketing information and they offer it to companies in terms of services. The Business Model of Consultancy Firms is through a fee, which is paid per service and per time the company uses the service. Each unit service has a standard price but usually discounts can be obtained in the price of the unit if packages are purchased in advance. e-NVISION will allow consultancy firms to provide new services and to offer existing services through electronic means.

6 Conclusions

From the construction sector SMEs point of view it is clear that there is a need for external services in the e-Business area. The main conclusion of the end user validation survey is that e-Business services, both integration and external, are needed to deploy the e-Business scenarios, according to the majority of the external user groups and industry actors from the Construction Sector that have been interviewed.

Nowadays, external organisations to the SME already offer similar services to those defined in the e-NVISION scenarios. The difference is that these services are provided through web portals or via e-mail or have to be requested manually, as in the case of the Chambers of Commerce where it is possible to request the financial state of a company. e-NVISION Services will be provided in the future with similar business models as the services already offered by these organisations.

Regarding semantic service technology, it can provide the business processes with flexibility and adaptability but, nowadays it lacks the level of standardization and maturity of other SOA areas, like BPEL and Web Services. In order to leverage the whole set of benefits of the semantic services there is a need for:

- A critical mass of semantic web services available on the market.
- Good infrastructure of semantic service registration, search and invocation..
- Mature set of tools to define, implement, discover and use semantic web services.

The e-NVISION results sustainability approach is based on looking for support and promotion from external organizations, including the European Construction Technological Platform (ECTP), National Construction Platforms of the countries involved in the project and Construction Associations involved in the project. The e-NVISION project has been included among the list of some major recent projects related to the ECTP SRA Implementation Action Plan [4]. The ECTP considers that e-NVISION has synergies especially with Items H6 (Collaboration support) and H8 (ICT enabled business models). Therefore, it is the intention of the project consortium to follow the research line in the framework of the ECTP SRA in order to refine the services definition and implementation.

Acknowledgements

e-NVISION project No. IST-028067, “A New Vision for the participation of European SMEs in the future e-Business scenario”, a STREP project partially supported by the European Commission under the 6th Framework Programme in the action line “Strengthening the Integration of the ICT research effort in an Enlarged Europe”. The consortium is composed by ROBOTIKER-TECNLIA, IBERMATICA, ASEFAVE, CSTB, BBS-SLAMA, EUROPARAMA, HRONO, KTU, ITERIJA, ASM, K-PSI, ATUTOR, PROCHEM, ZRMK, CCS, NEOSYS (http://www.e-nvision.org).

References

1. Bilbao, S., Sánchez, V., Peña, N., López, J.A., Angulo, I.: The Future e-Business Scenarios of European Construction SMEs. In: Cunningham, P., Cunningham, M. (eds.) e-Challenges 2007, Expanding the Knowledge Economy: Issues, Applications, Case Studies, pp. 1104–1111. IOS Press, Amsterdam (2007)
2. Angulo, I., García, E., Peña, N., Sánchez, V.: E-nvisioning the participation of European construction SMEs in a future e-Business scenario. In: ECPPM 2006, e-Business and e-work in Architecture, Engineering and Construction, Valencia-Spain (September 2006)
3. e-NVISION project, IST-028067, D2.1, e-Business Scenarios for the Future (January 2007)
4. European Construction Technology Platform (ECTP), Strategic Research Agenda for the European Construction Sector - Implementation Action Plan, July 20 (2007)
5. E-Business W@tch. The European e-Business Market Watch, Sector Reports No. 08-I and 08-II, ICT and Electronic Business in the Construction Industry, European Commission, Enterprise & Industry Directorate General (July/September 2005)
6. e-NVISION project, IST-028067, D2.2, SME Requirements and Needs for the future Electronic Business (January 2007)
7. e-NVISION project, IST-028067, D4.2, Semantic Context Component Architecture (February 2008)
8. e-NVISION project, IST-028067, D5.2, Internal Integration Design (February 2008)
9. e-NVISION project, IST-028067, D7.3a, External User Validation (February 2008)
10. e-NVISION project, IST-028067, D5.2a, e-NVISION Semantic Approach (February 2008)
11. OWL-S: Semantic Markup for Web Services, `http://www.w3.org/Submission/OWL-S/`

Business Patterns in Ontology Design

Freek van Teeseling and Ronald Heller

Be Informed, De Linie 620, 7325 DZ Apeldoorn, Netherlands
{f.vanteeseling,r.heller}@beinformed.nl

Abstract. Trends in ontology design show an interest in the development of Ontology Design Patterns. Most of these are derived from a scientific point of view. We argue that there is a category missing in these pattern libraries, being that of business related patterns. In this paper we describe how Be Informed[1] uses business patterns to develop commercial applications based on semantics and ontologies. Large scale applications ask for a structured approach, including the use of these patterns. We describe three examples of these patterns and show how they are used in existing applications.

Keywords: Ontology Design Patterns, Business Patterns, Taxonomies, Knowledge Systems, Commercial knowledge systems, Structuring Semantics, Metamodel.

1 Introduction

In recent years we have successfully implemented a number of knowledge driven applications, using the Be Informed product suite. These projects have been on a scale of less than one man year work up to 20 to 30 man year work. For an example, see [10]. As projects become larger, and more knowledge engineers and architects become involved, there is a growing need of a recognizable and manageable structure in the knowledge models.

In the field of object orientation and software modeling tools, we see of course the design patterns by Gamma et. al. [1] and Bertrand Meyer describing guidelines for designing OO based software [5]. The semantic community also recognizes the need of a recognizable and manageable structure in the knowledge models [2, 9], resulting in projects like the Ontology Design Patterns [7]. However, main focus of scientific work today is the collection and validation of patterns "inside" an ontology, called Content Ontology Design Patterns. They help modelers to define typical structures between individual concepts in an ontology.

In this paper we present a different category of patterns, the Business Ontology Patterns. These consist of a high level, often abstract pattern, describing a structure in an ontology that resembles the business domain and helps to put the ontology to use. The presented business patterns all resemble some "decision process" or problem solving direction. This function proves critical in the successful deployment of ontologies in a business domain, where execution of models is just as important as disclosure. We

[1] Be Informed is a Dutch independent company building Knowledge Based Systems, using the Be Informed Toolkit.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 183–189, 2009.

describe the patterns on a high level, and from this we translate them into strong-typed metamodels, to be used in ontologies in the Be Informed Product Suite.

One should note that there is no example given in e.g. OWL, simply because the Be Informed Tools use their own, strong typed syntax, and not OWL. One reason for this is that OWL currently lacks certain expressive power, like discussed in [4]. Also, Be Informed uses strong typed relations, which can't be "frozen" in OWL. Another important reason is the strong need to link to content from the models for presentation, which is very hard in OWL. Interchange between Be Informed and OWL and RDF is possible, but results in some information loss (like the content parts).

The Business Ontology Patterns we propose help in designing a structured and maintainable metamodel, as the base for knowledge models. In these model layers, we use patterns too, but they strongly resemble the Content Ontology Design patterns from ODP.org and are outside the scope of this paper.

2 The Business Pattern Framework

The framework we have adopted consists of 4 main areas, which will be discussed in the following section. First, we define a taxonomy as an ontology with only hierarchical relations such as "Subclass of" and "Instance of"; the typical parent-child relations. Ontologies use any kind of relation, including hierarchical, causal and semantic relations and can be multidirectional. Causal relations in Be Informed are strong typed and include "Requires", "Excluded by" and "Numerically Depends On". Obviously they can also have a semantical meaning in a model. We use the term "semantic relation" for those that do have meaning, but are not strong typed and are not used in the inferencers. The inferencers only use the causal relations, whereas navigational instruments can use both.

Typically an ontology, at least in Be Informed applications, show a structure of several core taxonomies linked together with causal and semantic relations. Basis for our patterns is that these core taxonomies represent the key concepts (and their subclasses and instances) in a business domain. This core is represented in the figure below as the circle. Following a pattern in this core we can use the relations between these taxonomies for application in one or more instruments for e.g. classification and deduction. All surrounding elements, and the taxonomies they represent are, in a way, secondary to these core taxonomies.

Secondary elements include conditions and constraints, like calculation models or a semantic representation of legal grounds in law or legislation. The category of properties consists of all elements which cannot be derived but somehow are asked to the end-user or linked IT system. They include variables and context taxonomies like e.g. countries, marital state or a range of possible colors. Computational grounds represent (often fixed) base values of facts like required age or minimum income. The left block in the framework represent constructs for disclosure of information, both in an end-user application (Publication Structures) and Model Validation (reference Structures). The bar at the bottom represents content that can be linked to the concepts and models. For example in case of legal decision-making this is often a link to the relevant (implemented) law and legislation A more thorough description of the entire metamodel framework can be found in [11].

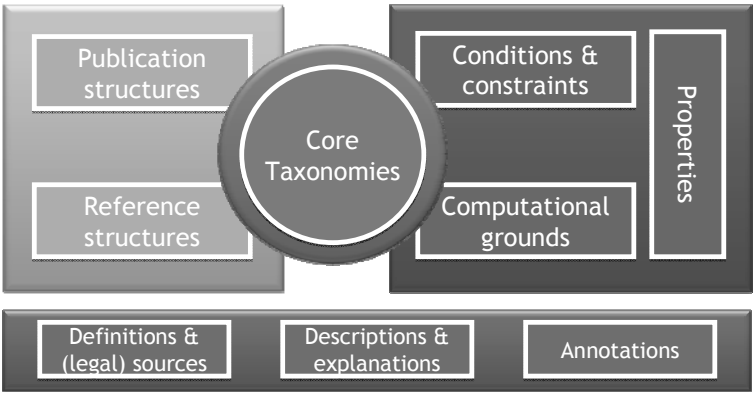


Fig. 1. The Metamodel Framework

3 Patterns in Core Taxonomies

For structuring core taxonomies we have derived a number of typical patterns, to be used in stereotyped applications. This helps the knowledge architect to recognize the key concepts in a typical domain, but also structure and interrelate them so that the inferencer can do its work without overcomplicating the models. In the following sections we describe three patterns, "From Profile to Product Advise", "Legal Decision Processing on permits" and "Selecting and Providing Health Care".

The origin of these patterns lie in numerous projects conducted with the Be Informed tools. Over 40 projects have been examined. In this examination we defined the "typical purpose" of a project, or part of it, and focused on finding recurring constructs in the models. Early projects did not use this strong typing of concepts, but applying learning lessons, newer projects used more and better meta-typed concepts. Here the patterns became more explicit as they often directly reflect a design decision. Matching the typical purpose and the patterns reveal the generic patterns described in this paper. Recent and new projects now start by looking at the pattern library before a meta-model is designed.

In the patterns described below only meta-level relations are depicted. They have a meaningful label while still representing the set of strong typed relations mentioned in the previous sections (requires, excluded by, implied by, etc.).

3.1 From Profile to Product Advise

The first pattern, shown in figure 2, is the simplest one. Main purpose is to support Product advice applications. Typically, these applications try to figure out what kind of customer it is dealing with, and from there, compile a set of advises and corresponding products the customer should buy.

The typical example here is that of an insurance company. In recent trends these companies are shifting towards a more customer-tailored insurance package. To do that, they need to know what type of customer they are serving, for instance a new

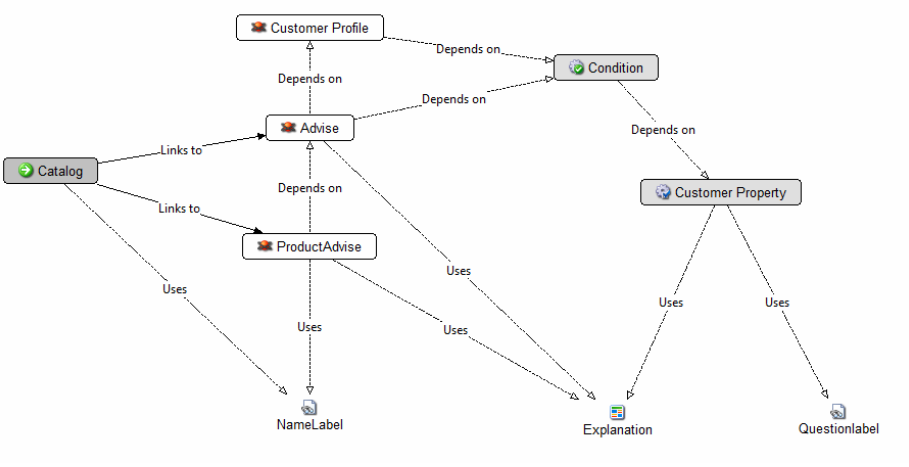


Fig. 2. The Pattern Profile to Product Advice

house owner (Customer Profile). Once they know this, they can advise the –potential-customer what problems they should fix, like covering calamities which can happen to a house and everything in it (Advise). Following that, the product advice is derived, a fire and/or glass insurance(Product Advise).

The pattern can be used for any application going from a client, customer or user profile, ending with specific products. Examples of this pattern can also be seen in several non-business applications, developed for Dutch government, telling citizens to do when confronted with a certain life event, like immigration [13], a close relative dying [8] or starting at a job while still in school [6]. They all reflect the pattern of user profiling, determining what help they need with the life event (Advise), and following that directing them to appropriate organization (Product Advise). Note that the first two examples are multilingual, thanks to the use of special language specific Name- and Question labels and language-specific explanations.

3.2 Legal Decision Processing on Permits

The pattern for a permit request is a bit more complicated. It is designed for organizations (typically government) dealing with a citizens request for a permit.

In these case the citizen often does not know what formal request he should address. That is why the pattern starts with an intention with which the citizen comes to the organization. This is converted into a formal request, derived from laws and regulations. By applying for the formal request the citizen has rights, but also certain obligations, like supplying proof for his situation. Depending on that, a decision is made. The pattern is depicted in figure 3.

Examples here are “I want to build something onto my house”, “I want to come to Holland to study”. These are translated into the formal request, like a building permit or a visa. A set of restrictions can then be compiled together with a list of documents

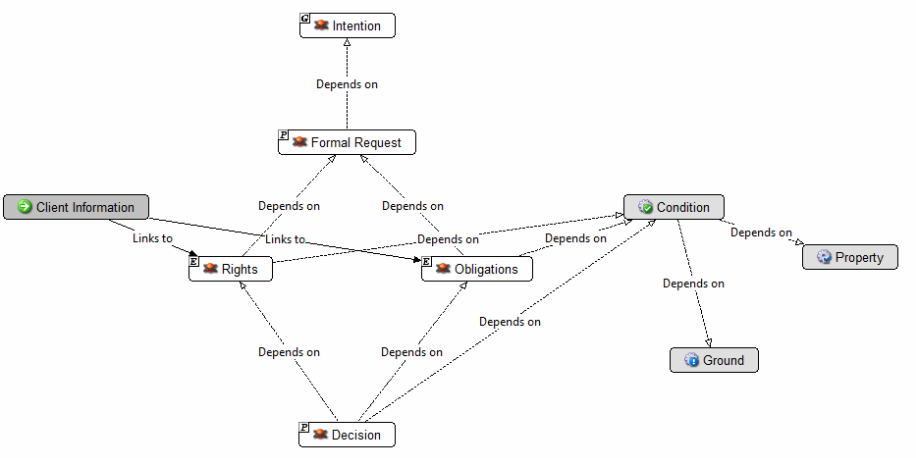


Fig. 3. The pattern legal decision processing on permits

to provide with the application (e.g. an architectural description of the construction work, or proof of diplomas). The first example can be found online as it is developed for the Dutch Ministry of Housing, Spatial planning and the Environment (VROM) [17]. It is worldwide the first online application where citizens can both find information and do a combined formal request to multiple governmental organizations, ranging from national to local [14]. The other example is currently in development for the Dutch Immigration Services [12], as part of an organization wide knowledge infrastructure.

3.3 Selecting and Providing Health Care

The last pattern (figure 4) we describe here is that of selecting proper treatment plans and providing applicable care for clients and patients. In this pattern it is very important to know what problems or illness the client or patient has. This is translated into an indication (formal description of symptoms), and a diagnosis (the illness). From this treatments plan can derived and a set of treatments, which in turn can be fulfilled by Care Products. The pattern also supports the general trend that a treatment can be provided by multiple suppliers, both institutional and commercial. Also, health care has a complex set of financial arrangements, ranging from commercial health care insurances up to governmental policies and funding. This is calculated in a calculation model, based on care policies, laws and regulations.

Examples of this pattern can be found in two applications developed by Be Informed: Providing care for multiproblem patients in e.g. addiction, financial trouble and domestic care problems, which get help from both local government and Health Care insurances in a combined effort [16]. The other example is a (Be Informed) internal training application, diagnosing and treating problems in Social Child Care (autism, learning disabilities, etc.).

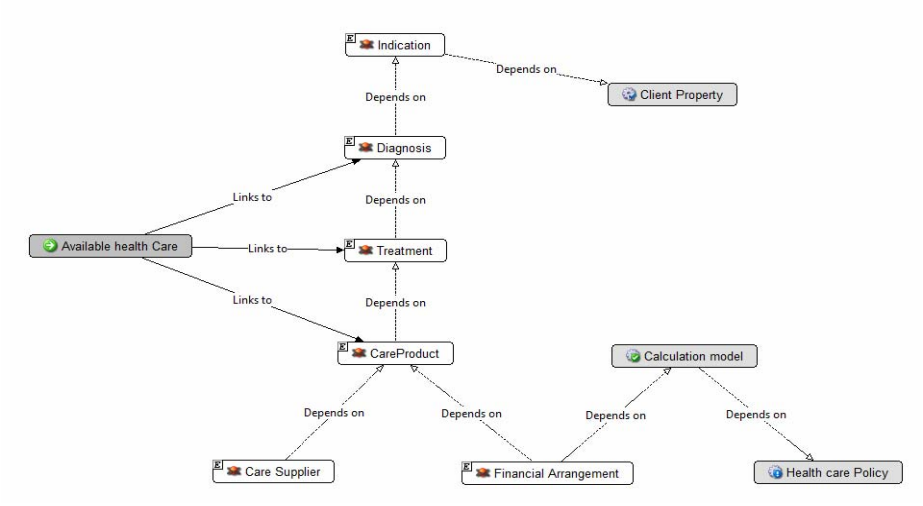


Fig. 4. The pattern selecting and providing health care

4 Discussion and Future Work

In the previous sections we have shown three examples of Business Ontology Patterns. We have described their structure and their relation to successful use in applications. Multiple projects have already made use of these patterns. They have proven very helpful, if not critical, in designing a maintainable complex ontology-based application. It enables knowledge architects to define a structure which can be easily adopted by individual modelers in the project.

Future work will focus on further expanding the library of business patterns. Already we have discovered three more patterns and are now working on formalization. These patterns come from ongoing projects and include “Dynamic treatment plans”, “Generation of legal documents” and the general structure of an online “knowledge repository”. We have identified that multiple projects already use these patterns, but future work needs to focus on a generic structure. Also, several existing patterns might well be reworked to related domains. E.g. the diagnosis for health care might also be applicable, slightly changed to technical domains such as system failure detection.

A drawback of the patterns described here comes from the public applicability of these patterns as they are not described in , and maybe not easily compatible to, other libraries like the ODP.org. Main reason for this that Be Informed, by choice, does not use OWL and RDF. To enable a wider use of these patterns, a description in OWL and/or RDF is useful. The fact that these patterns are on a high business level, is promising in terms of possibility to express the patterns in a variety of notations.

Acknowledgements

In writing this paper we describe work done by ourselves but also many of our colleagues. In finishing a lot of successful projects with Be Informed, they have given us

a large base of projects to derive the patterns from. So without them, these patterns, and thus this paper would not exist. Also we need thank several people in the scientific community who have encouraged us to publish commercial work in a scientific environment as they have convinced us that what we do might also be an interest to others. If you were one of them, thank you!

References

1. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design patterns – Elements of Reusable Object-Oriented Software. Addison Wesley, Reading (1994)
2. Gangemi, A., Euzenat, J. (eds.): Knowledge Engineering: Practice and Patterns. In: Proceedings of the 16th Conference on Knowledge Engineering and Knowledge Management. Springer, Berlin (2008)
3. Gitzel, R., Ott, I., Schader, M.: Ontological Metamodel Extension for Generative Architectures (OMEGA), University of Mannheim(D) (2004)
4. Hoekstra, R., Breuker, J.: Polishing diamonds in OWL2. In: Gangemi, A., Euzenat, J. (eds.) EKAW 2008. LNCS (LNAI), vol. 5268, pp. 64–73. Springer, Heidelberg (2008)
5. Meyer, B.: Object Oriented software construction. Prentice Hall, New Jersey (1997)
6. http://onderwijsenbijverdienen.overheid.nl/
7. Semantic Web portal dedicated to ontology design patterns, http://ontologydesignpatterns.org/
8. http://overlijden.overheid.nl http://www.w3.org/RDF/
9. Pressuti, V., Gangemi, A., David, S., de Cea, G.A., Suárez-Figueroa, M., Montiel-Ponsoda, E., Poveda, M.: A Library of Ontology Design Patterns. NeOn Deliverable 2.5.1 (2008)
10. Verbeek, J., Rensen, G.: A Citizen Centric Government Enabled by Semantic Technologies. In: Semantic Technology Conference, San Jose, CA, May 18-22 (2008)
11. Van Teeseling, F., Heller, R.: Patterns in Metamodelling. Be Informed White paper (2008)
12. Dutch Immigration Office, http://www.ind.nl
13. http://www.newtoholland.nl/
14. http://www.omgevingsloket.nl/
15. Catalog of OMG Modeling and Metadata Specifications, http://www.omg.org/technology/documents/modeling_spec_catalog.htm#MOF
16. http://www.regizorg.nl
17. Dutch Ministry of Housing, Spatial Planning and Environment, http://www.vrom.nl/
18. Web Ontology Language (OWL), specification, http://www.w3.org/2004/OWL/

Towards Models for Judging the Maturity of Enterprises for Semantics

Marek Nekvasil and Vojtěch Svátek

Department of Information and Knowledge Engineering, University of Economics, Prague,
Winston Churchill Sq. 4, 130 67, Prague 3, Czech Republic
{nekvasim,svatek}@vse.cz

Abstract. In recent years, semantic technologies have been included in broader and broader areas of application deployment, and their scope has been constantly expanding. The differences amongst them, however, are often vast and the successes of such investments are uncertain. This work provides a possible approach to the categorization of semantic applications and uses it to formulate a set of critical success factors of the deployment of these technologies in a business environment. Finally, it outlines how it is possible to formulate the maturity models of enterprises for preliminary assessment of the investments into semantic applications.

Keywords: semantic technology, critical success factors, process maturity.

1 Introduction

Before 2001, the web was regarded by the wider community as a mere conglomeration of static web pages, but then Tim Berners-Lee, former director of the W3 consortium[1], introduced in his most famous article [1] the concept of Semantic Web. In the following period semantic technology became popular and nowadays we find its applications in much broader areas than ever before: from applications that integrate data from different sources, support the search in a diverse range of data, derivate new relationships across heterogeneous databases, including the application support of social networking, management decision-making, annotating and indexing of any content, for up to such different tasks as information extraction from unstructured sources, and even so-called Business Intelligence 2.0 [5].

Such a wide range of semantic (i.e. knowledge-based) technologies is mainly caused by the fact that the general understanding of what can be considered as a semantic application is somewhat loose, thus there are no universally accepted definitions. For example, according to [9] any application that stores data separately from the meaning and content files, and in the same time does not have the meaning hard-wired into the application code, can be called *semantic application*. This concept includes the use of ontology languages (such as RDF[2], RDFS[3], OWL[4], etc.) and rule-based systems.

[1] http://www.w3.org/
[2] http://www.w3.org/RDF/
[3] http://www.w3.org/TR/rdf-schema/
[4] http://www.w3.org/TR/owl-features/

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 190–199, 2009.

The W3C Semantic Web Education and Outreach (SWEO) Interest Group collected and published case studies of existing applications and potential use cases that take advantage of semantic technologies in praxis [8]. Thanks to this overview it is possible to gain insight into the current state of semantic applications and their usability it the production environment.

In the pursuit of using the semantic applications in commercial sphere it is necessary to justify the respective investments. However there are many views on what the gains of such investments are. Some of these views are clear and straightforward, such as the analysis of financial characteristics and indexes which compare just the costs and revenues. Others are not so clear but are at least equally important; especially in this case where the benefits of the investment can only be quantified with great difficulties and very roughly. Such views include but are not limited to the added value for customers or the productivity increase of employees. Much more essential by the time of assessing the investment is estimating (or defending) its feasibility and determining the necessary conditions under which the whole project will not be loss-making.

However, as we have already mentioned, the notion of semantic application is very diverse from project to project, hence the conditions of feasibility and potential profitability cannot be set generally (yet, some overviews have also been published, see [7]) but it is necessary to identify some *categories of knowledge-based applications* in the first place. Only once these categories were identified, it would be possible to formulate the requirements, because each kind of semantic application can be substantially different.

This work aims at several objectives and identification of the most common categories of semantic applications is only the first one of them. After the applications have been categorized it will be possible to isolate some of the substantial properties according to the categories. While judging the gains on this level would still be very general, we will propose some possible c*ritical success factors* (CSFs [6]). Therefore the next objective is to establish the most important CSFs of deploying (and developing) the semantic applications.

Finally we will try to outline the manner of how it would be possible to formulate the *maturity models* for deployment of knowledge-based applications (in the sense of the maturity of enterprise processes, according to the original W. Humphrey's work [3]) for some types of such applications based on the aforementioned critical success factors (as a reference model [4]).

2 Categorization of Knowledge-Based Applications

As noted above the knowledge-based applications cannot be considered as a compact area of interest, because indeed they are very heterogeneous uses of the appropriate technologies. The individual applications can differ between each other in many aspects, be it the scale of the used database, number of interested parties, kind of inputs and outputs or the very subject of operation. Because of this the categorization of knowledge-based applications is a multidimensional question.

The particular dimensions (i.e. categorization criteria) however had to be identified. This is where we started the analysis of the mentioned case studies published by the W3C interest group [8] (by the time of publishing this paper 20 were taken into

account). We split the individual case studies amongst 7 workers from Department of Information and Knowledge engineering (every one of them interested in semantic technologies) and went through them in detail. Afterwards, every case study has been discussed in particular by the whole team. Thanks to comparing the individual cases the following aspects of differentiation of the semantic applications emerged (not all however have a direct impact on the forming of critical success factor – this will be considered in part 3). Although none of the analysts had a personal experience with the case considered and only a description was available, the results are credible because of the fact that the SWEO catalogue gathers together cases that represent more than single software a distinctive kind of applications. The categorization criteria we found are these:

- **Information sources.** The semantic character of considered applications directly implies that at least one knowledge model (ontology or taxonomy) has to be used. Some applications also use other knowledge models or even expect a variable knowledge base. Apart from that the applications can of course also use other data of various kinds. Knowledge-based applications can be divided according to whether they process structured knowledge, structured data or unstructured data.
- **Data source provenance.** Semantic applications can be distinguished according to whether the information they are working with arise in other systems (or are already available in a structured form) or whether they are created specifically for this system. If the data are created exclusively for the semantic system we can further distinguish the cases where this is done manually, automatically from other sources or as a side effect of other activities (such as normal user behavior).
- **Accuracy of inputs and outputs.** Considering the semantic applications we find different approaches of transforming inputs to outputs. Here the applications can be divided e.g. into those firmly relying on full precision of data, applications that expect that the data may be incomplete but do not expect them to be inconsistent and do not work with uncertainty, and finally, applications that include treatment of uncertainty.
- **Domain-specificity and reusability of applications.** Because of the separation of data from their meaning the semantic applications should be much less domain-dependent than conventional solutions, but even here there are exceptions, which include, for example, specific interfaces tailored to a specific domain or particular treatment of data on the application level.
- **Number and kind of users.** Users of semantic applications may constitute of unprofessional individual users, professional users (domain experts), knowledge experts and management. Applications can also be distinguished according to whether they are intended for individuals, working groups or thousands of users in social networks.
- **User × provider relationship.** Here we managed to identify several options for operating the applications: the user is an individual and operates the application for his/her own use; there are a few users and they are more or less equal subjects or form a social network and the operation is granted commercially, by the

community or non-profitably; the users are the customers of the provider; and finally the users are the employees of the provider. For the last two possibilities we can distinguish cases where the operation of the application is the core business of the company and where it is only a supporting process and can therefore be considered as a possible target for outsourcing. The cases when the operation is ensured by the community can be broken down by whether the operation is centralized or decentralized.

- **Frequency of access to the application and its availability.** Applications may be used continuously (24/7), at random, regularly or by a single opportunity. Furthermore, a distinction must be also made by the availability of such applications: either the application must be available constantly, in defined intervals or on demand (e.g., the reactive manual start of the application).
- **Subject of operation.** From the analysis of case studies we managed to identify several main types of activities of semantic applications. These are data indexing, data integration and reasoning. These activities are, however, in most cases the means rather than the purpose of the activity (the exception is the integration of heterogeneous data). From these, we can derive several other activities which support the main purpose of the application, for example, they are enabling better searching capabilities (indexing + integration), heterogeneous database browsing and navigation in the domain (integration + indexation), recommending new relations among entities (reasoning) and allowing the adaptability to change the systems' data structures (data integration).

By sorting the considered case studies we can find out how often some specific values of the proposed criteria are seen in the real-world applications. (The sorting was done by filling a prepared form by the responsible analysts in the first place and the results were then reviewed and normalized by one of the authors.) The relative frequencies of these values are shown below on Fig. 1.

Of course, one can imagine a semantic application that is classified by the mentioned aspects more or less arbitrarily, but in the presented case studies certain coincidences and clusters can be identified amongst the various aspects. This sorting thus enables us to and to identify and name some basic archetypes of semantic applications, based on examples clustering:

- **"Improved search engine".** These applications focus on indexing the data, often associated with integrating data from various other systems where the data are generated automatically. These automatically acquired data are also often accompanied by manual annotation. Applications of this type work with both structured data and unstructured data (using automatic filters and wrappers).

  The main benefit of these applications is to enable searching in heterogeneous data base and the creation of complex queries without the need for a priori knowledge of data structures. In other aspects, however, they can vary greatly, so such different applications as support for annotating and searching of files on a personal computer[5], a public portal for searching for findings of Chinese

[5] http://nepomuk.semanticdesktop.org

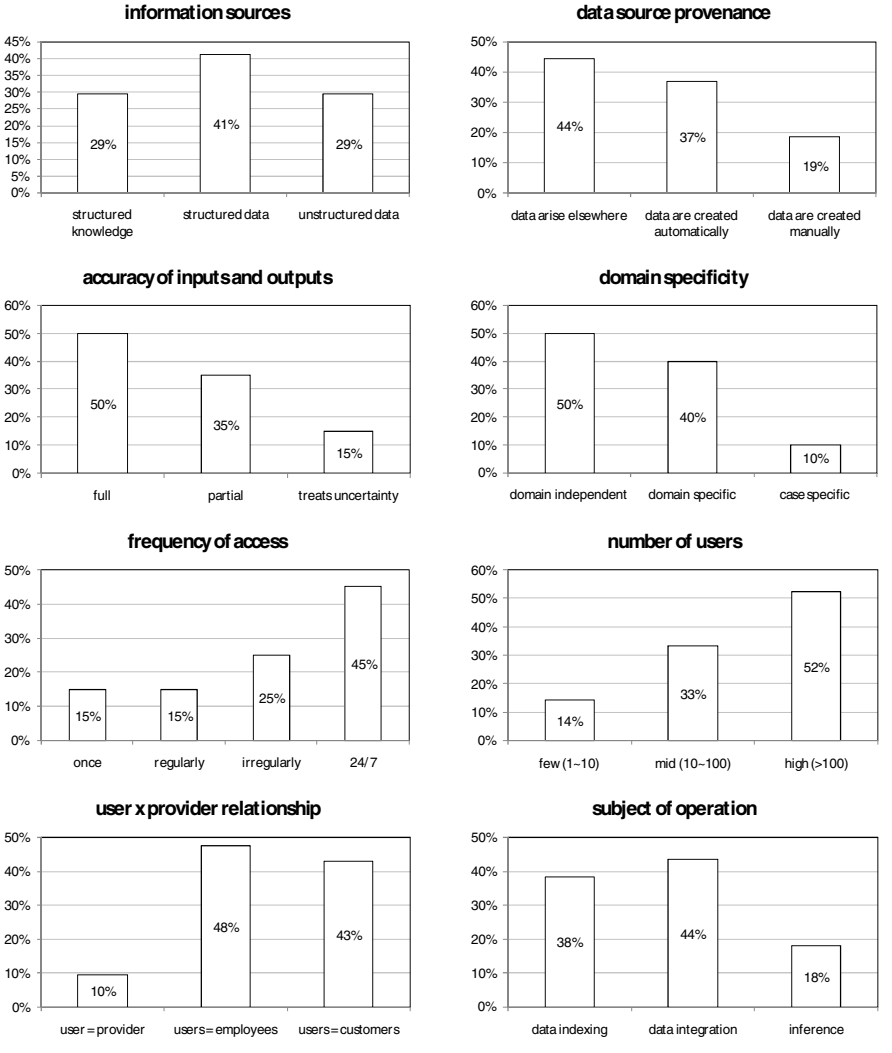


Fig. 1. Relative frequencies of certain criteria values

medicine[6] and management of sound recordings archives by a Norwegian radio station[7] can by classified here.

- **"Data-browsing interface".** These applications follow the abilities of the previous archetype, but enhance not only the possibility of displaying diverse content (videos, articles, chemical formulas, etc.), but also the possibility of visual data browsing, regardless of their structure. These applications are

[6] http://www.cintcm.com
[7] http://www.nrk.no/

mainly focused on the use by professionals and are operated either commercially for internal use or non-profitably to support a professional community (and simultaneously promoting the technology). Examples of this archetype can be e.g. systems for the aggregation of medical data, whether in order to facilitate the treatment of patients[8] or achieving savings in the development of new drugs[9] or portal for the association of programming knowledge by Oracle[10].

- **"Recommending system".** The nature of these applications is the derivation of new relationships between entities. Moreover, apart from all other types of source data these applications often utilize data that are automatically generated as a side effect of normal user activity, which enables, inter alia, to propose new relationships on the basis of the current users' context. The user is often the customer of the provider, be it either as a paid service, public service (e.g. designing of individual city tours in Zaragoza[11]) or a commercial way to personalize advertisement targeting together with the provision of services (such as a system for recommending services to users of mobile devices[12]). Applications in this category often work with uncertainty, thus one can also include a variety of expert systems.
- **"Data interchange framework".** Operations of applications in this category (because of their nature) are distributed, thus these knowledge-based systems "only" allow to unify the structure of data exchanged between the participants, regardless of their content. This content can evolve over time and be adapted to the needs of a particular bilateral exchange relationship and yet be transmitted in a standardized format. An example is the initiative for the establishment of semantic data interchange in the oil and gas industry[13].

Surely it would be possible to discover other archetypes of semantic applications; however, we consider these four to be the most usual. Of course there are also applications that cannot be assigned to any of these archetypes, as well as others which, on the contrary, lie in between two or more.

3 Critical Success Factors

As already indicated in the section 2, by the synthesis of the risks mentioned in the individual case studies [8] it is possible to outline the most frequent critical success factors in the development of the semantic applications and their deployment into the production environment. These factors are not universal, but each only applies to a particular group of applications given by the aspects of their categorization, referred to above. Critical factors for success of semantic applications identified so far are:

[8] http://www.pharmasurveyor.com/
[9] http://www.lilly.com
[10] http://otnsemanticweb.oracle.com/
[11] http://www.zaragoza.es/turruta/Turruta/en/index_Ruta
[12] http://www.w3.org/2001/sw/sweo/public/UseCases/SaltLux-KTF/KTF.pdf
[13] http://www.w3.org/2001/sw/sweo/public/UseCases/Chevron/

- **Correctness of the core ontology/taxonomy.** This factor holds for all knowledge-based applications and the more complicated and less volatile the used model is the more crucial is its correctness. Achieving this success factor entails the need for recruitment of high-quality analysts and knowledge engineers in the phase of development and deployment of the application, which involves considerable costs. The quality and reach of the used ontologies is not limitless; apart from the costs of creation it also has other more structural constraints (see [2]).
- **Sufficiently steep learning curve of end-users.** This applies to applications that have individual end users. Semantics used in this type of applications entails quite atypical method of control compared to standard applications and the learning curve is rising very slowly. Not only a comprehensive and intuitive user interface of the system is a must, but also clarity and accuracy of the outputs and results is vital for the users' work.
- **The potential of possible benefits to compensate the temporary reduction in productivity during implementation and learning** (as well as operating costs). **T**he benefits of the applications are very diverse and often very vague (in contrast with conventional solutions) and thus can be hardly estimated (and quantified) at the time of the deployment of the system. Operating costs are mostly comparable to conventional applications, but in the phase of deployment, it is necessary to count with temporary decrease of productivity of the users (see previous item). For an application to be successful, this temporary decrease should not be so serious that it overshadowed its potential benefits.
- **Will and discipline of all parties to use the same knowledge model. I**n case the operation of application is distributed, it is necessary that all interested parties use a central shared knowledge model. There is therefore a potential risk in terms of the need to negotiate on its form and content.
- **Synchronized distribution of central ontology.** Gradually, there may be modifications of the central knowledge model that arise subsequently and if the operation is distributed, it is necessary that these changes are properly disseminated amongst interested parties, or else this could lead to some inconsistencies. While these changes and modifications take place the previous item still holds.
- **Sufficient number of users.** If the respective semantic application is based on social data, its success is conditioned by the existence of a large enough number of users that produce this data. The risk in this case occurs in the form of necessary expenses for the promotion of an emerging system.
- **Users' motivation.** This critical factor occurs at two levels. The first is in the time of the introduction of a new application while the user experiences a negative stimulation in the form of the slow rise of the learning curve. The user thus lacks the motivation to learn to deal with the system in the first moment. Moreover the user that does the work is not always the one who benefits from it (discussed in [2]) which can be of a further burden. The second is in the actual phase of operation; a common source of data for the semantic systems of all sizes is manual annotation, whose creation is up to a certain point very labor-intensive for the users. Partial source of motivation may be a potential benefit of the better results or a facilitation of work in the future. In addition the user can

be motivated by the possibility of using the experience gained elsewhere, while even the most different semantic applications use similar technologies (e.g. SPARQL querying).

- **Sufficient supply of data.** For applications that use some reasoning having a sufficient data source is very essential for providing beneficial results (i.e. utilizing the added value of semantics). Even for applications based on data indexing, having enough data is critical to the success, because for small volumes of data they give similar results as traditional methods but with higher initial costs.
- **Diversity of sources and forms of data.** The greater the richness of the knowledge-modeling language (namely, its part actually used in the application), the more beneficial results can be produced by applications based on the derivation of new relationships. Likewise, the greater is the diversity of data content the more useful are the results given by applications performing semantic integration. The use of semantic technologies on trivial systems will therefore likely not pay off.
- **Maintaining at least the same accuracy of results as the sub-systems.** Applications that integrate data of some source systems are at risk of finding an inconsistency in the aggregated results. The functionality of semantic applications itself is not subject to consistent data, but the possible inconsistency should be expected in the design. Good estimation of the reliability of data sources is thus crucial at this point.
- **Reliability of parsers and wrappers.** If the application handles unstructured data, it is dependent on the output of parsers and wrappers of various content and, where appropriate, the natural language processing systems. Here again the same applies as in the previous paragraph, namely that it is necessary to correctly estimate the reliability of the information obtained in such a way.

Of course, these critical factors will be weighted differently it the scope of different applications. If, for example, the source application collects data automatically and passes the outputs to the user in almost natural language and in an appropriate context, we can expect a relatively steep learning curve, so that the period of reduced productivity is quite minimal and as a result it will be compensated enough even by minor benefits. These universal critical success factors can only be taken as starting points when considering a particular case.

4 Future Work – Maturity Models

Maturity models [3] have developed over the past two decades in order to enable to assess the readiness of enterprise to implement some kind of structural investment. Most commonly they are used in the deployment of any IT applications such as CRM systems, ERP and Business Intelligence. In our opinion it should be possible on the basis of the above aspects of categorization and the associated critical success factors to establish enterprise maturity models for the deployment of a certain type of semantic technologies. Although these models would be without factual content and without the target statement, they could be formulated by concrete examples and at least for

each archetype would thus make it possible to set a certain level of requirements for an enterprise, which should be met in order to consider the feasibility of the solution.

An example maturity requirement for the archetypal semantic search engine could look like this: *If an enterprise uses a single source of data and a proprietary data structure, then it is unprepared for the introduction of this kind of system. If it is using multiple systems with heterogeneous data structure, the introduction of search engines with semantic indexing can bring some improvements to the search results. The enterprise achieves next level of readiness if it uses more systems with a standardized data structure; in such case it can start thinking about the integration of these systems with a semantic data exchange, etc.*

Before formulating such exemplar maturity models the critical success factors need to be evaluated because their validity is vital. We plan on performing a detailed survey of the successful semantic technology projects (based on the SWEO catalogue) and test whether the critical factors hold. As a side effect such a survey will help quantify exact boundaries and limits in our dimensional categorization approach. Only after verifying the CSFs we can move onto formulating the maturity models.

5 Conclusion

This work outlines a dimensional method of categorization of semantic applications, and with the help of case studies published by the W3C interest group then identifies some of the basic archetypes of semantic applications. Given the proposed categorization we then formulated the most critical success factors for the deployment of semantic applications in a business environment, together with the validity limits of these factors.

Finally we also outlined how the proposed critical factors and the categorization can be used in the process of finding valid models for maturity of enterprises for semantics. The formulation of such specific models is the subject of our future work.

Acknowledgement

We wish to express our sincere thanks to our department colleagues Tomáš Kliegr, Jan Nemrava, Ondřej Šváb-Zamazal and Jan Zemánek, who helped us a lot with the survey of SWEO case studies, and, in particular, to Ota Novotný from the Department of Information Technology, who offered his valuable consultations on CSFs and maturity models.

References[14]

[1] Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web, Scientific American (May 2001), http://www.sciam.com/article.cfm?id=the-semantic-web

[2] Hepp, M.: Possible Ontologies: How Reality Constrains the Development of Relevant Ontologies. IEEE Internet Computing 11(1), 90–96 (2007)

[14] The hypertext links are valid on the date of February 1st, 2009.

[3] Humphrey, W.: Managing the Software Process. Addison-Wesley Professional, Massachusetts (1989)
[4] Novotný, O.: IS/ICT Management Reference Model. Revista de Engenharia de Computacao e Sistemas Digitais 3(3), 53–61 (2007)
[5] Raden, N.: Business Intelligence 2.0: Simpler, More Accessible, Inevitable, Intelligent Enterprise,
http://www.intelligententerprise.com/showArticle.jhtml?articleID=197002610
[6] Rockart, J.F.: Critical Success Factors. Harvard Business Review, 81–91 (1979)
[7] Sauermann, L.: Benefits of Semantic Web.. for you, DFKI GmbH (2008),
http://www.dfki.uni-kl.de/~sauermann/2008/04/benefits/
[8] Semantic Web Education and Outreach (SWEO) Interest Group: Semantic Web Case Studies and Use Cases,
http://www.w3.org/2001/sw/sweo/public/UseCases/
[9] Staab, S., Studer, R.: Handbook on Ontologies - Preface. In: International Handbooks on Information Systems. Springer, Heidelberg (2004)

EeLT 2009 Workshop Chairs' Message

Sławomir Grzonkowski[1] and Tadhg Nagle[2]

[1] Digital Enterprise Research Institute, National University of Ireland, Galway
IDA Business Park, Galway, Ireland
slawomir.grzonkowski@deri.org
[2] Department of Accounting, Finance and Information Systems,
Univeristy College Cork
T.Nagle@ucc.ie

1 EeLT 2009

Lately eLearning is experiencing massive disruption through the continuous introduction of emerging technologies. However, there still is substantial scepticism about the effectiveness of these technologies within eLearning consumer and provider communities. Nowhere else is this more evident than in the area where Web 2.0 meets eLearning. Only now wikis are becoming accepted as a valid platform for knowledge transfer, both to and from eLearning consumers. Yet, the industry (as a whole) still has to cross the chasm with regards to more sophisticated Web 2.0 technologies such as SecondLife. Within the IS and IT community there is always the argument that academic research has only limited impact on practitioners. However, as indicated in the previous edition of this workshop [1], there is a need for strong and objective research on providing guidance for the eLearning industry on how best to analyse, implement and utilise these new emerging technologies. Targeting this niche, Janc and Olejnik (2009) present a high-level framework through the development of a Web 2.0 language tool. This provides a increased understanding for the language domain of eLearning but findings can also be extrapolated to the wider industry.

Reference

1. Flejter, D., Grzonkowski, S., Kaczmarek, T., Kowalkiewicz, M., Nagle, T., Parkes, J. (eds.): BIS 2008 Workshops Proceedings: Social Aspects of the Web (SAW 2008), Advances in Accessing Deep Web (ADW 2008), E-Learning for Business Needs, Innsbruck, Austria, May 6-7. CEUR Workshop Proceedings, vol. 333. CEUR-WS.org (2008)

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, p. 200, 2009.

eLearning in the Web 2.0 Era – Lessons from the Development of the Lingro.com Language Learning Environment

Artur Janc[1], Łukasz Olejnik[2], and Paul Kastner[1]

[1] Lingro Inc, Berkeley CA 94702, USA
{artur,paul}@lingro.com
[2] Adam Mickiewicz University
Poznan, Poland

Abstract. Recent advances in Web technologies have enabled the creation of new kinds of eLearning applications which have the potential to supplement existing solutions. In contrast to traditional educational software often developed by large corporations leveraging their existing customer base and distribution channels, Web 2.0 educational platforms are commonly created by small technical teams with limited advertising and development budgets. In this work we use our first-hand experience from developing Lingro, a popular on-line language learning environment, to compare and contrast the traditional and new approaches for creating eLearning systems. We introduce Lingro as an example of a next-generation language learning application, and use it to illustrate Web 2.0 concepts and industry trends, as well as their relation to traditional practices. We provide a general overview of relevant development and business issues, analyzing changes in software development methodologies, ways of acquiring and creating educational content, learning styles and financial and marketing information.

Keywords: language learning, informal learning process, eLearning models, Web 2.0.

1 Introduction

Recent years have brought significant innovations in on-line applications based on Web 2.0 technologies, of which a large group are eLearning solutions. The conditions in which Web 2.0 software is created are often significantly different from the traditional model, due to several business and development methodology issues[1]. As the barrier-to-entry for building innovative educational platforms is constantly lowered, the understanding of mechanisms governing such Web-based eLearning applications is crucial. In this work we survey the differences in methodologies using our first-hand experience working on Lingro – an internationally-recognized on-line language learning environment[10,11].

Launched in November 2007, Lingro is a learning platform allowing students to read websites in foreign languages, with the ability to quickly translate

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 201–211, 2009.

unknown vocabulary with a single mouse click. Lingro provides dictionary information, including word translations and definitions, based on *open content* sources, and allows users to easily make additions and updates to the dictionary database. The system stores words translated by each user for future reference, and provides capabilities for creating word lists and using them in educational games. The development and business methodologies applied during the creation of Lingro closely reflect approaches taken by other Web 2.0 projects, and serve as examples in our analysis of the industry.

In this work we present the lessons learned during the development of Lingro and how they compare to traditional educational software and other startups in the Web 2.0 eLearning space. Specifically, we address issues such as software development methodology, methods of acquiring educational content, learning styles and marketing.

The approach we have taken in our analysis focuses on practicality and attempts to provide researchers and industry practitioners with concrete examples of differences between the old and the new approaches. Our work is the first attempt at describing and comparing those different methodologies for creating eLearning solutions based on a case study of a Web 2.0 application. In addition to contrasting traditional and current approaches, we also highlight the possibility of combining their capabilities to establish an eLearning base for the information society showing a need for new pedagogical approaches.

2 Background

Language learning software and services have long accounted for the majority of expenditures on education worldwide[17]. The demand for language abilities has been increasingly growing because of globalization processes – connecting individuals through networks (computer and social) poses a new challenge, resolved only by investing in language education. However, most established computer-based approaches do not utilize the full capabilities offered by Internet connectivity. We provide a brief comparison of former and current approaches to computer-based language learning.

2.1 Past Approaches

Traditional language-learning software for individuals was distributed as floppy disks, and later – CD-ROMs. The approach was introduced in the 1990s and enabled the use of personal computers in the language learning process. While it was a logical step in the development of immersive educational environments, most available software lacked any capabilities for customization and failed to acknowledge individual differences in learning styles and needs of users.

In the late 1990s, along with the growth in the popularity of the Internet, the rapid expansion of discussion boards such as Usenet and Web-based forums was observed. Users became able to interact with others and exchange of experience was facilitated. Services such as Web-based dictionaries and the DICT protocol

appeared[14], and allowed to query for a word definition or translation in an on-line database. These services, however, did not provide a significant learning component, and were most often used as reference sources.

The demand for concrete eLearning solutions was met with the notion of *computer-assisted language learning*[2], which was the first technological approach to learning using both multimedia and the Internet, and became integrated into the eLearning framework.

2.2 Current Landscape

Today, learning a language does not necessarily have to employ a formalized approach within a single service. It is possible to achieve basic competency in a foreign language by using one or more different specialized services, many of which exist in purely on-line form.

Examples of such platforms include on-line vocabulary games and short courses, Web-based language exchanges, and social networks. With the social network model being well-established and accepted, a new category of social bonds was established, helping foster the creation of on-line eLearning communities.

There are also new tools available which utilize blended learning approaches [3], significantly simplifying the teaching and learning activities. Much research in these areas is supported by programs sponsored by the European Union[4,5]. The success of a modern eLearning application depends on the ability to combine several seemingly unrelated concepts to create a cohesive and useful platform. Many recent sites combine aspects of a social network with personalized learning, provide a mix of formal and informal learning approaches, operate using both proprietary and open content, and use emerging (usually mobile) hardware platforms.

3 Overview of Lingro

We propose Lingro as a case study for this new approach to eLearning. Lingro is a free-to-use multilingual dictionary and platform for studying vocabulary which proved to be a powerful on-line Web 2.0 tool for learning languages[11]. Lingro's mission is to create an open on-line eLearning environment for language students worldwide.

There are two main reasons for using Lingro as a starting point for our analysis of issues affecting next-generation eLearning applications:

- We have extensive first-hand experience with development, business and marketing issues affecting Lingro, and have detailed data allowing us to analyze usage patterns.
- Lingro is in many ways a typical Web 2.0 educational application; the context in which it is developed, processes and target users closely reflect the circumstances of many other recent eLearning projects.

3.1 Organization and Capabilities

Lingro was designed to serve as a multilingual dictionary combined with a set of language learning tools to combine reference and learning components. The main goal of Lingro is to allow language learners to read websites and documents in the language they are studying and provide them with tools to organize and study relevant vocabulary. In contrast to machine translator software, such as Google Translate and Yahoo! Babel Fish, Lingro does not automatically translate texts, but displays them in the original language, translating only the foreign-language words and phrases requested by the user. Thus, the core function of Lingro is to facilitate learning by providing tools to reference and organize foreign-language vocabulary. Lingro currently maintains 121 dictionaries for 11 languages, with over 8 million translations and definitions.

The website is organized into four primary components: reference, review, games, and external tools.

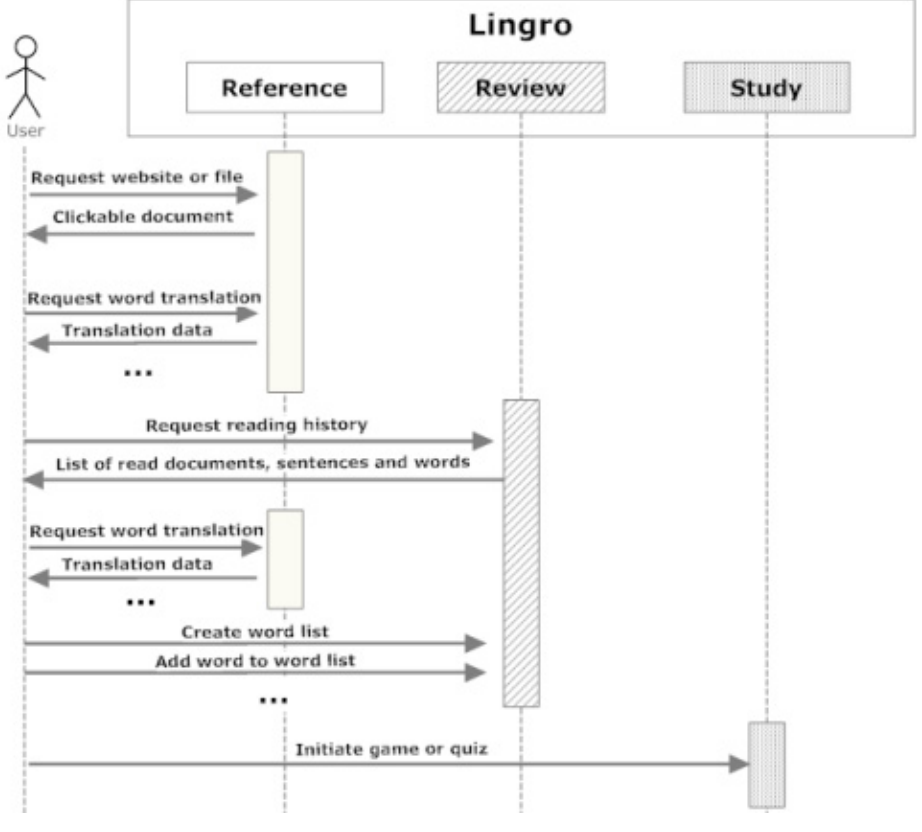


Fig. 1. Typical Lingro user workflow

- Reference component
  - *Web/File Viewer* systems, allowing language students to read any website or file in a foreign language, providing the functionality to check the meaning of any word with one click. Intuitive pop-ups suggest a meaning or translation. Each word is automatically added to the student's Word History.
  - *Dictionary* interface for checking word translations in all available dictionaries. The system searches for matching vocabulary as soon as a user starts typing into the input form and shows several relevant words.
  - *Dictionary builder* allowing users to contribute to the project by adding and improving translations and definitions.
- Review component
  - *Word list* tool for composing word lists which may then be used in vocabulary training activities.
  - *Sentence history*, containing contexts of words translated by each user, allowing users to review words in the sentence in which they originally encountered the unknown word.
- Study component, including flashcard exercises and games (in development).
- External tools
  - *Webmaster tools* to allow third-party sites to access information in Lingro's dictionary database.
  - *Browser plug-ins* to give users the ability to reference translations in a variety of ways.

A typical end-user workflow, initiated after visiting the `lingro.com` website, is shown in Figure 1.

4 Analysis

We now describe the main issues affecting eLearning projects in the Web 2.0 space and compare approach taken by such projects to traditional ones. We focus on the areas where the two methodologies show the most pronounced differences, as discussed in [1], as well as those from our first-hand experience. A summary of important differences is shown in Table 1.

4.1 Platform

The advent of Web 2.0 technologies caused a massive shift to using the World Wide Web as an eLearning platform [17]. The decision of many software authors to launch their products as Web applications available through a Web browser is itself responsible for many of the other factors differentiating next-generation applications from traditional educational software.

In such traditional software, a clear divide existed between off-line and on-line solutions. Products distributed as CD-ROMs rarely included any network capabilities; Web-based forums, on the other hand, were mechanisms for facilitating human interaction, without providing real learning solutions.

Similarly, products for mobile platforms (PDAs) focused on providing reference capabilities, including dictionaries, phrasebooks and machine translations, and lacked a strong learning component.

New technologies and associated processes are attempting to resolve many of the shortcomings of older approaches which affect the distribution of software to end-users, as well as usability and overall user experience. The main distinguishing features of such new projects are:

- Universal availability – it is immediately possible to use Web applications without installing any software.
  In the first month after launch, Lingro was tried out by over 85,000 people, of which 34,000 used it in the first week; initial users represented 163 countries and utilized all major operating systems (Windows, Mac and Linux).
- Continuous improvement – updates to Web applications can be made at any time, and immediately propagate to all users.
  Due to the centralized nature of Web applications, adding new functionality to existing websites is much easier than in the case of traditional software, as such changes don't require any user interaction. In the case of language learning projects, such improvements often include better localization, new content and features. Similarly, when *software bugs* are discovered, they can be handled according to their severity; particularly important issues in Lingro have been fixed within 15 minutes to 2 hours after being reported. However, the ability to quickly perform such upgrades translates in the need for constant maintenance.
- Hardware platform convergence – providing software for different platforms to interface with main Web-based system.
  An increasing trend among recent projects is to utilize the growing need for mobile solutions and provide software for various kinds of portable devices, most often cellular phones. This approach makes it possible to create a platform which can provide a basic service to the user without being tied to particular hardware. In the case of Lingro, in addition to the main Web-based system and browser plug-ins and bookmarklets, applications for the iPhone platform are being developed; each component connects to Lingro's Web system, receives updates and stores user data for future reference.

Using the Web as an application platform also affects several other business processes, including revenue generation and marketing; we discuss those issues later in this work.

4.2 Educational Content

The traditional approach to generating content for eLearning uses was, for many organizations, to build databases of proprietary content and utilize it in many products, or to license necessary information from third-parties.

In many cases, Web 2.0 startups use a similar approach by narrowing their focus (for example by only providing content for a few popular languages) and by licensing data (such as dictionary information) from other publishers.

Table 1. Comparison of Traditional and Web 2.0 Approaches

Category	Traditional approach	Web 2.0 approach
Platform	PC/CD-ROM	Web, mobile
Content	proprietary/licensed	open/user-generated
Learning style	formal/course-based	individual/informal
Interaction model	application → user	user ↔ on-line community
Distribution	store sales	on-line
Marketing	advertising	on-line advertising/word-of-mouth
Revenue	sales	on-site advertising

However, in the last decade, the idea of *user-generated content* was introduced with significant success. Such an approach relies on enabling users to create and enter data into a Web-based system; the most successful example is Wikipedia, a user-generated encyclopedia. Many eLearning platforms combine proprietary content with user additions.

The approach taken by Lingro was to utilize only *open content* [8] data, as a cost-effective solution for obtaining dictionary information. The data was gathered from two kinds of sources:

- open dictionary projects – dictionaries released under free licenses, such as Creative Commons and GNU Free Documentation License[1].
- additions and corrections made by Lingro users

To increase the rate of creating dictionary content, Lingro implemented a system allowing users to actively participate in the development, addition and modification of our dictionaries. Overall, we found that approach to be extremely beneficial for user engagement and publicity reasons – collecting and distributing open content data fosters the creation of a user community and encourages users to disseminate information about the service.

However, we also note that from a practical standpoint, enabling users to create content can sometimes negatively affect the quality of provided data. We found that potential problems with low-quality submissions can be solved by implementing a relatively simple moderation system.

4.3 Learning Style

Because of the growing number of Web-based language learning applications, it is possible to find software employing both formal and informal learning approaches. However, a general trend observed in Web 2.0 products is to enable informal learning styles; a possible reason for this is the desire to supplant existing software with capabilities provided by the Web platform, rather than replace it completely.

[1] Due to the large number of available *open* and *free* licenses such terms have no single accepted definition. For our purposes, an open license is one which explicitly gives rights to its user, rather than restricting them for the benefit of the copyright owner.

Some of the main advantages of a Web-based solution include the already discussed immediate availability of the application to any Internet user, and the ability to utilize network effects for increasing the number of users of the application. By giving users the freedom to decide which content to use for learning, it is possible to help them achieve their individual educational objectives. Allowing users to create their own content and publish it on the site fosters collaboration and benefits the application by providing other users with more educational choices, without direct involvement from the application's creators.

Many Web 2.0 eLearning startups are thus focused on building platforms that help users learn, rather than just provide them with educational content [9]. Applications such as social networks and on-line language exchange programs to connect learners with native speakers of a language have recently become extremely popular; another development is media-based learning, where users post their own video clips with particular learning objectives – all are examples of informal learning styles. The facilitation of interaction between (possibly geographically dispersed) users has had such broad implications, that many consider Web 2.0 to be more of a social revolution than a technological advancement [17].

Lingro does not aim to provide a formalized language learning path – the choice of educational objectives is left to the user. By allowing each learner to make decisions about the kind of texts to read and analyze, Lingro becomes an eLearning platform which can fulfill various learning needs.

The key features which Lingro adds to informal learning are a result from the following:

- Users can learn in an unstructured way, enabling informal learning,
- Users can learn while performing work-related tasks.

Such an approach allows Lingro, and other similar applications, to be used alongside more structured courses, to introduce informal learning into an otherwise inflexible learning environment.

4.4 Financials

Initial Development. The cost of developing a website capable of providing a service to a moderately-sized user community is extremely low – Lingro was created with little more than $10,000 in educational grants, in addition to software development costs. The maintenance of the infrastructure (including computing and network hardware) is delegated to an outside hosting provider, allowing the primary staff to focus on research, development and implementation of the application itself, rather than solving systems issues.

Marketing. With low development costs, many Web 2.0 startup organizations opt to spend a significant percentage of their budget on marketing and public relations to publicize their products. In the absence of existing customers and distribution channels, the need to reach the target customer group often requires innovative marketing strategies.

Well-funded startups often choose to combine purchasing on-line contextual advertisements with strategies employing word-of-mouth as a mechanism for spreading awareness about their products. The low cost of on-line ads (usually between $1-10 to display the ad to 1000 users) makes it possible to reach millions of potential customers with a limited budget; the advent of contextual advertising through services such as Google AdSense helps to target specific groups of users who might be interested in the offered product.

In the case of Lingro, because of the small operating budget, the sole strategy employed for reaching potential customers was word-of-mouth. Lingro initially contacted a small group of individuals interested in on-line language learning (including bloggers and operators of other eLearning websites) and asked if they were interested in reviewing the service. When several of them posted information about Lingro on their websites, new users tried the site and used their own blogs to let others know, in an example of *viral marketing*. An important factor in the initial marketing success of Lingro was the decision to use *open content* dictionaries, which drew much attention from people involved in the *Creative Commons* movement.

Relying on word-of-mouth, while potentially very effective, also requires the basic product to be available free of charge and with minimal impediment to use (for example, without requiring users to register on the website). We believe that many organizations will choose this mechanism in addition to other more active ways of reaching potential customers.

Revenue. Distributing an eLearning product as a Web application makes it difficult to directly charge customers for using the product. As most websites can be viewed free of charge, many potential users will only consider using content they do not have to pay for, even if it is of inferior quality to paid content. An anecdotal rule of thumb in the Web 2.0 industry is that of all users of a Web application, approximately 1% will create a *free account*, and that 1% of users with free accounts will be prepared to purchase *paid accounts*. In Lingro, out of 240,000 users visiting it in the first year, 2,133 created free accounts, confirming the first estimate; as Lingro does not offer premium paid accounts, we cannot verify the second estimate.

It is clear the traditional approach of requiring customers to purchase a product before using it does not suffice in the Web 2.0 space. A common solution is to support a free product by displaying on-line advertisements. Another approach is to suggest other products of interest to the site's users and collect referral fees.

5 Conclusion

In this work we provided an overview of select issues affecting eLearning projects in the Web 2.0 era. We described past and current approaches, and used the Lingro language learning platform to illustrate examples of current industry trends and practices. We also analyzed various methodologies, both those affecting business operations and developing software code as well as acquiring and creating educational content.

Lingro, like many other Web 2.0 applications, has been designed as a platform taking advantage of recent technological and social advances. Implemented as Web application, the Lingro system is able to provide a useful service to thousands of users with negligible distribution costs; the use of a high-level programming language enables developers to quickly add requested features for all users. By using open content dictionary data combined with user contributions, Lingro avoids the often prohibitive costs of generating or licensing educational content. The service provided by Lingro focuses on giving users the ability to follow an individual learning path; community-based aspects are currently developed to allow collaborative learning.

The approaches taken in the development of Lingro, while indicative of larger Web 2.0 trends, are, in some areas, a significant departure from traditional methodologies. Through pinpointing the differences between older approaches and current trends we hope to contribute to the understanding of many issues affecting next-generation learning platforms, and allow for the inclusion of Web 2.0 concepts in existing eLearning applications.

Acknowledgements. Authors would like to thank Katarzyna Kościelska for her invaluable input at various stages of the development of this work. We would also like to express gratitude to the Adam Mickiewicz University in Poznań, Poland for providing funding which allowed us to present our observations.

References

1. O'Reilly, T.: What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software. O'Reilly Media, Sebastopol (2005), www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html
2. Warschauer, M.: Computer Assisted Language Learning: an Introduction. In: Fotos, S. (ed.) Multimedia language teaching, pp. 3–20. Logos International, Tokyo (1996), http://www.ict4lt.org/en/warschauer.htm
3. Liotsios, K., Demetriadis, S., Pombortsis, A.: Blended Learning Technologies in Lifelong Education: Lessons Learned from a Case Study. In: Nejdl, W., Tochtermann, K. (eds.) EC-TEL 2006. LNCS, vol. 4227, pp. 634–639. Springer, Heidelberg (2006)
4. Autonomous Language Learning, http://www.allproject.info/
5. European Association for Computer-Assisted Language Learning, http://www.eurocall-languages.org/
6. Mantzari, D., Economides, A.A.: Cost Analysis for E-learning Foreign Languages. European Journal of Open and Distance Learning, http://conta.uom.gr/conta/publications/PDF/Costanalysisfore-learningforeignlanguages.pdf
7. May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148–1158. Springer, Heidelberg (2006)
8. Pawlowski, J.M., Zimmermann, V.: Open Content: A Concept for the Future ofE-learning and Knowledge Management? In: Knowtech, Frankfurt (November 2007), http://users.jyu.fi/~japawlow/knowtech_20070907finalwithcitation.pdf

9. Safran, C., Helic, D., Gütl, C.: E-Learning Practices and Web 2.0. In: International Conference on Interactive Computer Aided Learning (2007)
10. Lingro: Foreign-word widget, http://features.csmonitor.com/innovation/2008/07/16/lingro-foreign-word-widget/
11. Lingro helps you translate ginormous words, http://news.cnet.com/8301-17939_109-9966624-2.html
12. Wikibooks, http://wikibooks.org
13. Wiktionary, http://wiktionary.org
14. Dict, http//dict.org
15. The DICT Protocol, Request for Comments: 2229, http://www.faqs.org/rfcs/rfc2229.html
16. Google Translator, http://translate.google.com/
17. Downes, S.: E-learning 2.0, http://www.elearnmag.org/subpage.cfm?section=articles&article=29-1

ESHE 2009 Workshop Chairs' Message

Jorge Marx Gómez, Liane Haak, and Dirk Peters

Carl von Ossietzky Universität Oldenburg, Fakultät II - Department für Informatik,
Abteilung Wirtschaftsinformatik I / VLBA, 26111 Oldenburg, Germany
{jorge.marx.gomez,liane.haak,dirk.peters}@uni-oldenburg.de

1 Workshop

The 1st Workshop on Enterprise Systems in Higher Education (ESHE09) in conjunction with the 12th International Conference on Business Information Systems (BIS 2009) in Poznan, Poland is dealing with Enterprise Systems in form of Business Information Systems which are, as part of any business environment, becoming an integrated part of educational environments. Because of their potential of illustration, visualization and simulation of business and decision-making processes to students, the systems have potential for future pedagogic innovation within higher education. The main goal of this workshop is to bring researchers and practitioners together to explore the issues and challenges related to Enterprise Systems in Education and to show the actual activities in this area.

2 Introduction of the Papers

Magnusson, J.; Oskarsson, B.; Gidlund, A.; Wetterberg, A.: Process methodology in ERP-related education: A case from Swedish higher education
On the basis of a under-developed use of Enterprise Systems in the curriculum of a swedish university, the authors are presenting a case, which explores the potential use of a flow from business opportunities to general ledger as a means for achieving both ERP and business knowledge enhancement among the students. The approach – as a joint initiative between academia and industry – was applied and tested in Sweden during the fall of 2008.

Brehm, N.; Haak, L.; Peters, D.: Using FERP Systems to introduce Web Service-based ERP Systems in Higher Education
Introduced by the fact that Enterprise Systems like Enterprise Resource Planning systems are an important application area in higher education, the authors are facing today's problems occurring from difficulties in teaching. The contribution reports about using a Federated ERP system as an example for a Web Service-based ERP system to introduce aspects of Service Oriented Architectures in higher education.

Ask, U.; Magnusson, J.; Enquist, H.; Juell-Skielse, G.: Applied Business Intelligence in the making: An interuniversity case from Swedish higher education
The paper is about a joint initiative between academia and industry, which gives students the possibility to access real data during the curriculum from a medium-sized manufacturing company via full accounts. The students are given the task to identify

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 212–213, 2009.

potential problems with these accounts by using a specially designed BI solution. The purpose of this case is to present the outline and outset for the competition, together with some initial reflections on the setup-phase.

Hans, D.; Marx Gómez, J.; Peters, D.; Solsbach, A.: Case Study-Design for Higher Education - A Demonstration in the Data Warehouse Environment
The research presented in the paper focused on case study-design by analyzing existing methods and the developing of criteria's in order to give an insight how case studies can be designed to support didactics and education science aspects. The analysis of the scientific work in this field shows a gap between actual available case studies and approaches to increase the learning success in the field of education science. As exemplary the paper presents a case study in the data warehouse environment.

Schilbach, H.; Schönbrunn, K.; Strahringer, S.: Off-the-Shelf Applications in Higher Education: A Survey on Systems deployed in Germany
The goal of the contribution is to explore, if there is a steady growth for off-the-shelf applications in higher education as well as it already appears in many industries. The paper shows a survey among institutions of higher education in Germany. As a result there are a few dominant products that already automate the highly administrative process areas, but there are still a few process areas with a low degree of automation.

Process Methodology in ERP-Related Education: A Case from Swedish Higher Education

Johan Magnusson[1], Bo Oskarsson[2], Anders Gidlund[2], and Andrea Wetterberg[3]

[1] Centre for Business Solutions, School of Business, Economics and Law, University of Gothenburg, Sweden
johan.magnusson@handels.gu.se
[2] SYSteam AB
bo.oskarsson@systeam.se, anders.gidlund@systeam.se
[3] Jeeves Information Systems AB
andrea.wetterberg@jeeves.se

Abstract. There has long been a debate regarding the inclusion of IT into the curriculum for business students. With IT being a natural part of their coming working environment, the under-developed use of for instance Enterprise Resource Planning (ERP) solutions has suffered much critique. At the same time the process oriented approach has saturated the business environment. With ERP systems designed using a process oriented approach, this case explores the potential use of a flow from business opportunity to general ledger as a means for achieving both ERP and business knowledge enhancement among the students. As a joint initiative between academia and industry, a first attempt at using this approach was applied and tested in Sweden during the fall of 2008. The purpose of this case is to present the outline of the initiative, along with an evaluation and key lessons-learned.

Keywords: ERP education, Role-based, Process oriented, business education.

1 Introduction

In the spring of 2005, the industrial advisory council for the School of Business, Economics and Law at the University of Gothenburg in Sweden expressed a need for making IT an integrated part of the business education. With the purpose of decreasing the time-to-productivity for new-hires, the Centre for Business Solutions (CBS) was established as one of the schools six strategic initiatives.

Since 2005, more than 3000 students with practical ERP experience have left the school for promising careers within the industry. The work within CBS has focused on integrating the use of ERP and Business Intelligence (BI) solutions as a natural element in all courses at the school, as a pedagogical tool for closing the gap between academia and industry.

From the start, the focus of the work has been on simulating and visualizing business processes where the students can apply their attained theoretical knowledge. Hence, the process oriented approach has been of paramount importance for the success of the centre.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 214–219, 2009.

In October, 2008, a dialogue was initiatied between one of the members of CBS Industry Advisory Council and a co-worker at the centre. The outset was to create a series of sessions for the course "Applied Enterprise Systems" where master-students would follow a key business process from start to finish.

Early on the process from business opportunity to general ledger was identified as a starting point for increasing the students awareness of what business is and what the links to the ERP system are. After establishing support from an ERP vendor (Jeeves) and a consulting firm (SYSteam) for the practical aspects of running the sessions, an implementation and evaluation was conducted in December 2008.

2 Outline

After initial discussions, a general outline of the course element was agreed upon. This involved an introduction to process orientation and the links to ERP systems followed by an introduction to the assignment and outline of the sessions.

After this, the students got a crash-course in the ERP system with a clear focus on the business process at hand. The platform for this was a Microsoft Sharepoint solution where the consulting firm had built a complete process map with links to all the necessary user instructions for each process. At the same time, the students had unlimited access to a full installation of the ERP system with 30 user accounts (one per group). The students were organized into groups of four, and the assignments given focused on taking a business opportunity all the way through the process to general ledger (see below).

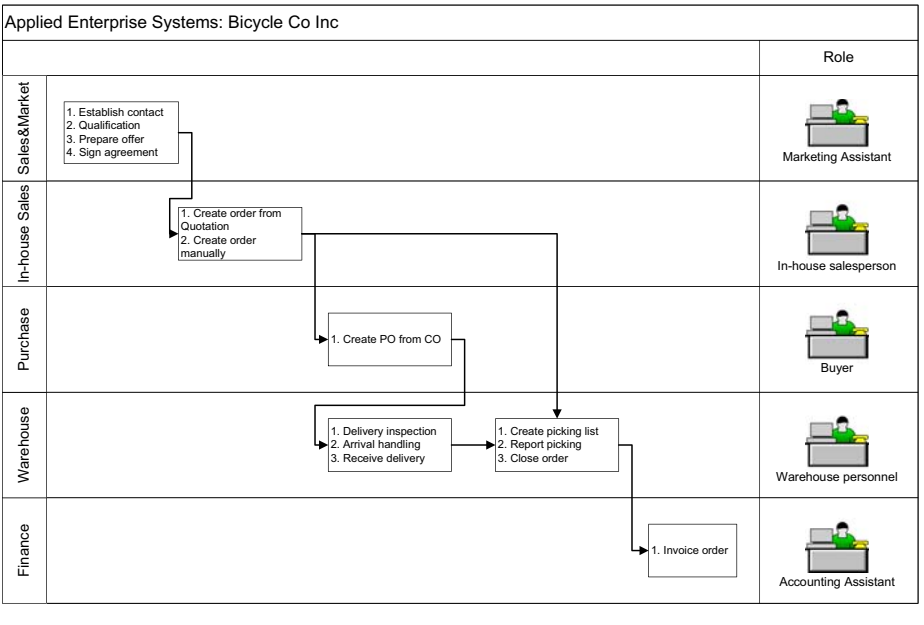


Fig. 1. The process in focus and its related functions and roles

This process was chosen as a means for illustrating the different business functions and roles involved in taking generating value for a company. As can be seen in the figure above, this involved the functions of Sales&Market, In-house sales, Purchasing, Warehouse and Finance and the roles of Marketing Assistant, In-house sales person, Buyer, Warehouse personnel and Accounting Assistant.

Apart from access to the ERP system, the students were also give access to a Sharepoint portal that had been developed by the consultancy firm and that was now part of the offer from the ERP systems vendor. This portal was organized as an interactive process map where the students could go from main process all the way down to the specific tasks that were involved in each sub-process. The students could also, from the process map, access user guides and tutorials specifically designed for the ERP system in focus.

The students were given a week (in parallel with the rest of the course) for completion of the flow from Business Opportunity to General Ledger, with one session in an "open-house" setting where three consultants were present for an afternoon in a computer-lab for those groups that needed extra assistance in completing the assignment.

After completion of the entire flow, a new session was held where the students were introduced to the difficulties in handing transactions between different functions and roles. Each of the groups were assigned a role such as Marketing Assistant, Warehouse personnel et cetera, and given a new log-in to the ERP system. This time, the system was scaled down and customized for the particular role, and the students were given an assignment in getting to know the role and which processes and functionality that was of relevance. Like in the first round, the students were given one session in an "open-house" setting with three consultants.

After completing the second assignment, the students were brought together into two large groups. Each of the larger groups was comprised of two groups per role, hence ten of the original groups (40 students).

Each of the larger groups were then placed in a separate room together with consultants and lecturers from the academy to enact a live Business Opportunity to General Ledger case. The consultants called representatives for the first role (Marketing assistant) to the floor, and gave the group a report of a meeting with a potential customer at a trade-fair.

While going through the input of all necessary data into the system (projected on a beamer for the rest of the participants in the large group), the group was asked to describe what they were doing and what functionality they were using. They were also asked to reflect on the implications of their handling of the system, with consequences of their choices constantly being questioned by the consultants and academics on site.

After completion of the process handled by the role of Marketing assistant, the next role in the complete process was invited on stage to go through the same procedure.

To summarize, the overall outline of the course element was:

1. Opening day
   - Introduction to process oriented business and Enterprise Systems(2h)
   - Introduction to process oriented Enterprise Systems (2h)
   - Introduction to system and exercise (1+3h)

2. Tutoring (3h)
3. Specialization
   – Introduction to Roles and exercise (1+3h)
4. Tutoring (3h)
5. Live case and roleplay
   – Introduction and complete inter-group process (4h)

All in all, the time spent in class was 22 hours, distributed over a two week time-period.

3 Evaluation

Apart from the in-class, qualitative evaluation that was conducted in relation to the three main sessions (1, 3 and 5 in the summary above), two quantitative evaluations were conducted after the first and last sessions. This evaluation was based on four criteria, namely relevance, quality, complexity and experience which were evaluated on a five grade Likert scale. In the two questionnaires, the response rate was (41%) and (38%).

On a general level, the students were positive towards the course element, mainly given the high level of perceived relevance of the exercise. The quantitative post-course evaluation gave the following scores (out of 5): Relevance (4,4), Quality (3,0), Complexity (2,2), Experience (3,3).

Concerning a question given to the students regarding if this exercise was a necessary part of an education within business administration, 100% of the students answered that this was the case in their on-line post-exercise evaluation.

The process and methodology applied within this course has since this application been diffused and adopted by other Swedish schools such as Umeå University (using Microsoft Nav) and the Royal Institute of Technology (using Microsoft Ax).

4 Lessons Learned

Given the short time-span from conception of the idea to the factual implementation, the course element was regarded as highly successful. The student's perception of this as being a highly relevant and necessary part of a business education is one of the indicators that were regarded as most promising for the continued development and inclusion of this in future courses.

Apart from establishing a proof-of-concept for the idea of process oriented ERP education, this first try also resulted in a number of lessons-learned that could help enhance the quality in this type of education. Many of these lessons-learned are considered general for most ERP related education.

Among the lessons learned, the following were seen as most useful for future development of the course element:

– Secure a 100% service level when it comes to access to the ERP system.

Given that the overall IT environment at the school was under re-design during the course, there were a lot of problems related to conflicts between different profiles,

security protocols and VPN-connections that were less than reliable. This was the main critique offered and something that for one part of the student body also resulted in them not seeing what the exercises really were intended for. The students had limited understanding of what was currently going on in the technical environment, and for some this resulted in them taking a back-seat approach to the course element.

– Package the overall methodology and assignments so that the students are given all the documentation before starting the course element.

With the short time-span between conception and implementation, the course element was in some aspects rather improvisational from the course leader's side. This resulted in a blurring of objectives and a difficulty in communicating the overall goals and pedagogic targets. This could be further improved by developing full documentation for the exercises, and an overall description of objectives that should be communicated to the students at the start of the course.

– Make the course element part of the regular examination for the course and make sure that the students have ample time for completing the exercises.

In the course, there was a strenuous work-load for the students. With the focus being on application of theory into the role of consultant, the course differs from what the students are used to from previously in their academic education. The late inclusion of the course element into the course resulted in the course administration not being able to assign any credits to the course element. This in turn resulted in a cope-out from several of the students that were highly motivated to gain a high grade for the course. Since their performance in the course element did not impact their overall grade for the course, they felt that they could not prioritize the exercises. This could be further improved by replacing one of the other cases in the course and thereby making it part of the grading.

– Prepare a number of preliminary, theoretically related questions for the students as they enter Phase II of the course element.

In the second phase of the course element, the students were asked to act within a particular role and reflect on the process that this role was involved in and the functionality needed. They were also asked, in the final session, to reflect on the theoretical aspects of what they were doing, such as for instance discussing the relationship between how the role of Byer impacted stock-turnover et cetera. This linking of the different key performance indicators to the choices made by different roles could be made more interesting if the students prepared for a discussion. Given the limited information that the students received before the last session, this discussion was somewhat limited, even though it was moderated by consultants and academics.

– Involve consultants with both ERP systems knowledge as well as business knowledge in the course element.

The consultants and vendor assigned a lot of time and resources to the course element, and this is something that was perceived as highly positive by the students. The message sent from the consultants was that this was a good investment, given that they would later meet the students both as potential employees and as future buyers of their products and services. This gave the students the perception that what they were currently studying was highly relevant and improved their employability. The

consultants match between systems and business knowledge was also highly valuable, since they could highlight certain aspects that were out of reach for the academic lecturers involved.

– Keep the focus on processes and roles.

The use of specific roles assigned to different functions in the company in question was perceived as highly valuable. With a substantial data-set available form a real-life company, the richness of the data was sufficient for making the task of rummaging around in the particular role interesting to the students. This was seen as a key ingredient in making the students increase their knowledge regarding what constitutes a business and its inner workings.

5 Contact and Further Information

University of Gothenburg: www.gu.se
School of Business, Economics and Law: www.handels.gu.se
Centre for Business Solutions: www.handels.gu.se/cfa
Jeeves: www.jeeves.se
SYSteam: www.systeam.se

Using FERP Systems to Introduce Web Service-Based ERP Systems in Higher Education

Nico Brehm, Liane Haak, and Dirk Peters

Carl von Ossietzky Universität Oldenburg, Fakultät II - Department für Informatik,
Abteilung Wirtschaftsinformatik I / VLBA, 26111 Oldenburg, Germany
{nico.brehm,liane.haak,dirk.peters}@uni-oldenburg.de

Abstract. Enterprise Systems like Enterprise Resource Planning (ERP) systems are an important application area in higher education and getting more and more complex. Therefore problems occur from different perspectives and increase the difficulties in teaching. Existing software products cannot be used to teach specific technical aspects, like e.g. Web Services because it is still research in progress. This contribution reports about using a Federated ERP (FERP) system as an example for a Web Service-based ERP system to introduce aspects of Service Oriented Architectures (SOA) in higher education.

Keywords: ERP Systems, Federated ERP Systems, FERP, Web Services, SOA.

1 Introduction

Enterprise Systems in form of Business Information Systems (BIS) are getting more and more important for educational environments. The technology aspects are highly complex in many areas and the resulting problems are depending on different perspectives in computer science and business economics. This interference and complexity make the teaching and learning in this field quite difficult. For this reason systems practically approved are needed to give the students the chance to get their own practical experience. Enterprise Systems have the capability for future pedagogic innovation within higher education. Their potential results from the possibilities in illustration, visualization and simulation of business and decision-making processes to students. The main goal of using enterprise system as e.g. Enterprise Resource Planning (ERP) systems in higher education is to prepare them for real work life and to give them practical experience in the application of these technologies. Another objective is focused by software developing companies like SAP® or Microsoft®, e.g the students should get in touch with their special products as early as possible, so that they already know these products if they need to work with it later or are in the position to decide about investments. Our interests belong to the area of Business Information Systems, to the variety of ERP technologies and how they could be used to teach our university students. Hereby the focus lays besides the actual standard software on new concepts resulting from up to date research, e.g. Federated ERP (FERP) systems. This contribution will show the benefit from using the FERP system to introduce Web Service-based ERP systems in higher education with the help of an exercise used within the lecture *ERP Technologies.*

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 220–225, 2009.

2 Technical Background

Before introducing the concept of FERP systems and how they can be used for teaching in higher education it is necessary to outline the main technical concepts behind: ERP systems and Web Services.

2.1 ERP Systems

ERP systems can be defined as standard software systems that integrate operational application systems of different enterprise sectors. These sectors can be for example customer service, finance or sales, which act as different components within one ERP system (see Fig. 1). The main goal of ERP systems is to integrate the data, functions and processes of all components of an enterprise. Because of the high complexity of ERP systems the following problems occur [1]:

- The price-performance ratio is dependent to the potential benefit an ERP system is able to generate
- in the majority of cases, not all installed components/functions are needed
- a high-end computer hardware is required
- expensive customizing is necessary

Due to these problems normally only huge enterprises can apply such complex ERP system to provide business logic of all its sectors. However, small- and medium sized enterprises (SME) deploy several different smaller business application systems in parallel [2]. This parallel operation often causes problems which jointly arise from insufficient system integration. Moreover the potential of each business application system is not exploited [3].

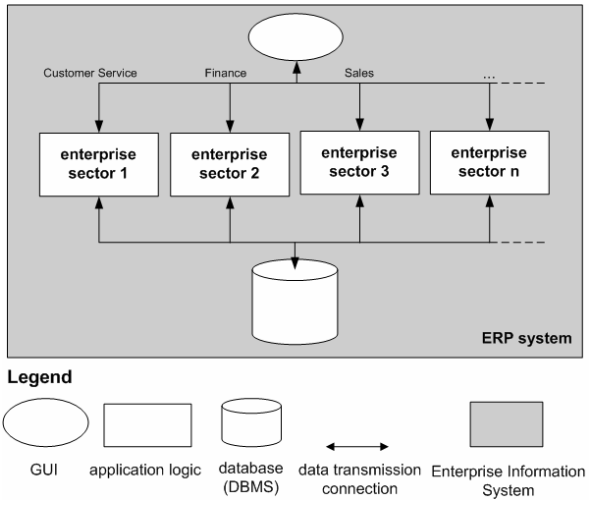


Fig. 1. Architecture of a conventional ERP system [1]

2.2 Web Services

Web Services are software systems which support interoperable interactions between machines over a network. Its interfaces are described in the Web Service Description Language (WSDL) to allow systems the interaction with the Web Service [4]. According to the paradigm of SOA, Web Services enable the development of distributed applications by offering single services which can be interconnected and reused. Furthermore, the integration of existing software is possible. Due to the allocation via a network and the access by open standards, another advantage of Web Services is the operating system and programming language independency [5].

3 FERP Systems

To face the problems which come along with setting up and operate a conventional ERP system, there is a need of a system in which SME can fulfill its requirements. One possible solution is the usage of FERP systems. A FERP system allows a variable assignment of business application functions to software providers. The overall functionality is provided by an ensemble of standardized subsystems that all together appear as a single ERP system to the user. Different business components can be developed by different vendors [1].

In the approach the application logic of ERP systems is encapsulated in a multiplicity of Web Services, which allows the separation of local and remote functions whereby no local resources are wasted for unnecessary components. Furthermore, in FERP, single components are executable on small computers which subsides the installation and maintenance costs by decreasing the degree of local system complexity [6]. According to the usage of Web Services and following the multi-layer paradigm of modern information systems by aiming at the separation of the application logic from the presentation layer and the database layer, the vision of this approach is to allow the application of business logic components in a distributed manner.

As shown in Figure 2, the FERP reference architecture consists of several subsystems which are interconnected. The different subsystems which are specific for the Web Service functionality are described in the following section, for more details of the whole concept please see [1].

- *FERP Workflow System (FWfS)*
  The FWfS represents the central component within the FERP reference architecture and coordinates all business processes inside the FERP system. The main task of the FWfS is to organize process definitions and to control business processes. The business processes are described in an appropriate XML-based workflow language and can be implemented via a FERP workflow editor [1].
- *FERP Web Service Consumer System (FWCS)*
  The FWCS is the subsystem which provides methods to call Web Services in order to execute the business logic via the business processes. All possible types of FERP Web Services are specified by the FERP WS standard which describes the Web Service operations as well as the input and output parameters [1].

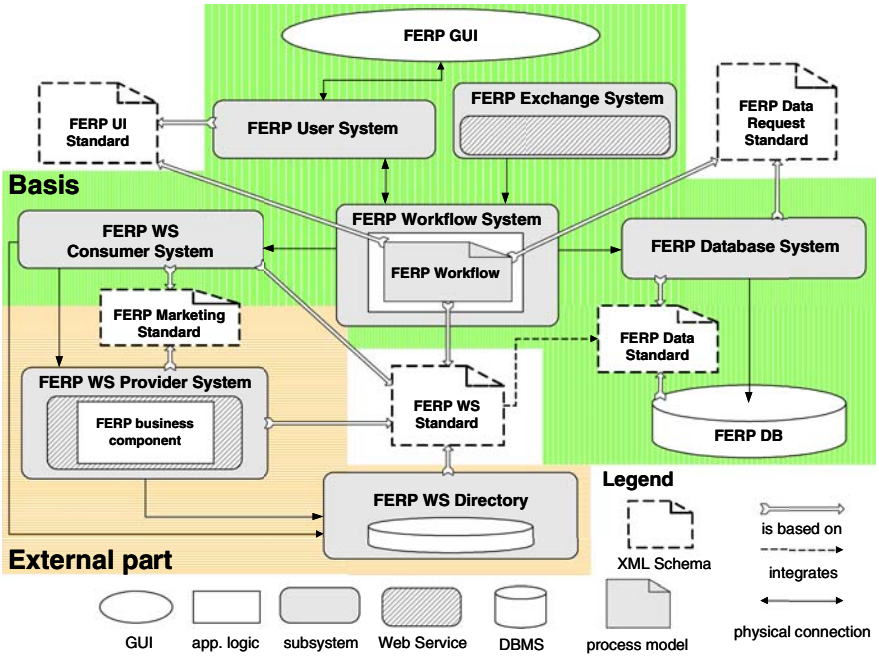


Fig. 2. Reference architecture of an FERP system with relation to necessary standards [1]

– *FERP Web Service Provider System (FWPS)*
  As another subsystem of the FERP reference architecture, the FWPS implements the functionality of Web Services according to the FERP WS standard. It offers the possibility to publish Web Services and handles incoming Web Services requests.
– *FERP WS Directory (FWD)*
  The FWD saves the Web Service descriptions in a format of a so called Web Service Secure Marketing Language (WSSML) and includes a link to the technical description of the Web Service (WSDL-File) as well as marketing-specific descriptions for helping users to make a decision about using the Web Service or not, e.g. price.

4 FERP Systems in Higher Education

FERP is a theoretical approach which shows how problems of conventional ERP systems could be solved by using new technologies like Web Services. Furthermore there is the prototype *FERP X ONE* as open source software available as basis for a Web Service-based ERP System. Thus, we used it in our lectures to teach the students in aspects of these new technologies, e.g. ERP Technologies and Electronic Business.

The objective of the lecture ERP Technologies is to teach the main theoretical and technical aspects in this application area, e.g. the implementation of Web Services. Therefore we invented exercises which deepen and clarify the problems step by step to increase the understanding and the interaction with these technologies. After the

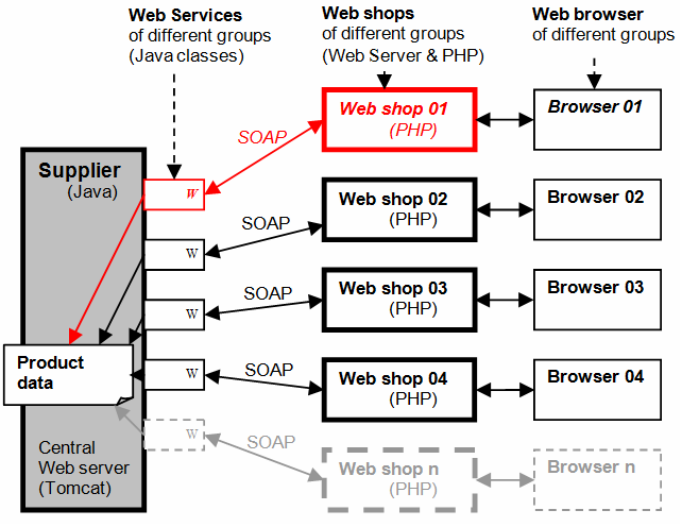


Fig. 3. Scenario of a Web Service exercise

introduction of the theoretical aspects of SOA and a detailed overview about the FERP Architecture, we give the students a practical exercise to learn more about the implementation. One scenario given in these exercises is shown in Figure 3.

The main task of the students in this exercise was to develop a Web Service for request and display the time of delivery of products. Therefore each student was already an owner of a web shop (implemented with PHP before) and partly prepared Java components were given as basis. The goal was to extend the functionality of the both systems in that way, that it allows a communication between the supplier (left side) and the web shop (middle) using the Simple Object Access Protocol (SOAP). All Web Services are managed by one central server and each student has to develop his program logic locally first and transfer it later to this server. In the next step they to extend the operation from getDeliveryTime to the parameter distance and weight to show as a result how the proposed delivery time is depending on the distance of the shop from the customer and from the weight of the product.

At the end of the exercise the students were able to implement a Web Service communicating with a web shop and an ERP system (FERP) via Web Service and to change different parameters. They used the Eclipse Platform[1], Tomcat Web Server[2], JAVA and PHP and learned why WSDL is needed.

5 Summary

FERP represents an approach for a Web Service-based ERP system which does not exist in practice, but where actual research in computer science and economics is

[1] www.eclipse.org
[2] www.tomcat.apache.org

directed to. Due to this fact, FERP offers a very good chance to let students participate in ongoing research activities by teaching them in concepts beyond the state of the art. Our contribution shows only one small example how this approach could be used for explanation and training in the area of Web Services-based systems. The main advantage of this approach is that students can work with an environment which considers newest technologies and is already approved. Thus, it is possible to teach them about the whole architecture as well as only in some specific parts like the implementing a Web Service as it is shown in our case.

The application area of this Enterprise System and its potential is much wider, e.g. in the area of Electronic Business etc. Actually we are working in an extension to include Semantic Web Services. The FERP prototype is open source software and consequently offers the potential for future scenarios in other lectures. Therefore the inclusion of FERP in higher education could be seen as collaborative work which is done in the future by more then only one university. In future work we are interested to build up a network of universities and research institutes to share our experience, software and cases. Within this, it is possible to extend and improve the technical environment as well as to benefit from each other in teaching actual cases.

References

1. Brehm, N., Marx Gómez, J.: Web Service-based specification and implementation of functional components in Federated ERP-Systems. In: Abramowicz, W. (ed.) BIS 2007. LNCS, vol. 4439, pp. 133–146. Springer, Heidelberg (2007)
2. Brehm, N., Heyer, N., Marx Gómez, J., Richter, B.: Das ERP-KMU-Dilemma und Anforderungen an Service-orientierte Architekturen zur Nutzung von Verbesserungspotentialen. In: Tagungsband der Multikonferenz Wirtschaftsinformatik 2008. München/Garching, Germany (2008)
3. Brehm, N., Marx Gómez, J.: Federated ERP-Systems on the basis of Web Services and P2P networks. International Journal of Information Technology and Management (IJITM) (2007)
4. Haas, H.: Web services glossary. Technical Report (2004), `http://www.w3.org/TR/2004/NOTE-ws-gloss-20040211/#webservice` (visited February 27, 2009)
5. Fensel, D., Lausen, H., Polleres, A., de Bruijn, J., Stollberg, M., Roman, D., Domingue, J.: Enabling Semantic Web Services: The Web Service Modeling Ontology. Springer, Heidelberg (2006)
6. Brehm, N., Lübke, D., Marx Gómez, J.: Federated Enterprise Resource Planning (FERP) Systems. In: Handbook of Enterprise Systems Architecture in Practice, London, UK (2007)

Applied Business Intelligence in the Making: An Inter-University Case from Swedish Higher Education

Urban Ask[1], Johan Magnusson[1], Håkan Enquist[1], and Gustaf Juell-Skielse[2]

[1] Centre for Business Solutions, School of Business, Economics and Law, University of Gothenburg, Sweden
urban.ask@handels.gu.se, johan.magnusson@handels.gu.se, hakan.enquist@handels.gu.se
[2] Royal Institute of Technology, Stockholm, Sweden
gjs@kth.se

Abstract. There has long been a debate regarding the inclusion of IT into the curriculum for business students. With IT being a natural part of their coming working environment, the under-developed use of for instance Enterprise Resource Planning (ERP) and Business Intelligence (BI) solutions has suffered much critique. As a response to this, the Centre for Business Solutions and the Scandinavian Academic Network for Teaching Enterprise Systems (SANTE) have created a joint initiative together with the industry. Through making the full accounts from a medium-sized manufacturing company available to the students through a specially designed BI solution, the students are given the task to identify potential problems with the accounts. The assignment is intended to be run in the form of a competition, where the students from different Swedish universities compete in analyzing the company in a given time-frame. The purpose of this case is to present the outline and outset for the competition, together with some initial reflections on the setup-phase.

Keywords: BI, Business intelligence, education, game, competition.

1 Introduction

In the spring of 2005, the industrial advisory council for the School of Business, Economics and Law at the University of Gothenburg in Sweden expressed a need for making IT an integrated part of the business education. With the purpose of decreasing the time-to-productivity for new-hires, the Centre for Business Solutions (CBS) was established as one of the schools six strategic initiatives.

Since 2005, more than 3000 students with practical ERP experience have left the school for promising careers within the industry. The work within CBS has focused on integrating the use of ERP and Business Intelligence (BI) solutions as a natural element in all courses at the school as a pedagogical tool for closing the gap between academia and industry.

CBS has since its formation built a pedagogic platform that consists of a selection of ERP and BI solutions that are made available to the students and faculty. Apart from access to the systems via Browser, terminal server and Virtual PC (VPC), the platform also contains documentation and educational material from the different

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 226–230, 2009.

vendors (Agresso, Microsoft, Jeeves, SAP and Hogia) as well as various material developed by the academy itself. This platform was awarded the Microsoft Customer Excellence Award EMEA in 2007.

SANTE (Scandinavian Academic Network for Teaching Enterprise systems) was formed in 2001 with the intent on introducing and including enterprise systems (ERP, BI etcetera) into the curriculum of higher education. The network consists of representatives from all universities in Sweden.

In 2006, SANTE requested that the infrastructure and platform developed by CBS be made available to the other universities in SANTE. In 2007, a model for inter-university collaboration surrounding the platform was presented and the first university, Umeå University was signed. Since then, a total of nine universities have joined the collaboration, sharing material, costs and labor to keep the infrastructure up-and-running.

In the fall of 2008, the first discussions regarding a practical business intelligence session were initiated. These discussions focused on the necessity for students of business to be given the possibility to work hands-on with real-life companies. With the technical possibilities inherent in the inclusion of enterprise systems into the curriculum, the idea to have a competition between different student groups where they competed in finding irregularities and potential problems for a company was formed.

After this, discussions took place between one of the consulting firms that support CBS (SYSteam) and representatives from the academy. The topic for the discussion was whether or not it would be possible to gain access to a real-life case, where the students would be able to work with a real, yet anonymous, company's complete set of transactions.

The consulting firm had previously developed a data-set where they had taken all the transactions from a medium-sized manufacturing company and made these anonymous. They had also created a script that could change the dates for these transactions, so that the transactions would always remain current.

This data-set was made available to the academics for further analysis and the consulting firm agreed to be a partner in the design of the competition. This involved engaging consultants from the firm that were well adept with the company in question, so that a long-list of potential problems could be developed.

2 Outline

The competition would involve students with an interest in business and firm analysis at all universities in Sweden. The students would be introduced to the data-set through a, for this particular case developed, BI solution with a set of OLAP cubes and KPIs already defined and ready for use. Access to this solution would be browser-based to make the threshold for getting started as low as possible.

As a second step, representatives from other universities with a potential interest in the competition were contacted. This was done through the SANTE, where universities such as Lund, Umeå and the Royal Institute of Technology were invited into the design of the assignment.

The Royal Institute of Technology (KTH) has been one of the fore-runners in ERP related education in Sweden, with a bachelor-program up and running from the turn of the new millennium. Through early discussions with KTH, the previous experience from conducting scorecard related cases was found to be a key ingredient in the

design of the competition and the assignment that was to be given to the students. The highly mathematical/analytical approach part of the KTH culture was also seen as a valuable ingredient in creating an assignment that could cater to students from both business and engineering.

The intention was to conduct the competition in the beginning of the summer of 2009. Through introducing the concept to the media and the universities within SANTE, the marketing aspects of the competition were assessed as being feasible with a limited budget.

The competition would consist of one assignment that would last for a total of three weeks. No limit for the total amount of participating students was set, and a price in the form of a monetary sum was set up through CBSs Industry Advisory Board.

The assignment was to identify and describe issues and consequences in Company Xs current operations. No limitation as to focus from the students was set, with the intent of making the competition interesting to students within all aspects of business. This implies that no matter if the students major in finance, HR, management, management accounting or logistics, they would be able to find interesting issues in operations.

Discussions were also held with several BI vendors whose products could be used as a means for accessing the data-set. The idea was to design an interface in a BI solution that could form the starting point for the students in their analysis.

The consulting firm designed a set of OLAP-cubes and KPIs that could form the starting point for a thorough analysis of the company in question. In parallel with this, they also constructed an introductory e-learning lecture that the students could use to get started with the BI solution. This included links to further assistance when it came to the basic functionality of the solution, as well as pointers in regards to how the students could go about to initiate their analysis.

Discussions were also held regarding how to assess the final reports from the students. Through previous discussions regarding which potential problems and issues that were present in the data-set, a long-list had been developed and these issues had also been graded in regards to complexity and importance/impact for operations. Finally, a dual means of assessment was agreed upon.

This dual approach consisted of one quantitative assessment where the reports were compared to the long-list, and one qualitative assessment where representatives from both industry and academia were invited to participate in a judging panel. This panel would form the final body responsible for establishing a winner out of the participants.

3 Timeline

As the figure above shows, the span of the entire project stretches from November 2008 till September 2009. During this period of time, the following actions have been/will be taken to ensure the successful completion of the competition. With this being a live project in the making, much of the project is yet to be finalized.

- Conception: The basic idea was developed in dialogue between the industry and academia.
- Formation: A formation of an initial team was initiated and the team, in turn began to assess the amount of resources necessary for the competition as well as establish senior buy-in from their respective organizations.

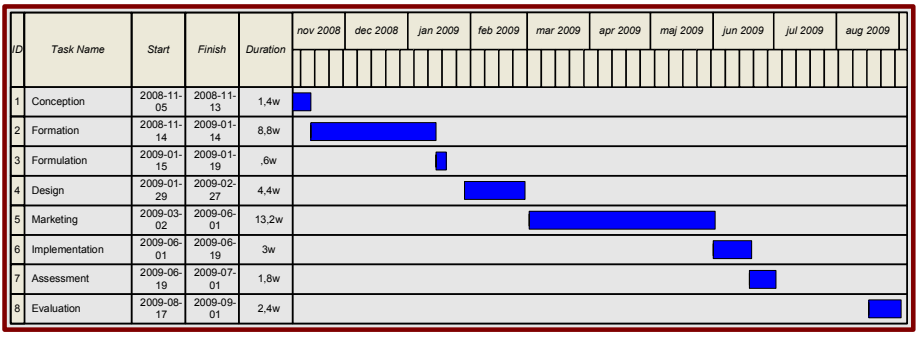


Fig. 1. Timeline of the project

- Formulation: The intent and objectives of the competition were discussed and elaborated upon in order to arrive at a conclusion in respect to what value could be derived from the project from the involved stakeholders. This included both academia, students, media and the industry. The main value and benefits that were identified for the industry was marketing- and branding-related gains, as well as a means for effectively scanning the Swedish student body for employable students. From the academic and student perspective, increased employability, exposure and pedagogical innovation were identified as the main wins.
- Design: The design phase involved working with both the technology and the analytical aspects of the task. On the technology side, an appropriate technological infrastructure ensuring easy access and a low, educational threshold was needed. With CBS having a well developed and strong infrastructure, a choice was made to use this as a starting point in the design of the technology. Secondly, an appropriate BI solution was needed as a front-end to the data-set made available by the consultancy firm. The choice stood between prior collaborators with CBS such as Microsoft (Performance Point Server) and SAP(Business Objects), and new, potential collaborators such as Targit, Qlickview and Microstrategy.
- Marketing: Early on preliminary contacts with the media was established regarding the possible implementation of a game-like construction in the summer of 2009. At the same time, marketing also focused on getting the news out to the faculty at other universities in Sweden. This was done through SANTE and SANTE Academy, the two platforms for academic collaboration regarding ERP existing in Sweden. After securing the interest from a set of large universities regarded as fore-runners in ERP inclusion and technology enabled higher education, individuals from these Universities were also involved as participants in the final design of the competition.
- Implementation: The implementation phase will consist of signing up the students/groups that will participate in the competition. At the same time, a Sharepoint site for making all communication between the participants and the competition administrators transparent so that there will be no asymmetries in the amount of information that the students receive. During early discussions the idea about releasing information regarding the case throughout the entire competition was lifted. This may be one form of working with the administration of the competition that

could propel the depth of analysis that the students reach to new levels and avoid the risk of losing groups that have a hard time with their analyses. With the competition running over a three-week period, we expect that the students will get into the deeper levels of analysis in week two and three, after first getting acquainted with the data and the interface. There was also a discussion regarding how the students should report their findings. One idea was for the students to be assessed on the basis of the scorecard/dashboard that they have developed throughout the competition. At the time being, no consensus has been reached as to if the reporting should be in report or dashboard format.

- Assessment: As previously noted, the assessment of the final reports will be made through a dual approach focusing on both quantitative as well as qualitative aspects of the findings from the students. The results of this assessment will then be communicated to the students and a final session will be held where all participants are invited, where the consultants go through the case pointing out all the key findings that the students found, as well as other aspects that might have not surfaced.
- Evaluation: The competition will be evaluated by the students as well as the administration team. The results will be communicated to all stakeholders, with a recommendation as to whether or not the competition should be run again and the potential design changes that would need to be implemented.

4 Contact and Further Information

University of Gothenburg: www.gu.se
School of Business, Economics and Law: www.handels.gu.se
Centre for Business Solutions: www.handels.gu.se/cfa
Jeeves: www.jeeves.se
SYSteam: www.systeam.se

Case Study-Design for Higher Education - A Demonstration in the Data Warehouse Environment

Daniela Hans, Jorge Marx Gómez, Dirk Peters, and Andreas Solsbach

Carl von Ossietzky Universität Oldenburg, Fakultät II - Department für Informatik, Abteilung Wirtschaftsinformatik I / VLBA, 26111 Oldenburg, Germany
{daniela.hans,jorge.marx.gomez,dirk.peters, andreas.solsbach}@uni-oldenburg.de

Abstract. In the field of economics the work on case studies is a major part in practical tertiary education since the turn of the century. The need for integrated information and communication systems among different enterprise sectors has developed over time. Accordingly, higher education has to reflect this development. The research presented in this paper focused on case study-design by analyzing existing methods and the developing of criteria's. That gives an insight how case studies can be designed to support didactics and education science aspects. The analysis of the scientific work in this field shows a gap between actual available case studies and approaches to increase the learning success in the field of education science. The results of the research will help to design case studies under the consideration of didactics and education science aspects, as exemplary case study the paper presents a case study in the data warehouse environment.

Keywords: Case Study, Higher Education, Data Warehouse, Didactics.

1 Introduction

Since the mid of the nineties the need of integrated information and communication systems among different enterprise sectors has developed over time. The systems had to organize the rapid growing mounds of data and manage them as much suitable as possible at the same time. Besides covering their normal functionality, like supporting business processes of an enterprise in a technical way, information systems became a tool in assisting experts and the leading management for making their decisions concerning the business context. Therefore the data warehouse technology came into operation. Using these analytical information systems, crucial data can be extracted from different data sources in order to enable a powerful reporting based on a huge amount of analytical-relevant data.

In the field of economics the work on case studies is a major part in practical tertiary education since the turn of the century. The American Harvard-Business-School can be seen as the initiator concerning the use of case studies in the environment of higher education. The work on case studies was introduced in 1908 in conjunction with law courses. Because of the casuistics of the lawyers, the lecturers left the traditional lecture methods and went over to a more practical way of learning. They

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 231–241, 2009.

focused on the discussion of practical cases taken from the real economic life and arranged their lectures thus more authentic, practical and application-oriented. The didactic of case studies was born and over the years there grew a huge collection of different case studies in the university, which contains more than 1000 written and tested case studies today. Later on, the didactic of case studies was furthermore applied to train interdisciplinary competencies especially in Europe. In 1954 E. Kosiol introduced the work with case studies at the Freie Universität Berlin to increase the practical work within the higher education. Basically there are existing four different types of case studies, which vary in the way the case is presented, how the information of the case is prepared, how a possible solution can be found and how the case can be solved finally: The Case Study Method, the Case Problem Method, the Case Incident Method and the Stated Problem Method [1]. The main idea of case studies are the intention to present complex situations including their problems within the business environment and to determine the learners to a mostly independent work on that specific case. Therefore it is important, that the learners develop their own solutions in a strategic way and make a reasonable decision. To sum up, case studies offer the possibility for an activity-oriented education with the focus on self-directed learning [2, 3]. Have case studies the ability to introduce current complex situations in the business environment considering the state of the art of information system?

2 Assessment Criteria and Related Work

Presenting a complete catalog of criteria for evaluating case studies would go beyond the scope of this paper. Among others the following assessment criteria could be possibly included in a catalog of criteria:

- Learning target: Is it possible to reach the learning targets, which were formulated in the teaching note?
- Target audience: Does the case study as a whole fit to the target audience? Is the case study receiver-oriented?
- Expenditure of time: Can the case study be executed in the given period of time?
- Formulation: Are language, word choice and expression appropriate to the given content and the target audience?
- Motivation: Does the scenario of the case study motivate the audience? Where are the motivating or discouraging aspects?
- Illustration: To what extend is the content of the case study uptaken via visual, acoustic and haptic senses?
- Achievement of results: Is there an achievement of the results in order to obtain the knowledge, which was created during the work with the case study?

In comprehension with the case study "Ist-Daten-Analyse" by Marx Gómez et al. [4] the mentioned assessment criteria illustration will be exemplary clarified in detail.

In opposition to a real case, the scenario in a case study can be presented or illustrated in the form of screenshots or graphics as shown in Fig. 1. Because of the high authenticity this is the ideal display format. By using labels and numberings inside the graphics it is possible to visualize the chronological order of the steps the learner has to pass through. Using texts besides these graphics is very helpful as well, because the

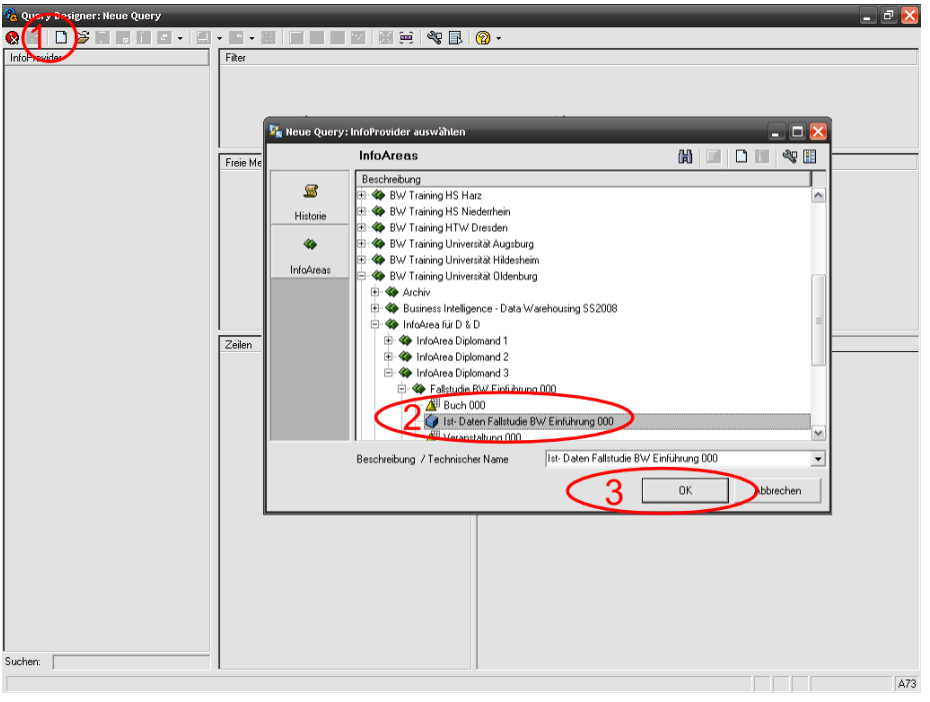


Fig. 1. SAP® BEx Query Designer - Creating query (data warehouse case study)

describing character results in a better understanding. The learner has to be guided through the case study by a consistent structure in these descriptions. For example there has to be some highlighted text for the important instructions like for the values that have to be entered or instructions that need to be executed. Relevant names of fields inside the graphics can be highlighted, too. This is also for a better understanding and a higher degree of clarity.

3 Conception of Case Studies

The background for the created concept in this work was given in the book "Grundlagen der Fallstudiendidaktik" by Perlitz and Vassen [5]. The concept of these authors was extended in this work, while not relevant aspects were not taken in consideration anymore. The four-divided conception of case study design is shown in Fig. 2. The initial point of the case study design is the motivation and information search that results in the end of the concept in the creation of a teaching note.

3.1 Motivation/Information Search

There can be two different reasons for the need of creating a case study. On the one hand the author can have a special teaching or learning target in mind which underlies

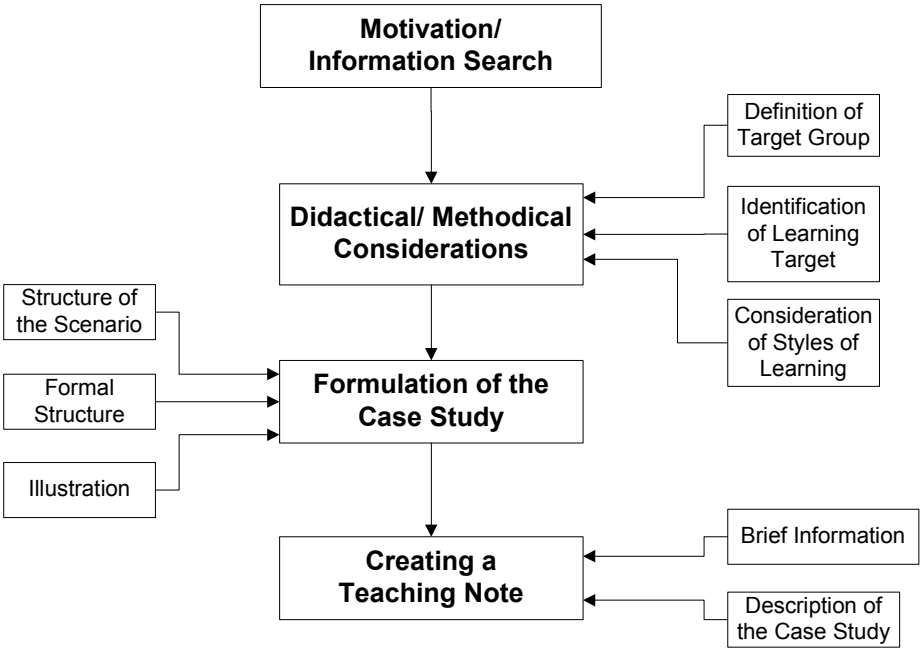


Fig. 2. Conception of case study design

a specific subject. On the other hand the motivation can arise from an article inside a newspaper, from publications of enterprises or from something similar than that.

For the first reason the information search is a prerequisite for creating a case study. Therefore it is necessary to find an appropriate enterprise scenario. Basic information about products, the market situation, the environment and international interdependencies concerning the enterprise have to be taken into consideration. Business magazines, annual reports, statistics and so forth have to be gathered. Consequently it is necessary to find an enterprise which can be taken into collaboration which allows it, to collect all these information and supports the author of the case study in doing his work.

For the second reason the author has to check first, if the present situation corresponds with his teaching or learning target. For the case that the described situation fits into his idea, he needs to find an enterprise as information source as described above. After the first contact and the agreement on the target, the way of proceeding and the basic parameters like data privacy or responsible contact persons, the gathering of information inside the enterprise can begin. The acquisition of additional information material like memos or letters which looks not relevant in the context at all, should be taking into consideration too, because they make the case study more realistic [5].

3.2 Didactical/Methodical Considerations

After completion of the information search phase the author has to decide on a learning target which wants to be addressed by his case study and for which target audience the case study has to be designed. In general, the target audience results from the

location of the case study as well as the occupational activity of the author. During the consideration of the learning target(s) it is necessary to decide whether the focus is on creating knowledge, the learning of special methods, techniques and processes for solving problems or on the impact on behavior and activity. Moreover the style of learning has to be determined: Shall the case study be used within a lecture or an exercise, which can be executed in the form of individual-, pair- or team work [5]? The problem or task has to arouse interest in the reader; he has to be motivated to find a possible solution and to spend time in working on the case study. This can be achieved by an authentic correspondence to reality. The case study has to be geared to the interest of the students or learners and although challenging for motivating the processor to deal with the problem, the conflict or the decision-making process [6]. Besides these aspects, it has to be avoided that the learner is going to be overstrained by a complex learning environment or multiple contexts. It is necessary that the acquired knowledge and the learned abilities can be transferred on different contexts and not only to stay on a single view. Therefore it can be useful to advise the learners to project the acquired knowledge on other problem situations. Not only the learning in multiple contexts but also the reflection of the teaching content from different perspectives, assists the flexible usage of knowledge [7].

3.3 Formulation of the Case Study

The work on the case study should arouse interest in the reader to deal with the underlying problem. To fulfill this condition, the learner has to be motivated to deal with the given case study and to work it through. Therefore it is necessary to design and formulate the case study in a way an incentive can be offered. For this reason it is helpful to introduce the enterprise at the beginning of the case study in a brief summarization. The historical development, market- and environment situation, acting people and the main problem of the case study should be introduced at this point [5].

It is essential to consider the prior learning of the learners, their previous knowledge about theoretical and practical experience and their psychical development. For the acquirement of new knowledge and abilities the learners need to have the possibility to understand the learning content, to integrate their previous knowledge and to build up the connection between different learning areas [7]. Therefore the case study has to be understandable and it should be geared to the knowledge, skills and abilities of the students. Depending on the initial situation of the learners, the learn process has to be designed differently. If there is a heterogeneous learning group the configuration of the case study has to be sophisticated. Processing the case study should take place in the given time frame wherefore the complexity has to be adjusted accordingly.

The text flow should be organized schematically and well structured; in order to prevent the reader from scrolling backwards oftentimes, what would result in a discouragement or a lack of concentration. In the best case, processing the case study is in a chronological order. The heading of the case study should be chosen in a way, that there are no hints given to the area of expertise or the handled problem, because this decreases the efficiency of the work by foreclosing. This applies especially to case studies, which put the focus on developing the knowledge of the learners whereby the case studies are extremely well structured [5].

In addition to the language medium, a well designed case study is characterized by clearness and good illustration. Thus, the case study has to be prepared straightforward and interesting. It is recommended to use graphics, cartoons, photos, diagrams, symbols or the like. If there is an extensive amount of material, it should be placed in the appendix, but it always has to be balanced, if the usage of the additional material in this kind of detail is necessary for reaching the learning target. If there is too much information, the learner will get clobbered which results in a d discouragement. Another increase of attractiveness can be reached via the usage of film and TV. Small sequences of realistic recorded cases including interviews of involved people for example, can improve the motivation of the learners in a remarkable way. The Internet as an information and communication platform can be useful during the work on case studies as well. Through this usage of new communication technologies and media advancement in media literacy can be expected in parallel [1]. By designing the learning area, which is a major part of the learning process, it is necessary, to animate the learners to execute the learning processes as self-directed as possible. To which degree the learner is going to acquire knowledge in a self-directed way, depends on different environment factors like the learning aid, the learning room and the learning environment as a whole. The arrangement of learning methods and -techniques and the learning material and media are measurements for the quality of the learning processes [6, 7]. Depending on the learning target the last chapter should contain a brief summary and a repetition of the decision-making situation. This is refused very often in the literature, if it is not an explicit condition of the learning target [5, 8]. But however, a summarization can be useful just for recapitulation.

3.4 Creating a Teaching Note

One problem of case studies is that they are often not allocated to the public. For this reason a case study method is often not used in lectures. Therefore it is necessary to publicize high quality, already tested and approved case studies. But only publicizing the case studies is not enough, if they are not documented in an appropriate way. Thus, the teaching note provides brief information for other teaching staff. It provides information about the content, the structure and the functionality. This allows the teacher to gain a short insight into the case study. By using the teaching note the teacher can recognize, if the usage of the case study within his lectures is reasonable or not. According to Perlitz and Vassen the structure of a teaching note can be divided into the following sections [5]:

1. Brief information
– Special field
– Target group and styles of learning
– Scope of the case study (total and textual)
– Duration
2. Description of the case study
– Dimension, branch and area of expertise of the enterprise
– Story, content of the case study
– Raising problems

3. Details for the work
– Details for the analysis of the case study
– Possible solutions and strategies
– References

The special field can be the overall labeling of the university domain, the internal or corporate seminar. An accurate definition of the target group consists of e.g. term of the students or professional qualification as educational requirement. The scope of the case study should include the number of pages, graphics, tables and the amount and kind of appendix e.g. forms and financial statements. The duration should clarify the time demand of every phase in the concept as reading, analyzing and executing and should be recorded in the duration. The seminar participants should be able to do the phases without a pressure of time which could downsize the ability to focus on the case study. The description of the case study should include information about context and exemplary enterprise (e.g. branch or scale).

References to the methodically approach, educational objective and teaching targets have to be included in details for the analysis. The teaching note as one main aspect in the presented approach should mention the references of the developed case study to support the classification and preparation of the case study [5].

4 Data Warehouse Case Study

This chapter contains the concept for a teaching unit in which the case study is integrated. This concept is divided into the following sections: Considerations about the target learning group, didactical considerations and construction of the teaching lessons in detail (technical analysis, learning targets, methodical and didactical considerations, lesson timetable). Topic of the teaching unit is the introduction of the SAP® BW software.

The target group of the case study is defined to address students in the field of computer sciences, business informatics, economics and business studies. Furthermore, learners of vocational schools, which want to deal with the SAP software in the context of their education are although address by the case study.

In the course of the case study the learners should receive an insight to the work with the SAP BW. The core of the case study work is the data modeling, the data collection and the reporting in SAP BW. Every part should be handled in three hours.

The teaching unit is separated into six lessons à 90 minutes each. The superior learning targets are illustrated in the following Table 1.

Table 1. Teaching unit content and the planned time-frame

Teaching Unit Content	Planned Time-Frame
The learners are modeling the data. Creating InfoArea, InfoObject-Catalog, InfoObject and InfoCubes.	First and second teaching lesson
The learners are modeling the Application Components, InfoSources and DataSources and are extracting Master and Transaction Data through Flat Files.	Third and forth teaching lesson
The learners are creating a Query and a Report by using the Query Designer.	Fifth and sixth teaching lesson

For the first and the second teaching lesson the didactical and methodical considerations could look like the following:

The first lesson should provide an introduction into the new topic. According to Gagné, during the organization an appropriate learning structure has to be considered. While planning the learning sequences it is necessary to take care of the hierarchical setup of concept formation, knowledge acquisition and problem solution. At the beginning the relevant terms have to be learned. After that, it is necessary to clarify whether and how these terms are in relationship to each other. Not till then it is possible to use these rules in order to find a solution. The introduction to the teaching lesson starts with a brainstorming, where the knowledge of the learners related to topic Data Warehousing is collected at the blackboard. During this proceeding the teacher gets an idea, on which level of knowledge the learners are moving. The brainstorming is guided by the teacher. In the context of a group puzzle the learning group is divided into smaller groups (master groups) which consist of four up to six members. Every master group gets the same learning material, which can be definitions or descriptions to the topic of data modeling in SAP BW like an abstract from a book for example. Afterwards the learning material is shared equally within the master group. After this every member of the group is working on his material (InfoCube, InfoObject (Characteristics, Key Figures, etc.), InfoArea and InfoProvider). During this work every member is allowed to consider other information resources (e.g. Internet), besides the already available material which was shared in the group. After this individual work phase all the members of the master groups which were working on the same topic are getting together in a so called expert group, in order to deepen their acquired knowledge. They are discussing their gathered work, answering questions to each other and are creating a presentation about their topic. Thus, the single learners are becoming experts in their field. After this expert round the expert are returning to their master groups. The expert knowledge, which was elaborated before, is shared with the whole master group by using the former created presentation. In this way, every member of the master group is informed about every special topic very well. This group puzzle is a type of group work, which assists in the acquisition of the knowledge base. The idea of this learning method is to communicate knowledge by explaining the learning matter [9]. The teacher plays a passive role in this kind of learning method. He/she is only assisting when difficulties in understanding occur.

For internalization the basic knowledge, characteristics and functionality of the software, it is necessary, that the learners are working self-contained and action-oriented. Therefore, after the group puzzle a working phase at the PC takes place. The learners are working at the first part of the case study, which was created by the teacher. In addition to the increase of decision-making and responsibility, the activity of the learners is strengthened. The work on the case study is done in individual or partner work. While encouraging the social competence and the ability to work in a team, a possible existing shortage of space can be avoided. By contrast, an individual work is more effective, because in opposition to watch a partner, the required knowledge to solve the tasks is more brought forward. The work on a case study enables a guided learning by discovering. There is only a learning help available, which avoids frustration. Furthermore the case study offers possibilities to differentiate, because the learners can configure the learning speed by themselves. The teacher stays passive during the working phase and only assists, if problems occur. Apart from the teacher,

Table 2. Course plan of the first teaching lesson

Time	Phase	Teaching content	Style of learning	Media
5 min.	Introduction	Welcome	Classroom teaching (frontal)	
10 min.	Introduction	Brainstorming	Classroom teaching (discussion)	Blackboard
15 min.	Knowledge Development	Dividing the learners into master groups, knowledge acquisition	Individual work	Working material, PC
15 min.	Deepening	Discussion with other experts, Preparing the presentation	Expert group	Working material, PC
15 min.	Presentation	Knowledge presentation in master groups	Group work	PC, Beamer
90 min.	Practice	Case study work at the PC	Individual or partner work	PC

learners, which show a strong performance in the course, can although assist their colleagues when problems occur. In this case, it is necessary to look after, that these learners are only advising and not solving the problems on their own. Especially in large learning group this method is recommended.

At the beginning of the second teaching unit, the learners have the time to continue their work on the case study. Learners, who already completed the part of data modeling, can go on with the work on the second part of the case study, the data collection in SAP BW. For the conclusion of the first part the knowledge should be reflected and discussed in classroom. Special issues can be picked out as a theme for the discussion in the whole class, single learners can present special activities without using the case study as guidance and the learners can ask questions which were not answered until now. At the same time, in this phase, the learning material is reflected once again. Shortfalls can be discovered by selective questions and a knowledge examination is possible in that way as well. Moreover the teacher can impact the motivation of the learners by giving commendation and criticisms.

5 Conclusion and Future Work

The use of teaching methods as the case study method can increase the learning success. A passive learning method as presentations supports very well the initial learning and aim-oriented learning however the case study method as a representative of active learning supports the motivation and the learning efficiency and should be considered as part in the lesson planning.

With the aid of the conception of case study design as one result of the paper the developed data warehouse case study shows how to create a case study and a lesson plan in the topic of data warehouse environment. The theoretical evaluation of the conception of case-study and data warehouse case study show that the parallel usage of theory and praxis (as tools like the SAP® Business Information Warehouse) increasing the learning success by using the procured knowledge in the same seminar. The data warehouse case study presents an approach how a case study should be structured to be able to introduce current complex situations in the business environment considering the actual state of the art. The reference list can be seen as an example that a gap between current available cases studies and approaches how to create case studies exist. Available references of the last decade cover the didactical and methodical considerations of the process to create case studies but the requirements of the current business environment are not considered for this process and case studies in the data warehouse environment are not available.

Next steps in the practical evaluation and following scientific papers should be an evaluation by using the created case study and criteria's in a seminar. The feedback from seminar participants could be used for following objectives:

- To measure the learning success of the seminar participants (one-time only passive learning methods and one time a lesson plan with a mix of active and passive learning methods)
- To measure the required time to create a comparison of required and planned time for each phase (e.g. individual work, group work, lesson discussions or execution of the case study)
- To measure the scope of information (insufficient or too much information)
- To measure the complexity of the case study

The performed research can only be seen as first step to increase the learning methods and learning success of using case studies for higher education. Further research is required, with a detailed conception and example which has been evaluated from theoretical and practical view in detail.

References

1. Kaiser, F.J.: Grundlagen der Fallstudiendidaktik - Historische Entwicklung - Theoretische Grundlagen - Unterrichtliche Praxis. In: Kaiser, F.J. (ed.) Fallstudie. Klinkhardt, Bad Heilbrunn/Obb (1983)
2. Liening, A., Paprotny, C.: Fallstudienarbeit in der Ökonomischen Bildung (2005), http://www.wiso.uni-dortmund.de/wd/de/content/forschung/publikationen/downloads/unido_wd_08.pdf (visited October 21, 2008)
3. Weitz, O.B.: Fallstudienarbeit in der beruflichen Bildung. Verlag Dr. Max Gehlen, Bad Homburg (1996)
4. Marx Gómez, J., Rautenstrauch, C., Cissek, P., Grahlher, B.: Einführung in SAP Business Information Warehouse. Springer, Heidelberg (2006)
5. Perlitz, M., Vassen, P.J.: Grundlagen der Fallstudiendidaktik. Peter Hanstein Verlag GmbH, Köln (1976)

6. Krapp, A., Weidenmann, B. (eds.): Pädagogische Psychologie. 4. Vollständig überarbeitete Auflage. Verlagsgruppe Beltz, Weinheim (2001)
7. Brettschneider, V.: Entscheidungsprozesse in Gruppen. Theoretische und empirische Grundlagen der Fallstudienarbeit. Klinkhardt, Bad Heilbrunn/Obb (2000)
8. Alewell, K., Bleicher, K., Hahn, D.: Entscheidungsfälle aus der Unternehmenspraxis. Gabler, Wiesbaden (1971)
9. Frey-Eiling, A., Frey, K.: Das Gruppenpuzzle. In: Wiechmann, J. (ed.) Zwölf Unterrichtsmethoden Vielfalt für die Praxis: Vielfalt für die Praxis, Beltz (2006)

Off-the-Shelf Applications in Higher Education: A Survey on Systems Deployed in Germany

Henry Schilbach, Karoline Schönbrunn, and Susanne Strahringer

Technische Universität Dresden, Faculty of Business Management and Economics
Business Information Systems Group
01062 Dresden, Germany
{Karoline.Schoenbrunn,Susanne.Strahringer}@tu-dresden.de

Abstract. Off-the-shelf applications have seen a steady growth in many industries. The goal of the paper is to explore whether this also applies to higher education as a domain where institutions are facing new challenges and thus might be on the verge of modernizing their application landscape. The results of a survey conducted among institutions of higher education in Germany show that there are a few dominant products that already automate the highly administrative process areas. However, there still are a few process areas with a low degree of automation. Surprisingly, users of home-grown systems do not necessarily plan to move onto "better" solutions even though at least two modern integrated products show promising interest among German institutions of higher education.

Keywords: off-the-shelf application, campus management system, enterprise resource planning system, higher education, student lifecycle, survey.

1 Introduction

Off-the-Shelf applications have seen a steady growth over the years. Enterprise resource planning (ERP) systems as a predominant category within off-the-shelf applications were initially deployed by large manufacturing firms. Today, market saturation in capital-intensive manufacturing industries is high and dissemination in other industries, such as service industries is quite remarkable. Penetration into non-typical domains has started in industries where automation was traditionally high but based on home-grown legacy applications rather than off-the-shelf systems. Banking is an example of an industry where willingness to replace custom-built legacy applications with ERP systems has seen a steady growth in recent years. [1] Lately, vendors additionally started targeting non-typical domains with low degrees of automation such as the higher education (HE) sector.

In service industries such as banking or higher education traditional ERP systems primarily comprise industry-independent back-office functionalities whilst leaving core-business processes untapped. In contrast to manufacturing or retail the scope of ERP functionality in service industries is smaller and rather focuses on non-specific processes, such as accounting or human resources (HR). Thus, industry-specific core functionality is either provided by specific add-ons to ERP systems or independent

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 242–253, 2009.

industry-specific applications such as core banking solutions or campus management systems. However, these industry-specific off-the-shelf systems with a focus on core business processes may also reach for typical ERP functionality by comprising support for generic administrative processes (e.g. accounting, HR) as well. Thus, the differentiation between these different types of off-the-shelf systems is increasingly blurred. In a domain such as higher education we additionally observe the phenomenon that apart from transactional systems there is a need of collaborative functionality to support learning processes or alumni relations. These types of systems usually are portal or groupware-based and non-transactional thus requiring totally different underlying platforms. The heterogeneity of applications deployed in higher education thus seems to be high and is probably causing a rather fragmented application landscape and integration problems. Therefore, the goal of our research is to give a first overview on the current situation in German institutions of higher education. Our research questions can be stated as:

1. What is the degree of application-support (automation) in core-business processes in German institutions of higher education?
2. What are the predominant off-the-shelf applications deployed to support core-business processes in German institutions of higher education?

The rest of our paper is structured as follows. In section two we provide a short overview on the core business process areas we differentiate in higher education. As we intend to analyze higher education institutions (HEI) in Germany only we give a short overview (section three) on the background and products of three important vendors of off-the-shelf applications targeting HEI in Germany. Eventually, section four describes the methodology and results of a survey we conducted in the second half of 2008.

2 Core Business Process Areas in Higher Education

In order to further analyse application support we have broadly categorized core business process areas in higher education. The business process areas we differentiate are:

- Learning, i.e. processes to support students' learning
- Teaching, i.e. processes to support the delivery of teaching with sub-processes such as
  - class scheduling
  - room allocation
  - course evaluation
- Student Lifecycle, i.e. processes centering around students' records and lifecycle with sub-processes such as
  - admissions & enrollment
  - student records
  - examinations
  - alumni relations

Our categorization is rather coarse-grained and might not be totally complete. However, for the purpose of our study we were in need of a rather simple and straightforward categorization which is aligned along typical application areas rather than a value chain approach as suggested in [2].

3 Major Vendors and Products

In order to give a first overview on products deployed in German HEI we want to mention three vendors one might assume to play a major role.

The *Hochschul-Informations-System (HIS) GmbH* was founded in 1969 by the Volkswagen Foundation as a non-profit organisation and was taken over in 1975 with the Federal and state governments as shareholders. HIS has been offering IT-support to higher education administrations for over thirty years. This know-how regarding decision-making, work and organisational structures in German higher education is unique to HIS and is used to continually update and develop their software, which is widespread in German HEI. [3] HIS' offerings comprise a portfolio of self-contained modules that run independently and offer interfaces for data exchange and in the case of some modules for web-based access. Recently, HIS added a new product, HISinOne, to their offerings which is "a web-based integrated software system that covers all business processes of universities of different sizes and forms of organization" [4].

A new player in the market, the *Datenlotsen Informationssysteme GmbH*, founded in 1993 in Hamburg started developing an integrated campus management solution, CampusNet in 1999. The solution aims to be comprehensive and is continuously being enhanced. [5] First adopters of CampusNet were smaller (private) universities but the company has managed to win new customers among the bigger public schools also.

SAP, as one of the major ERP vendors offers an industry solution for higher education and research [6]. Being a German vendor with a very high market share in traditional ERP industries SAP's solution certainly needs to be taken into account when looking at the German market. Although SAP's ERP system is used in some of the universities' administrations covering the typical industry-independent back-office functionality such as accounting, HR and eProcurement [7] its industry-specific counterpart supporting HEI's core processes has not gained foothold in German universities yet. This might be due to the fact that SAP's industry solution was originally developed with the help of universities in the US and the UK [8] [9].

4 Systems Deployed in German Institutions of Higher Education

4.1 Survey Design

As off-the-shelf applications in higher education can be considered an emerging field, we opted for a descriptive exploratory research approach in order to gain a first understanding of system dissemination and heterogeneity within this specific domain. In selecting appropriate research objects, we confined our study by focusing on German institutions only. As mentioned in the previous section two major German vendors do not offer their products on a broad international basis. We therefore consider the German market to be somehow different from other countries and therefore wanted to specifically explore it before starting international comparisons.

We used a questionnaire as our research instrument. Prior to finalizing the questionnaire we conducted telephone interviews with university staff comparable to those we intended to address in order to pre-test parts of the questionnaire and further refine other sections.

Based on a complete list of higher education institutions in Germany (as of November 2007, see [10]) we excluded 10 institutions from our list for which size in student numbers was not available. For the remaining 345 institutions we checked websites to identify the right person to address which were positions like head of IT or staff explicitly responsible for campus management systems. As we were not successful in the case of 25 institutions we finally addressed 320 institutions via a personalized email sent to the person identified including a link to the online-questionnaire.

Data collection took place in September-October 2008. By the end of October 126 questionnaires had been returned. Hence, we were able to achieve a 39,4% response rate.

4.2 Sample Characteristics

Institutions participating in our survey have student numbers ranging from 31 to 45.363 with an average of 7.760 students. Table 1 shows the institutions' distribution along types and ownership structures. The numbers imply that our sample might be biased with public institutions and universities being slightly over-represented. As public universities are the most common type of higher education institutions in Germany we feel that this bias does not strongly threaten the sample's validity.

Table 1. Selected sample characteristics

Categorization of institutions	Number of institutions	Number of participating institutions
By type		
Universities of applied sciences	172 (49,9%)	54 (42,9%)
Universities	116 (33,6%)	49 (38,9%)
Universities of music and arts	57 (16,5%)	23 (18,3%)
By ownership		
Public	233 (67,5%)	105 (83,3%)
Private	70 (20,3%)	14 (11,1%)
Church	42 (12,2%)	7 (5,6%)
	345 overall	126 participating

4.3 Selected Survey Results

When comparing the three process areas, learning, teaching, and student lifecycle, 85,2% of participating HEIs find application support for the student lifecycle very important compared to 47,5% in teaching and 38,5% in learning (see Fig. 1). Although learning seems to be the area where application support is not deemed as important as it is in the other areas it still is at the same level with at least teaching when combining "very important" and "important". Although HEIs might consider teaching a field where application support is not mandatory they have understood that supporting their students' learning processes on an excellent level must not be neglected.

This assumption is also supported when comparing application support's importance with actual support. Fig. 2 shows the portion of HEIs that already deploy applications in the mentioned process areas. 67,5% of participating HEIs already have systems in place to support *learning* processes which is a higher portion than the sub-process

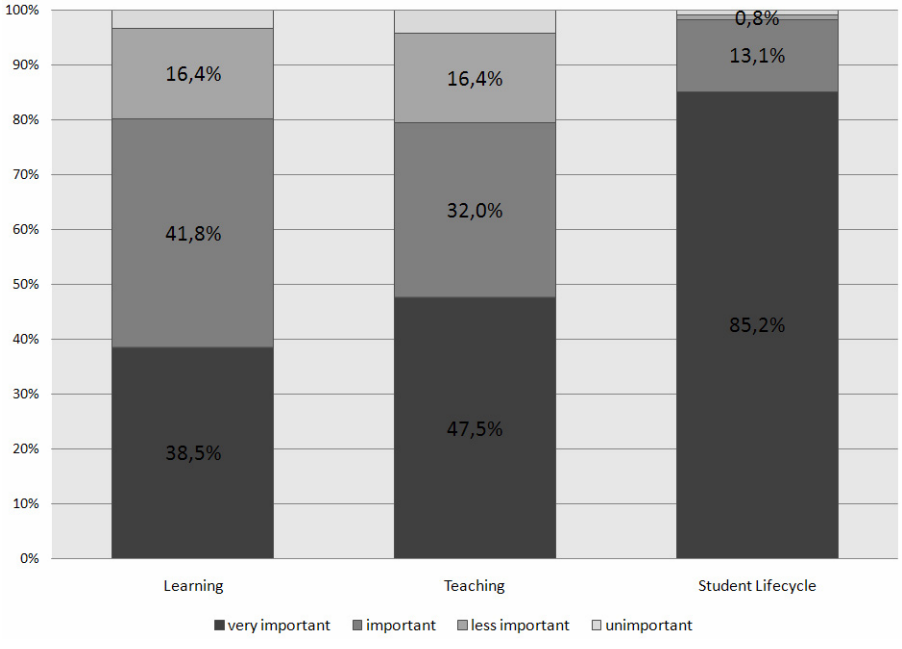


Fig. 1. Importance of application support in three core process areas

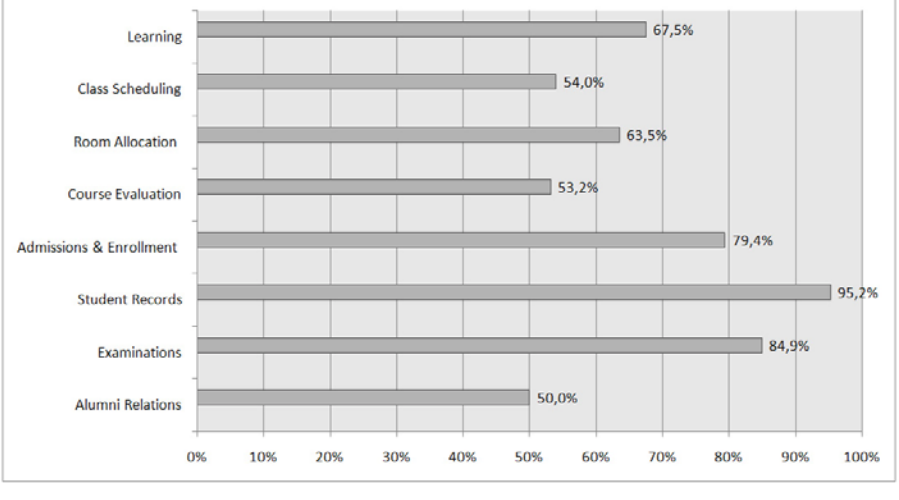


Fig. 2. Current state of application support in core business process areas (ratio of HEIs deploying applications to support process areas at sub-process level)

areas of teaching (class scheduling 54%, room allocation 63,5%, course evaluation 53,2%) which again underlines that learning support might not be mandatory but is considered important.

Within the sub-process areas of *teaching* 63,5% of HEIs use applications to support room allocation compared to lower numbers in class scheduling and course evaluation. In the case of course evaluation this might be due to the fact that courses are not yet systematically evaluated in every German HEI and even if they are this might be organized in a highly decentralized manner with smaller units such as chairs or departments being responsible for their evaluation on a voluntary basis using paper based-solutions without any application-support. The higher numbers of schools deploying applications in support of room allocation in contrast to class scheduling might be due to the fact that rooms are centrally allocated in many German HEIs, in many cases with one central unit being responsible for this task whereas class scheduling is usually coordinated at the school/faculty or department level. Another reason might be that automated scheduling with optimization features is a very sophisticated task schools do not ask for as long as they somehow cope with manual planning processes that yield satisfying results. A typical German public university was used to offering many optional courses without students registering for courses beforehand so that little scheduling with respect to avoiding overlaps was done. As registration for courses is the exception rather then the rule class scheduling does not deal with mass data. However, while moving into modularized programs with less flexible offerings German universities will certainly need to move on to more sophisticated class scheduling processes even if course registration will not become the rule.

In the process area *student lifecycle* not surprisingly many HEIs (95,2%) manage student records and related functionality such as re-registration via database applications. In order to avoid media breaks and considering the large numbers of students applying most schools try to automate as much as possible in the early stages of the student lifecycle with 79,4% of HEIs using applications during admissions & enrollment. And again these processes are usually dealt with in a centralized manner throughout the whole institution thus making automation worthwhile even for smaller institutions. That 84,9% of participating HEIs deploy applications in exam administration can probably be attributed to the fact that these tasks are highly standardized even if exam administration is done at the school/faculty level and thus is not fully centralized throughout the whole institution. Another reason might be that exam planning, registration and grading must be highly formalized, and error free and thus should avoid risks of error-prone manual processes. Additionally, exam administration requires preparation of documents such as transcripts on the basis of individual students and also needs to report grades to students individually and confidentially. These requirements can hardly be met in an efficient manner without application support.

Not surprisingly alumni relations is the sub-process area with the lowest number of HEIs (50%) deploying applications. This is certainly due to the fact that German HEIs have only recently started to professionalize alumni relations. Typical for German HEIs are clubs/societies organized by alumni themselves without formal integration into the institution. Most of these clubs are organized around faculties or degrees and do not unite the overall institution's alumni. All these aspects might be reasons why still many HEIs do not deploy applications themselves to support alumni relations. But with more and more schools moving on to a formalized institutional level alumni management this will certainly change in a few years.

In order to be able to analyze vendor/product penetration into the different areas we also asked HEIs for the specific applications they are currently deploying (see Table 2).

Table 2. HEIs using home-grown vs. off-the-shelf applications

		Teaching			Student Lifecycle			
Process areas / Applications	Learning	Class Scheduling	Room Allocation	Course Evaluation	Admissions & Enrollment	Student Records	Examinations	Alumni Relations
Home-grown applications	13,5%	10,3%	15,1%	19,8%	3,2%	5,6%	4,0%	37,3%
Off-the-Shelf applications:	54,0%	43,7%	48,5%	33,4%	76,2%	89,7%	81,0%	14,7%
HIS (Modules)	23,0%	16,7%	18,3%	0,8%	64,3%	74,6%	65,1%	0,8%
CampusNet	5,6%	4,8%	4,8%	4,0%	3,2%	4,8%	4,8%	1,6%
SAP	0,0%	0,0%	0,0%	0,0%	0,8%	0,0%	0,0%	0,0%
HISinOne	0,0%	0,8%	0,8%	0,0%	0,0%	0,0%	0,0%	0,0%
Other	25,4%	21,4%	24,6%	28,6%	7,9%	10,3%	11,1%	10,3%

Table 2 shows the dominance of off-the-shelf applications, especially the HIS module-based solution in the highly administrative process areas such admissions & enrollment, exam administration and student records where mass data needs to be processed efficiently. The integrated solution HISinOne is too young to be widespread and SAP has not yet gained foothold in the German market. The only integrated solution that has started gaining ground is CampusNet. Although its penetration is still low it is remarkable that it is deployed in every process area by at least a few HEIs implying that CampusNet is a very comprehensive system that covers all mentioned areas.

Off-the-Shelf systems' penetration is very low in alumni relations where most HEIs use home-grown applications. One HEI uses a HIS module for alumni management that HIS stopped developing in 2005 in order to include the functionality into their new product HISinOne [11]. It is also noteworthy that although CampusNet seems to be a comprehensive system its functionality might not be that convincing in the field alumni relations.

The second area where off-the-shelf systems' penetration is quite low is course evaluation. Apart from CampusNet almost none of the explicitly mentioned products is deployed. However, 28,6% of HEIs use other products which in the case of course evaluation is a group of products dominated by one system, a product called EvaSys[1] from Electric Paper, a vendor specialized on automated document solutions. Although the system can be used for online-questionnaires its success is probably mainly due to the bulk-processing of paper questionnaires.

[1] http://www.electricpaper.biz/products-en/evasys-en.html

Another process area where the "other" category should be looked at in more detail is learning. In this area six HEIs use a system called Moodle[2], four ILIAS[3] and another four Stud.IP[4] with all solutions being open source systems.

In the process area class scheduling seven HEIs use a system called S Plus from UK-based Scienta[5]. The system also supports room allocation and was explicitly designed to enable European universities to meet Bologna-induced challenges. Especially its automated timetabling seems to be highly sophisticated and might be the reason for the products' success in Germany despite its not being low-priced.

In the process area room allocation three HEIs use a system called UnivIS[6] which encompasses some of the other process areas as well. However, HEIs seem to deploy it mainly in this field.

Due to several reasons such as new products being put on the market and new bologna-induced requirements challenging universities we assumed many German HEIs might be on the verge of modernizing their systems. Hence, we also asked HEIs whether they planned to change their current system. Fig. 3 illustrates HEIs' plans in the process areas learning and teaching, whereas fig. 4 focuses on the student lifecycle.

In most process areas the users of HIS modules seem to be waiting for HISinOne and planning to adopt this new product generation from their current vendor. This is especially the case in the highly administrative sub-processes of the student lifecycle.

The analysis also shows that none of the HEIs using CampusNet has plans to adopt another system. This might be attributed to HEIs satisfaction with the system but is probably mainly due to the fact that it is a fairly new product.

HEIs using no systems at all until now show quite some reluctance in adopting a system in the near future. Although in a few areas these HEIs seem to be waiting for HISinOne as well there are still others with a low tendency to move onto automated processes. Also home-grown application users do not show as much inclination to change their system into an off-the-shelf solution as one might have expected.

Surprisingly, SAP as one of the major ERP vendors worldwide with an extremely high market penetration in other industries especially in Germany does not seem to be able to position its higher education solution among German HEIs. Although SAP's ERP system does have quite some relevance when looking at HEIs supporting processes such as accounting and HR its industry solution seems not to be able to catch up with its generic back-office counterpart in the near future. Also surprisingly, very well known systems that are quite successful in other countries such as Blackboard do not manage to penetrate the German market. With open source products being quite common in the field of learning pricing might be considered a major inhibitor.

The questionnaire also included a few free text fields asking participants for comments they wanted to add. The statements underline the results presented so far. The

[2] http://www.moodle.de/
[3] http://www.ilias.de/
[4] http://www.studip.de/
[5] http://www.scientia.com/uk/
[6] http://www.univis.de/

Bologna-induced reorganization of study programs is in most cases the main reason for the introduction or modernization of systems. Problems with current systems institutions mentioned are e.g. insufficient data quality, missing support by lecturers and difficult configuration and tailoring of systems. Participating HEIs also criticize missing experts or staff being able to align technology with organizational needs, a lack of faculty-wide solutions caused by inflexible systems and too little open source initiatives. Most of the participating HEIs pin their hopes on HISinOne – expecting easy configuration, simply handling and holistic coverage.

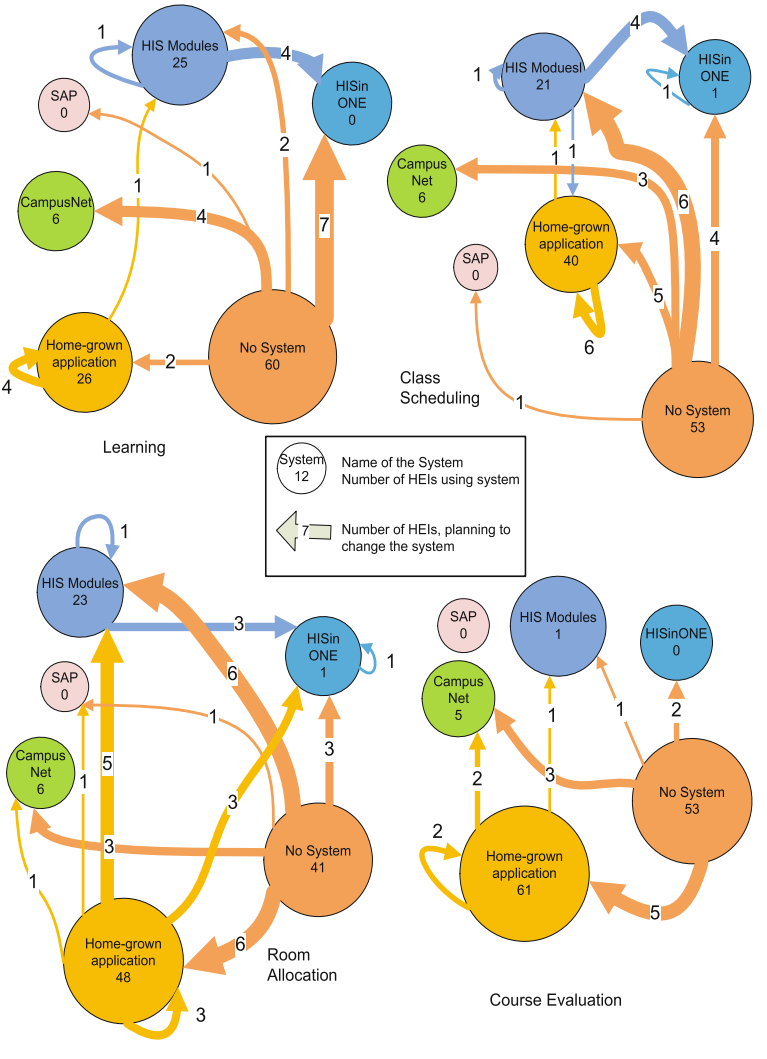


Fig. 3. HEIs' plans to change applications in process areas *learning* and *teaching*

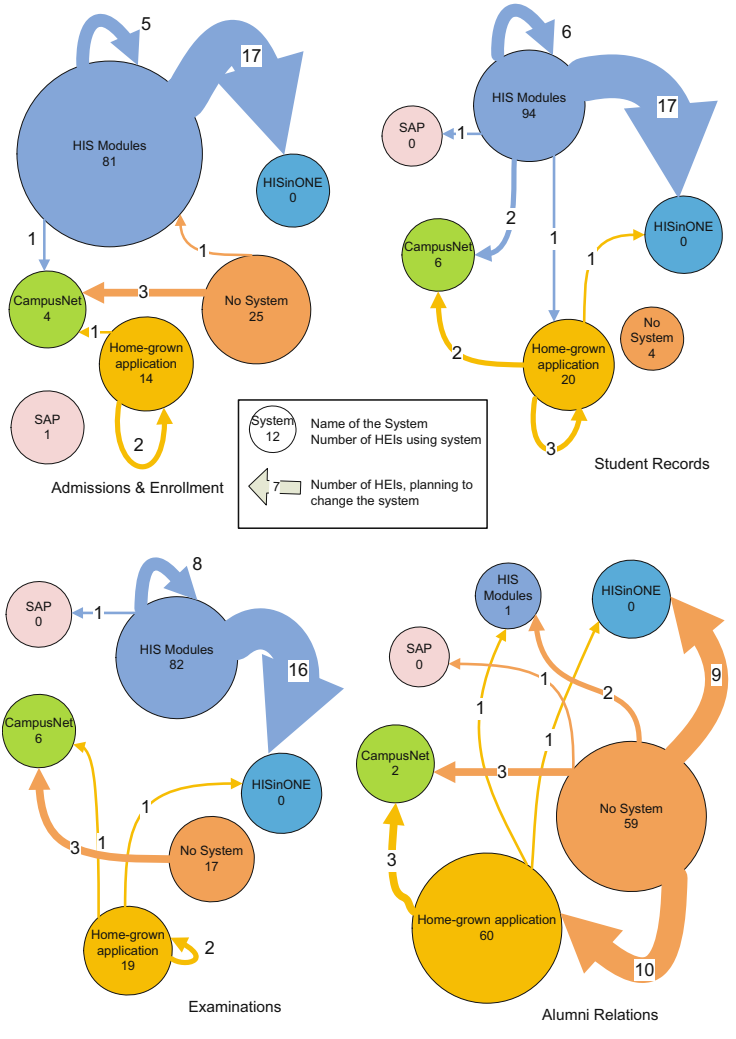


Fig. 4. HEIs' plans to change applications in process area *student lifecycle*

5 Conclusion

Our study showed that off-the-shelf application support in German HEIs is quite widespread in the highly standardized core-process areas along the student lifecycle (excluding alumni relations). This area is dominated by a module-based solution from HIS. Surprisingly, with three promising candidates on the market integrated solutions are still rather an exception than the rule. HEIs adopting HIS modules seem to be waiting for their vendor's new product HISinOne. A few other institutions have adopted CampusNet with some more planning to do so. Although HEIs ask for open source solutions the only field where these play a major role is the process area learning.

The obvious need for high-quality low-cost solutions somehow shapes the German market with little opportunities for vendors with higher priced products. Such vendors only seem to have chances with specific functionalities where a highly sophisticated product may be seen as a unique selling proposition with advantages that cannot be met by integrated solutions. We observed this phenomenon in the process areas class scheduling and course evaluation.

Overall we felt that the trend to modernize application landscapes is not at the level we had expected beforehand. The only field where we observed a strong trend towards higher degrees of automation is alumni management with many institutions having no system at all. With German HEIs recognizing how important student relationship management [12] is – especially when moving from a one- (diploma) to a two-cycle system (bachelor/master) and with concepts such as lifelong learning at hand – this follows our expectations. However, alumni reations will not necessarily become a domain of off-the-shelf applications with many HEIs using home-grown systems already and others planning to do so.

What needs to be further analyzed in the bigger German public universities is the autonomy that will be granted to faculties and departments with respect to application support. Although many German HEIs are used to being highly decentralized with high degrees of autonomy granted to faculties and schools in these decisions this may also be one of the major factors inhibiting integrated solutions and seamless university wide processes. Centralization even across universities such shared service centers may be an option to better benefit from economies of scale and may help unleash synergies.[7] Although this might not be appropriate in all process areas there certainly are a few where this model could save resources urgently needed in other areas.

References

1. Fuß, C., Gmeiner, R., Schiereck, D., Strahringer, S.: ERP usage in banking: an exploratory survey of the world's largest banks. Information Systems Management 24(2), 155–177 (2007)
2. Zielinski, W.: ICT and Competitive Advantage in Higher Education: Applying Value Chain Model and Balanced Scorecard for University. European University Information Systems - EUNIS, Manchester (2005)
3. HIS: HIS Organisational Profile, `http://www.his.de/english/organization` (15.1.2009)
4. HIS: About HISinOne, `http://www.hisinone.de/english` (15.1.2009)
5. Datenlotsen: CampusNet, `http://www.datenlotsen.de/index.php?se=399` (15.1.2009)

[7] An example showing that universities have started making advantage of such concepts are HEIs in Saxony who centrally run and improve a learning management system which is based on OLAT (http://www.olat.org), an open source system originally developed at University of Zurich. BPS (https://bildungsportal.sachsen.de/), a spin-off from 11 universities, acts as an OLAT service provider for higher education in the state of Saxony. BPS operates and supports a large OLAT installation under the brand OPAL for its university clients and offers the same services also to other clients in the field of higher education in Germany.

6. SAP: SAP für Hochschulen & Forschungseinrichtungen, http://www.sap.com/germany/industries/highered/businessmaps/index.epx (15.1.2009)
7. Rode, C.: Releasewechsel auf mySAP ERP: Rechenzentrum legt Basis für sichere Zukunft (2005), http://www.apentia-online.com/UP/Apentia/files/Article/Realtech_Rechenzentrum_legt_Basis_f_r_sichere_Zukunft.pdf (15.1.2009)
8. Pollock, N., Cornford, J.: Customising industry standard computer systems for universities: ERP systems and the university as a 'unique' organisation. In: 2nd IEEE Conference on Standardization and Innovation in Information Technology (2001)
9. Pollock, N., Cornford, J.: Fitting Standard Software to Non-Standard Organisations. In: Proceedings of the 2002 ACM Symposium on Applied Computing (2002)
10. Statistisches Bundesamt: Anschriftenverzeichnis deutscher Hochschulen (ohne Verwaltungsfachhochschulen), Excel-Datei des Statistischen Bundesamtes, Bereich VI B – Hochschulstatistik, Stand November (2007)
11. Dettmer, M.: Vorstellung des HIS – Alumni-Moduls ALU (2007), http://www.his.de/pdf/SOSZUL2007/A20.pdf
12. Hilbert, A., Schönbrunn, K., Schmode, S.: Student Relationship Management in Germany – Foundations and Opportunities. Management Revue 18(2), 204–219 (2007)

LIT 2009 Workshop Chairs' Message

Piotr Stolarski and Tadeusz Tomaszewski

Poznan University of Economics, Department of Information Systems,
al. Niepodleglosci 10, 60-967, Poznań, Poland
{P.Stolarski,T.Tomaszewski}@kie.ue.poznan.pl

1 Introduction

It is our greatest pleasure to present to the Readers the post-proceeding publication of the papers presented during the second edition of Legal Information Technology Workshop in Poznań on 28th April 2009.

This time the Workshop took place in conjunction with the 12th International conference on Business Information Systems. It was a one-day event spitted into two thematically different sessions.

At this 2nd edition, out of 8 papers allowed to the review process, 5 of them have been accepted. It is however important to say that this moderately high level of acceptance rate (62.5%) was resulted by the equally high review notes which were independently granted by at least 3 reviewers from the Programme Committee[1] only.

Another vital issue is that – in accordance with the factual workshop formula – the publications printed on the following pages are not only the simple copy of originally submitted texts, but they are the result of the reflection of discussions and substantive comments made by workshop participants during the Workshop itself and accompanying events. Therefore they represent enriched and hopefully broadened or improved versions of those papers.

The field of Legal Information Technology is becoming more and more prosperous after a period of limited interest from the research audience. Now it is becoming to play a central role in the search of innovation form the side of legal and knowledge scientists, legal practitioners and e-government solution developers.

In addition to this trend another phenomenon becomes visible. The weight of Legal Informatics as a discipline begins to move towards some smaller but better-defined groups of problems. This movement effects in possibility of setting apart portions of knowledge or sub-disciplines.

The paper of Stede and Kuhn is focused on legal office automation and processing of selected legal documents. This task is also important for the next step legal information retrieval systems, which on the other hand is the topic of the text authored by Bianchi et al. The results obtained by them are telling yet eventually the authors admit that ideally, solutions for such type of systems ought to use ontologies. Latter knowledge models are the main issue of the article by Stolarski and El Kharbili. The authors show that other usage possibilities for legal ontologies will probably soon emerge. But regardless of new ways of employment of ontological knowledge, those more

[1] In some cases of largely diversified notes we were taking advice of another additional reviewer.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 254 – 255, 2009.

typical like support for expert systems as knowledge storage solution are also vital. This is mainly stressed by the text of Zurek and Kruk. In their short paper they present their approach towards tax law support for agricultural purposes. But on top of the system itself some extraordinary reasoning techniques are exemplified. Another perspective on Legal Information Technologies shows us Nawrot. This time she presents the legal office practitioner' attitude towards social-Web connected casus. In fact the number of incidents and breaches of Law in the Internet is growing rapidly. Thus this issue will generate a number of challenges in the near future.

To the special attention of the Reader we would like to recommend the invited paper of prof. John Zeleznikow. As a internationally renown legal IT and AI specialist he give us a fresh and inspiring perspective on the Legal Technologies' research in general, asking about the matters of justice and fairness (the fundaments of any modern legal system) in the computer legal aid systems.

Above all we would also like to share our deepest thanks to all the participants and authors whose work and attendance enabled us to coin this Workshop into a very successful event. Special warmest thoughts we direct to the Programme Committee members. Without their wise judgments we would not be able to select interesting and discussion-triggering works. We strongly believe that the next-year Legal Information Technology Workshop which will take place in Berlin will produce even more impact in the community.

Thinking about the future we encourage you to take a closer view at the content provided this year.

The Need to Incorporate Justice into Negotiation Support Systems

John Zeleznikow

School of Information Systems, Victoria University, Melbourne, 8001, Australia
john.zeleznikow@vu.edu.au

Abstract. Over the past twenty five years there has been a movement towards resolving legal disputes through mediation and negotiation rather than litigation. Perceived benefits of this move towards Alternative Dispute Resolution include disputants having more control of the dispute and potential solutions, reduced costs and speedier decision making. If Alternative Dispute Resolution becomes the norm for resolving legal disputes, then we must ensure that the negotiation support systems that we develop utilize legally fair paradigms. But how can we develop measures, or at the very least principles, for the development of legally just? Through an examination of bargaining in the shadow of the law and principled negotiation we suggest principles which when applied, will encourage fairness and justice in the development of negotiation support systems. Such principles include transparency, bargaining in the shadow of the law and the need for discovery. We also illustrate the pitfalls of using such principles.

Keywords: Negotiation Support Systems, Justice, Bargaining in the Shadow of the Law, BATNAs.

1 Introduction

The development of '*fair*' and '*just*' negotiation support systems and Online Dispute Resolution environments should lead to an increasing confidence in the use of e-commerce. But there are no accepted norms on how to measure what is '*fair*' and '*just*' negotiation support?

Our interest in justice and negotiation arose because of our research into plea bargaining [14]. In both Australian and United States criminal law jurisdictions, a defendant can appeal a decision if they believe the judicial process was flawed. However, when negotiating about pleas, a participant cannot challenge the decision. The reason for this situation is that unlike in a trial, the defendant has pleaded guilty and thus admitted that he committed the crime. This situation becomes problematic in the admittedly few cases where a person accepts a plea bargain even though they did not commit the crime. The defendant may plead guilty because he was offered a heavily reduced sentence (e.g. no jail time) and he felt the probability that he would be found guilty[1] is reasonably high.

[1] Often because of poor legal support.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 256–267, 2009.

Thus, it is very difficult to undo an '*unfair plea negotiation*'. But it is also essential that it be possible to reverse unfair decisions.

Alexander [1] has argued that in Australian Family Law, women tend to be more reluctant than men to continue conflict and are more likely to wave their legal rights in a mediation session. McEwen et al [15] believe family mediators focus upon procedural fairness rather than outcome fairness. Phegan [19] argues that differences in power between men and women lead to negotiated results that favour men. Field [8] argues that as victims of domestic violence increasingly find themselves in the mediation context, specific strategies are needed to protect their interests and ensure their safety.

As a further example of bargaining imbalances, Alexander argues that because women tend to be more reluctant than men to continue conflict, if their major goal is to be the primary care giver for their children, they may reach a negotiated settlement, which whilst acceptable to them is patently unjust. The wife may for example, give the husband the bulk of the property, in return for her being granted the primary care of the children. Whilst such an arrangement may meet the goals of both parents, it might not meet the paramount interests of the children, who could be deprived of subsequent financial resources.

Bargaining imbalances can thus produce *unfair results* unless mediators overcome them. But should mediators try to redress imbalances? And how can we determine what are fair results?

It is vital that we develop '*fair*' and '*just*' negotiation support systems. Indeed, one of the barriers to the uptake of Online Dispute Resolution relates to users' concerns about the fairness and consistency of outcomes achieved by any Online Dispute Resolution approach. Pierani [20], in discussing Online Dispute Resolution in Italy, argues that as with Alternate Dispute Resolution models, Online Dispute Resolution systems need to be impartial, transparent, effective and fair.

Re notions of fairness, take for example a marriage in Australia where the couple have been married for fifteen years and have three children, one of whom has special needs. Suppose the husband works full-time, whilst the wife is not employed outside the house and is a full-time carer for the husband and children. Suppose they own a house valued at $400,000 with a mortgage of $250,000. Further, the husband earns $45,000 per annum.

Given that that this is both a low income and low asset marriage (the common pool is let us say $180,000) the wife might be expected to receive 70% of the common pool. Were she to retain the house she would need to pay the husband $45,000. The husband would need to pay Child Support which is mandated by the relevant law.

In many circumstances, the fact that the husband has a low income and is paying substantial child support, may mean that he cannot afford to pay rent. He might thus be forced to return to living with his parents. Australian men's groups have vigorously protested at what they perceive as injustices.

Are such results fair or just? The answer depends on how we measure fairness. If we measure fairness by meeting the interests or needs of both parents equally, then the answer is clearly no. In Australia, our notion of justice focuses upon meeting the paramount interests of the children. In Australia, this principle of the paramount

interests of the children outweighs other principles of justice, including the interests of parents and other concerned parties. Hence the solution suggested above, is eminently fair according to Australian Law.

Australian Family Law is one domain where interest-based notions of mediation can conflict with notions of justice. In such domains, the use of negotiation support systems that attempt to equally satisfy both parties is limited.

One lesson learned from the evaluation of family law disputes is that suggested compromises might conflict with law and justice. This problem can arise where a fully automated Online Dispute Resolution environment is used in which resolution is based on consensus. Nevertheless, we believe that an Online Dispute Resolution environment may still play a positive role in the family-law setting.

One safeguard for use of Online Dispute Resolution in fields such as family law may be required certification of the result by a legal professional. A comparable field is discrimination law where conciliation is used to ensure the principles of the legislation are not compromised by the interests of the parties.

2 Bargaining in the Shadow of the Law and Principled Negotiation

2.1 Bargaining in the Shadow of the Law

Traditional Negotiation Support Systems have focused upon providing users with decision support on how they might best achieve their goals [22]. Traditionally, negotiation support systems were not designed to model legal disputes. They focus upon business transactions, where the parties can walk away from failed negotiations without conducting further discussions.

In legal domains, the failure to resolve a dispute, often leads to litigation. A fundamental issue arises, in construction negotiation support systems in legal domains: namely *is the system being developed solely concerned with supporting mediation or do we also need to consider issues of justice?* How can we balance the importance of issues of justice with the need to support mediation? When issues of justice are not reflected in the outcome of the mediation process, bargaining theory has its limitations. Bargaining imbalances can thus produce *unfair results* unless mediators overcome them.

This is especially true when negotiating about charges and pleas in the domain of criminal sentencing. In this domain, the two parties often have very different resources, a well supported prosecution versus an impoverished defence. Further, the consequences of an unfair negotiation can be dire – the incarceration of an innocent defendant, cannot easily be reversed.

Plea bargaining brings administrative efficiency, certainty and reduced costs.

So how can we ensure that the benefits of plea bargaining are maintained whilst fairly allocating punishment? The issue of fairness in negotiation is important in all domains of law, but no more so in sentencing where the adversaries are not two individuals, but an individual and the state.

Priest and Klein [21] claim that the potential transaction costs of litigation provide an incentive for nearly all legal suits to settle.

Galanter [10] claims:

In the federal courts, the percentage of civil cases reaching trial has fallen from 11% in 1962 to 1.8% in 2002. In spite of a five-fold increase in case terminations, the absolute number of civil trials were 20% lower in 2002 than it was 40 years earlier.

In writing about the Vanishing American Trial, Galanter argues that whilst litigation in the United States is increasing, the number of trials decided by US judges has declined drastically. Two of the reasons for this phenomenon are because average trials are getting longer and more complex and litigants are using alternative forms of Dispute Resolution.

Most negotiations in law are often conducted in the shadow of the Law i.e. bargaining in legal domains mimics the probable outcome of litigation. Mnookin and Kornhauser [17] introduced the bargaining in the shadow of the trial concept. By examining the case of divorce law, they contended that the legal rights of each party could be understood as bargaining chips that can affect settlement outcomes.

Cooter et al [6] discuss Bargaining in the Shadow of the Law for civil cases. This model now dominates the literature on civil settlements.

2.2 Principled Negotiation

Walton and Mckersie [29] propose that negotiation processes can be classified as distributive or integrative. In distributive approaches, the problems are seen as "zero sum" and resources are imagined as fixed: *divide the pie*. In integrative approaches, problems are seen as having more potential solutions than are immediately obvious and the goal is to *expand the pie* before dividing it. Parties attempt to accommodate as many interests of each of the parties as possible, leading to the so-called *win-win* or *all gain* approach.

Most negotiation outside the legal domain law focuses upon interest-based negotiation. Expanding on the notion of integrative or interest-based negotiation, Fisher and Ury [9] at the Harvard Project on Negotiation developed the notion of principled negotiation. Principled negotiation promotes deciding issues on their merits rather than through a haggling process focused on what each side says it will and will not do. Amongst the features of principled negotiation are: separating the people from the problem; focusing upon interests rather than positions; insisting upon objective criteria and knowing your BATNA (Best Alternative To a Negotiated Agreement).

Knowing one's BATNA is important because it influences negotiation power. Parties who are aware of their alternatives will be more confident about trying to negotiate a solution that better serves their interests. is unwilling to reconsider the offer, walking out is a very sensible option. BATNAs not only serve a purpose in evaluating offers in the dispute, they can also play a role in determining whether or not to accept a certain dispute resolution method. Mnookin [16] claimed that having an accurate BATNA is part of the armory one should use to evaluate whether or not to agree to enter a negotiation. Comparing the possible (range of) outcomes with alternative options encourages parties to accept methods that are in the interests of disputants and enables them to identify those that are not. It is likely that most parties, to some extent, test the values of their BATNAs when assessing whether or not to opt for a certain dispute resolution method.

3 Incorporating BATNAs and Bargaining in the Shadow of the Law to Support Fair Negotiation in Australian Family Law

In their development of a three step model for Online Dispute Resolution, [13] evaluated the order in which online disputes are best resolved. They suggested the following sequencing:

1. The negotiation support tool should provide feedback on the likely outcome(s) of the dispute if the negotiation were to fail – i.e. the BATNA.
2. The tool should attempt to resolve any existing conflicts using dialogue techniques.
3. For those issues not resolved in 2, the tool should employ compensation/trade-off strategies in order to facilitate resolution of the dispute.
4. Finally, if the result from 3 is not acceptable to the parties, the tool should allow the parties to return to 2 and repeat the process recursively until either the dispute is resolved or a stalemate occurs.

If a stalemate occurs, arbitration, conciliation, conferencing or litigation can be used to reach a resolution on a reduced set of factors. This action can narrow the number of issues in dispute, reducing the costs involved and the time taken to resolve the dispute.

Lodder and Zeleznikow's model, in suggesting providing advice about BATNAs, facilitating dialogue and suggesting trade-offs, focuses upon E-Commerce applications. They claimed that their research assumes that disputants focus upon interests. But as we shall discuss in chapter five, the notions of Bargaining in the Shadow of the Law and BATNAs have important implications for developing just negotiation support systems.

3.1 Enhancing Interest Based Negotiation: The Family Winner and Asset Divider Systems

Bellucci and Zeleznikow [3] supported interest based negotiation in their Family Winner system. They observed that an important way in which family mediators encourage disputants to resolve their conflicts is through the use of compromise and trade-offs. Once the trade-offs have been identified, other decision-making mechanisms must be employed to resolve the dispute. They noted that:

- The more issues and sub-issues in dispute, the easier it is to form trade-offs and hence reach a negotiated agreement, and
- They choose as the first issue to resolve the one on which the disputants are furthest apart – one party wants it greatly, the other considerably less so.

Family_Winner asks the disputants to list the items in dispute and to attach importance values to indicate how significant it is that the disputants be awarded each of the items. The system uses this information to form trade-off rules. The trade-off rules are then used to allocate issues according to a '*logrolling*' strategy

Family_Winner accepts as input a list of issues and importance ratings that represent a concise evaluation of a disputant's preferences. In forming these ratings, the system assumes that the disputants have conducted a comparison of the issues. As noted by [26], bargainers are constantly asked if they prefer one set of outcomes to

another. Thus Sycara suggests considering two issues at a time, assuming all others are fixed. Family_Winner uses a similar strategy in which pair-wise comparisons are used to form trade-off strategies between two issues.

The trade-offs pertaining to a disputant are graphically displayed through a series of trade-off maps [31]. Their incorporation into the system enables disputants to visually understand trade-off opportunities relevant to their side of the dispute. A trade-off is formed after the system conducts a comparison between the ratings of two issues. The value of a trade-off relationship is determined by analyzing the differences between the parties, as suggested by [18].

When evaluating the performance of the Family_Winner system, family law solicitors at Victoria Legal Aid had one major concern – that Family_Winner in focusing upon mediation had ignored issues of justice; which is equated in the Australian family law jurisdiction with what is in the best interests of the child. They claimed that Bellucci and Zeleznikow had focussed upon the interests of the parents rather than the needs of the children. If the parents' negotiation is based upon what they perceive as the best interests of the child, then indeed family law negotiation is interest based. However if one accepts the arguments that women tend to be more reluctant than men to continue conflict and are more likely to wave their legal rights in a mediation session, then some safeguards need to be incorporated into the negotiation process, to ensure that the interests of the children are paramount.

The evaluators noted that the wife may for example; give the husband the bulk of the property, in return for her being granted the primary care of the children. Whilst such an arrangement may meet the goals of both parents, it does not meet the paramount interests of the children, who will be deprived of subsequent financial resources.

Family Law is one domain where interest-based notions of mediation can conflict with notions of justice as defined in the Act[2]. In such domains, the use of negotiation support systems that attempt to equally satisfy both parties is limited unless we also incorporate notions of justice.

The Queensland Branch of Relationships Australia wants to use a modified version of Family_Winner to provide decision support for their clients. The application domain is concerns agreements about the distribution of marital property. Instead of Family_Winner attempting to meet both parents' interests to basically the same degree, mediators at Relationships Australia determine what percentage of the common pool property the wife should receive (for example 60%). A major issue of concern to Relationships Australia is how to equate the percentage of property with the interests of the couple. It is not necessary that there be a direct connection between the financial value of an item and the points-value that each party in the dispute attaches to the item. Indeed, a major issue in dispute may involve determining the value of the item. For example following a divorce, the husband may agree that the wife should be awarded the marital home. In this case it would be in his interests to overvalue the house (say he suggests it is worth $1,200,000) whilst it is in the wife's interests to undervalue the house (say she suggests it is worth $800,000).

[2] Commonwealth Family Law Act (1975). See www.austlii.edu.au/au/legis/cth/num_act/fla1975114/s1.html Last accessed August 18 2008.

Unlike the Family_Winner system, the AssetDivider system allows users to input negative values. This development is necessary because family mediation clients often have debts (such as credit card debts and mortgages) which are as much items in the negotiation as assets.

Further, to ensure that AssetDivider proposes an acceptable solution, it might be necessary to include as a universal issue in all disputes, a cash variable payment item. For example, where the wife has identified that her highest preference is to retain the family home, an outcome might provide for her to keep the matrimonial home and the mortgage[3]. In order to reach an acceptable settlement, the wife might need to make a cash payment to the husband. Hence the requirement that a variable appear in the output is stipulated.

A further limitation of the AssetDivider system (arising from its similarity to the AdjustedWinner algorithm) is the need for users to enter numerical values. Whilst disputants can probably linearly order the significance to them of all items in dispute, it is unrealistic to expect them to give a numerical value to each item. But it is not unreasonable for the users to assign a linguistic variable to each item. A seven point Likert scale which can then be converted into points is suggested:

The development of AssetDivider allows the concept of interest-based negotiation as developed in Family_Winner to be integrated with notions of justice. The advice about principles of justice can be provided by decision support systems that advise about BATNAs or human mediators. But how can we develop fair BATNAs?

3.2 Developing BATNAs: The Split Up System

In the Split-Up project, Stranieri et al [25] wished to model how Australian Family Court judges exercise discretion in distributing marital property following divorce. The Split-Up system Section 79(1) of the *Family Law Act* (1975) empowers judges of the Family Court to make orders altering the property interests of parties to the marriage but does not lay down procedural guidelines for judicial decision makers. In practice, judges of the Family Court follow a five-step process in order to arrive at a property order:

1. Determine which assets will be paramount in property considerations. This is referred to as common pool property distribution.
2. Determine a percentage of the property to be awarded to each party.

The Split-Up system implements steps 1 and 2 above, namely the common pool determination and the prediction of a percentage split. According to domain experts, the common pool determination task (Step 1) does not greatly involve the exercise of discretion, in stark contrast to the percentage split task (Step 2) as directed graphs from domain experts.

To collect data for the Split-Up system, Stranieri, Zeleznikow, Gawler and Lewis read family court judgements. Values for relevant factors were extracted from each case. Ninety-four variables were identified as relevant for a determination in consultation with experts. The way the factors combine was not elicited from experts as rules or complex formulas. Rather, values on the ninety-four variables were to be

[3] A negative item.

extracted from cases previously decided, so that a neural network could learn to mimic the way in which judges had combined variables.

Whilst the Split—Up system was not originally designed to support legal negotiation, but it can be directly used to proffer advice in determining your BATNA. Suppose the disputants' goals are entered into the Split—Up system to determine the asset distributions for both W & H. Split—Up first shows both W and H what they would be expected to be awarded by a court if their relative claims were accepted. The litigants are able to have dialogues with the Split—Up system about hypothetical situations which would support their negotiation.

Bellucci and Zeleznikow [2] give an example of a divorcing couple who had been married twenty-years and had three children. The husband worked eighty hours per week whilst the wife did not engage in employment outside the home. They entered three scenarios into the Split—Up system. The system provided the following answers as to the percentages of the distributable assets received by each partner.

Resolution	H's %	W's %
Given one accepts W's beliefs	35	65
Given one accepts H's beliefs	58	42
Given one accepts H's beliefs but gives W custody of children	40	60

Clearly, custody of the children is very significant in determining the husband's property distribution. If he were unlikely to win custody of the children, the husband would be well advised to accept 40% of the common pool (otherwise he would also risk paying large legal fees and having ongoing conflict).

Hence, while Split-Up is a decision support system rather than a negotiation support system, it does provide disputants with their respective BATNAs and hence provides an important starting point for negotiations.

4 Principles for Developing Fair Negotiation Support Systems

Having examined interest based and principled negotiation and bargaining in the shadow of the law as well as family mediation and bargaining about charges and pleas, we now wish to develop a framework for developing just negotiation support systems.

4.1 Fairness Principle 1 – Developing Transparency

As we have seen from a discussion of negotiating about pleas and charges, it is essential to be able to understand and if necessary replicate the process in which decisions are made. In this way unfair negotiated decisions can be examined, and if necessary, be altered. The same statement holds in family mediation.

There is wide spread support for the development of transparent processes in dispute resolution. For example, at the commencement of all mediation conferences,

Relationships Australia (Queensland) clearly indicate to the disputants, how the process will be managed. They follow the model discussed in Sourdin (2005):

Opening, Parties' Statements, Reflection and Summary, Agenda setting, Exploration of Topics, Private Sessions, Joint negotiation sessions and Agreement/Closure

Whilst this model of mediation is regularly used in primary family dispute resolution and most commercial mediation, it is not used in negotiations about charges and pleas. The model involves a formal turn-taking exercise, which is inappropriate in less formal settings of negotiation such as plea bargaining. Some of the criticism of plea bargaining is that it does not, in general, follow a formal transparent model. However, even in plea bargaining, changes are occurring.

To improve the dilemma of plea bargaining, Wright and Miller [30] introduce the notion of *prosecutorial screening*. The prosecutorial system they envisage has four interrelated features: early assessment, reasoned selection, barriers to bargains and enforcement. Bibas [4] argues that a great gulf divides insiders and outsiders in the criminal justice system, whether in litigation or negotiating about charges and pleas. He claims that by making decision-making more transparent, the criminal justice system can help minimize injustices.

Even when the negotiation process is transparent, it can still be flawed if there is a failure to disclose vital information. Such knowledge might greatly alter the outcome of a negotiation. Take for example the case of a husband who declares his assets to his ex-wife and offers her eighty per cent of what he claims is the common pool. But suppose that he has hidden from his ex-wife, ninety per cent of his assets. Thus, in reality, he has only offered her eight per cent of the common pool.

Discovery, the coming to light of what was previously hidden, is a common pretrial occurrence. As Cooter and Rubinfield [7] and Shavell [24] point out, in litigation, the courts may require that a litigant disclose certain information to the other side; that is, one litigant may enjoy the legal right of discovery of information held by the other side. Interestingly enough, Shavell claims that the right of *discovery* significantly increases the likelihood of settlement, because it reduces differences in parties' information. This benefit is often lost in a negotiation.

The failure to conduct adequate discovery can be a major flaw in negotiation. But how can we conduct sufficient discovery without losing many of the benefits that derive from negotiation?

4.2 Fairness Principle 2 – Bargaining in the Shadow of the Law and the Use of BATNAs

As discussed previously, most negotiations in law are conducted in the shadow of the law i.e. bargaining in legal domains mimics the probable outcome of litigation. These probable outcomes of litigation provide beacons or norms for the commencement of any negotiations (in effect BATNAs). Bargaining in the Shadow of the Law thus provides us with standards for adhering to *legally just* and *fair* norms.

By providing disputants with advice about BATNAs and Bargaining in the Shadow of the Law and incorporating such advice in negotiation support systems, we can help support fairness in such systems.

For example, in the AssetDivider system, interest based negotiation is constrained by incorporating the paramount interests of the child[4]. By using Bargaining in the Shadow of the Law, we can use evaluative mediation (as in Family Mediator) to ensure that the mediation is fair.

The Split_Up system models how Australian Family Court judges make decisions about the distribution of Australian marital property following divorce. By providing BATNAs it provides suitable advice for commencing fair negotiations.

There is however a certain danger in promoting transparency and Bargaining in the Shadow of the Law for negotiation support.

- **Disputants might be reluctant to be frank**.
- **Mediators might be seen to be biased**.
- **The difficulty and dangers of incorporating discovery into negotiation support systems** – discovering appropriate information is complex, costly and time consuming. As previously noted, Katsh and Rifkin [12] state that compared to litigation, Alternative Dispute Resolution includes advantages of lower cost; greater speed and a less adversarial process. By insisting upon certain basic levels of discovery, we might lose these benefits.
- **The inability to realise the repercussions of a negotiation** – often disputants focus upon resolving the dispute at hand. They fail to realise that the resolution they advocate may have larger scale repercussions.

Thus, our proposed principles for developing fair negotiation support systems also have some drawbacks.

5 Conclusion

We have seen that one of the major concerns from disputants using alternative dispute resolution is about the fairness of the process and of the outcomes when confronted with a superordinate ideal of fairness (such as the paramount ideal in Australian Family law being the best interest of the children). Without negotiation procedures being seen as fair and just, there will always remain legitimate criticisms of the process. But how can we measure the fairness of alternative dispute resolution procedures?

Whilst meeting disputants' interests is a vital part of the negotiation process, it is also incorporate principles of justice into negation support systems.

Through an examination of the relevant literature in a variety of domains – including international conflicts, family law and sentencing and plea bargaining – and an in depth discussion of negotiation support tools in Australian Family Law, we have developed a set of important factors that should be incorporated into 'fair' negotiation support processes and tools. These factors include: Transparency; Bargaining in the Shadow of the Law and BATNAs; and Limited Discovery.

Incorporating these factors, does however have some drawbacks for the development of negotiation support systems.

[4] In this case, Relationships Australia (Queensland) input into the system what percentage of the Common Pool both the husband and the wife will receive.

References

1. Alexander, R.: Family mediation: Friend or Foe for Women. 10:1 Australasian Dispute Resolution Journal 8(4), 255 (1997)
2. Bellucci, E., Zeleznikow, J.: Representations for decision making support in negotiation. Journal of Decision Support 10(3-4), 449–479 (2001)
3. Bellucci, E., Zeleznikow, J.: Developing Negotiation Decision Support Systems that support mediators: a case study of the Family_Winner system. Journal of Artificial Intelligence and Law 13(2), 233–271 (2006)
4. Bibas, S.: Transparency and Participation in Criminal Procedure. New York University Law Review 81, 911–966 (2006)
5. Brams, S., Taylor, A.: Fair Division, from cake cutting to dispute resolution. Cambridge University Press, Cambridge (1996)
6. Cooter, R., Marks, S., Mnookin, R.: Bargaining in the shadow of the law: a testable model of strategic behavior. The Journal of Legal Studies 11(2), 225–251 (1982)
7. Cooter, R., Rubinfeld, D.: An Economic Model of Legal Discovery. Journal of Legal Studies 23, 435–463 (1994)
8. Field, R.: A feminist Model of Mediation that centralizes the role of lawyers as advocates for participants who are victims of domestic violence. The Australian Feminist Law Journal 20, 65–91 (2004)
9. Fisher, R., Ury, W.: Getting to YES: Negotiating Agreement Without Giving. Haughton Mifflin, Boston (1981)
10. Galanter, M.: The Vanishing Trial: An Examination of Trials and Related Matters in State and Federal Courts. Journal of Empirical Legal Studies 1(3), 459–570 (2004)
11. Honeyman, C.: Patterns of bias in mediation. Missouri Journal of Dispute Resolution, 141–150 (1985)
12. Katsh, E., Rifkin, J.: Online Dispute Resolution: Resolving Conflicts in Cyberspace. Jossey-Bass, San Francisco (2001)
13. Lodder, A., Zeleznikow, J.: Developing an Online Dispute Resolution Environment: Dialogue Tools and Negotiation Systems in a Three Step Model. The Harvard Negotiation Law Review 10, 287–338 (2005)
14. Mackenzie, G., Vincent, A., Zeleznikow, J.: Negotiating about charges and pleas – balancing interests and justice. In: Climaco, J., Kersten, G., Costa, J.P. (eds.) Proceedings of Group Decision and Negotiation, Proceedings – Full Papers, INESC Coimbra, Portugal, pp. 167–180 (2008) ISBN: 978-989-95055-2-0
15. McEwen, C., Rogers, N., Maiman, R.: Bring in the Lawyers: Challenging the Dominant Approaches to Ensuring Fairness in Divorce Mediation. Minnesota Law Review 79, 1317–1412 (1995)
16. Mnookin, R.: When Not to Negotiate. University of Colorado Law Review 74, 1077–1107 (2003)
17. Mnookin, R., Kornhauser, L.: Bargaining in the shadow of the law: The case of divorce. Yale Law Journal 88, 950–997 (1979)
18. Mnookin, R., Peppet, S., Tulumello, A.: Beyond Winning: Negotiating to Create Value in Deals and Disputes. The Belnap Press of Harvard University Press (2000)
19. Phegan, R.: The Family Mediation System: An Art of Distributions McGill Law. Journal 40, 365 (1995)
20. Pierani, M.: ODR Developments under a consumer perspective: The Italian Case. In: Proceedings of Second International ODR Workshop, pp. 43–45. Wolf Legal Publishers, Nijmegen (2005)

21. Priest, G., Klein, B.: The Selection of Disputes for Litigation. Journal of Legal Studies 13, 1–55 (1984)
22. Raiffa, H.: The Art and Science of Negotiation: How to Resolve Conflicts and Get the Best Out of Bargaining. The Belknap Press, Cambridge (1982)
23. Raiffa, H., Richardson, J., Metcalfe, D.: Negotiation Analysis: The Science and Art of Collaborative Decision Making. The Belknap Press, Cambridge (2002)
24. Shavell, S.: Economic Analysis Of Litigation And The Legal Process, Discussion Paper No. 404, John M. Olin Center For Law, Economics, And Business, Harvard University, Cambridge, Ma (2003) ISSN 1045-6333
25. Stranieri, A., Zeleznikow, J., Gawler, M., Lewis, B.: A hybrid—neural approach to the automation of legal reasoning in the discretionary domain of family law in Australia. Artificial Intelligence and Law 7(2-3), 153–183 (1999)
26. Sycara, K.: Machine Learning for Intelligent Support of Conflict Resolution. Decision Support Systems 10, 121–136 (1993)
27. Thompson, L.: Information Exchange in negotiation. Journal of Experimental Social Psychology 27, 161–179 (1991)
28. Thiessen, E., McMahon, J.P.: Beyond Win-Win in Cyberspace. Ohio State Journal on Dispute Resolution 15, 643 (2000)
29. Walton, R., Mckersie, R.: A Behavioral Theory of Labor Negotiations. McGraw-Hill, New York (1965)
30. Wright, R., Miller, M.: The Screening/Bargaining Tradeoff. Stanford Law Review 55, 29–117 (2002)
31. Zeleznikow, J., Bellucci, E.: Family_Winner: integrating game theory and heuristics to provide negotiation support. In: Proceedings of Sixteenth International Conference on Legal Knowledge Based System, pp. 21–30. IOS Publications, Amsterdam (2003)
32. Zeleznikow, J., Bellucci, E., Vincent, A., Mackenzie, G.: Bargaining in the shadow of a trial: adding notions of fairness to interest-based negotiation in legal domains. In: Kersten, G., Rios, J., Chen, E. (eds.) Proceedings of Group Decision and Negotiation Meeting 2007, Concordia University, Montreal, Canada, vol. II, pp. 451–475 (2007)
33. Zeleznikow, J., Stranieri, A.: Split Up: The use of an argument based knowledge representation to meet expectations of different users for discretionary decision making. In: Proceedings of IAAI 1998 — Tenth Annual Conference on Innovative Applications of Artificial Intelligence, pp. 1146–1151. AAAI/MIT Press (1998)

Building-Up a Reference Generic Regulation Ontology: A Bottom-Up Approach

Marwane El Kharbili[1,*] and Piotr Stolarski[2,*]

[1] ARIS Research, IDS Scheer AG, Altenkesselerstrasse 17, 66115, Saarbrücken, Germany
Marwane.Elkharbili@ids-scheer.com
[2] Poznan University of Economics, Department of Information Systems,
al. Niepodleglosci 10, 60-967, Poznań, Poland
P.Stolarski@kie.ue.poznan.pl

Abstract. The aim of this paper is to develop a Reference Normative Ontology. This generic ontology is being developed in order to allow modeling policy-based regulations. Such modeling is for creating models of norms and rules and can be used by a larger framework for compliance checking. The modeling of regulations using this ontology can be showcased for semantic business process management. Our study is based on 3 domain case-specific ontologies describing norms and behavioral standards within Web portals. The analysis of those concrete, small ontologies allows building an abstract, generic ontology which is destined to cover core aspects most of other specific regulation examples. This ontology is designed to grow by integrating additional case-specific normative documents and serve as the core component of a generic regulation modeling ontology.

Keywords: legal ontology, business policies, legal constraints, semantic web.

1 Introduction

Regulations are necessary to express which behavior in which context is to be expected from an entity by another entity. These regulations can be of various types, food regulations such as the FDA[1], financial regulations (such as BASEL II or SOx), or contracts between several organizations and individuals conducting business. Regulations are also in the wide majority of cases only available as natural language text, hard to understand for non domain or law experts, and containing potential contradictions and lacks of precision. Understanding these regulations for enforcement, auditing, implementation or training purposes requires some level of formality in representing regulations.

The goal of this work is to present a pragmatic, bottom-up approach for creating a generic policy-based and semantic regulatory framework for modeling regulations. Such a framework can thus be used in further research by other semantic frameworks for modeling regulatory aspects. Initial results have been integrated with a policy

[*] Both authors are first authors and contributed equally to this research.
[1] Food and Drugs Administration.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 268–279, 2009.

ontology developed in the SUPER project[2] on SBPM, which allows us to claim that the semantic regulatory framework introduced here can allow modeling regulations for semantic business processes. In the following, we first explain our approach more in detail in Section 2. Section 3 follows by presenting the case regulations on the basis of which our regulatory framework has been designed. Section 4 then gives a short overview on the state of the art of research on regulation modeling and Section 5 finally introduces the current state of work in developing our generic regulation ontology. This paper then proceeds in Section 6 to a discussion with an overview of related work and Section 7 contains our concluding remarks.

2 The State of Research in the Legal Domain

In the field of legal knowledge management and representation the problem of representing legal knowledge in the form of variety of knowledge bases or ontologies has been vastly recognized [3]. As a consequence a number of generally elaborated methods of ontology engineering have been tested to produce legal ontologies. Some of those methods were also the subject to create specific solutions for legal domain embedded tasks of building semantic knowledge repositories.

Legal ontologies are created to fulfill numerous aims [4] as well as support different functions. Those aims and functions delimit the outline and structure of the ontology as well as define content and expressivity. Thus, methods used to construct the knowledge models are to reflect the needs and intentions of constructors.

Legal knowledge models are rather specific ones even in the world of semantic knowledge modeling. The main reason for that is that legal ontologies must very properly reflect the very meaning of legal terms. Legal terms on the other hand are exceptionally sensitive to even subtle changes in the way they are used or defined. Therefore very subtle nuances and differences have to be handled in the process of ontology engineering as their expression has major influence on usefulness of the generated models. Going further, legal knowledge should be considered as a domain requiring high expertise – the knowledge sources, i.e. texts – constitute highly-specialized corpuses which constitute an obstacle in ease of processing and as a result make the engineering process costly.

Probably the first influential resume of works in the field of ontology creation with legal knowledge is the work of Visser and Bench-Capon [5]. This paper sums up the results of works done by – among others - Valente (1995), McCarty (1989), Visser (1995), Kralingen (1995, Visser, Kralingen, et al. 1997). This article of 1999 apart from presenting the results enumerated works goes even further in formulating a large number of accurate remarks and guidelines in proper constructing of legal knowledge models. For instance Visser and Bench-Capon citing their earlier works give the list of minimal features of any adequate ontology:

1. Epistemological adequacy (understood as epistemological clarity, intuitiveness, relevance, completeness and discriminative power),
2. Operationality (encoding bias, coherence, computationability),
3. Reusability (task-and-method reusability, domain reusability).

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

Those criteria are built on the earlier deliberations of Gruber (Gruber, 1995). The authors of the cited work also elaborate on different types of legal ontologies referencing them to general views on types in ontology engineering in general.

Another text, which is very relevant for our purposes is [6]. Although the paper is less focused on knowledge models methods it is on contrary mainly connected with the idea of building frameworks on the basis of legal ontologies. Those framworks are designed to make comparison and harmonization of legislation components. The text also raises the problem of differences between legal systems and diversified ways of expressing norms and regulations. This problem is also present in our findings and should be taken into consideration in later stages of this work.

The cited paper turns the attention of the reader to a number of interesting factors. Like, for instance, the enlistment of frames of regulations comparison and harmonization:

- Policy comparisons,
- Forecasting and reconstructions,
- Migrations.

The authors also pay much attention to the problem of legal concepts similarity. The similarity of ontological elements is widely explored in the works connected to the field of ontology matching (see for instance [7]). It is however issue that is much more difficult to handle within the legal applications as the similarity measures in this case should take into account the whole system of norms which the elements is connected with. Finally the authors give a formal introduction to representing systems of norms through ontologies.

The article of Bourcier et al. [8] shed a light on specific techniques and methodology of building legal ontologies. The authors begin with the statement, that the design of such ontologies raises knowledge management but also jurisprudence issues. They then start to present the effects of work of two groups of 25 researchers which aim was to produce two types of legal ontologies.

The group used and shared different resources in the course of the engineering. The text stresses also the fact of taking into account the various points of view on law, which is essential when doing this type of conceptual work. They consider a 3-dimensional space for situating this modeling problem. The dimensions should reflect three types of settings:

- Legal point of view,
- Action point of view,
- Logical point of view.

The authors of this text suggest also bearing in mind another 3-dimensional space when building legal ontologies. In the other case the axes are connected with the ontology goals perspective. This means that knowledge experts should take into account the factors of: Lexicographic vs. computational modeling, Coping with various levels of details by presenting upper-level concepts opposed to descriptive ontology, More or less decision-making approach.

There is also one more important differentiation in the cited paper. It is the distinguishing between legal operational engineering and legal cognitive engineering.

The work of [9] deals with the issue of accurate representation of provisions in the perspective of sophistication and complexity of legal regulations. The authors turn the attention to two excluding – in their opinion – options of modeling legal knowledge:

- To create sophisticated models capable of representing rules of arbitrary legal complexity,
- To focus on a subset of individual legal rules which are more amenable to simplified computational representation from a legal theoretical perspective.

Surden, together with his co-authors claims that certain norms coming from given legal systems or branches of regulation tends to be more challenging to represent and apply in computable contexts than others [10]. They suggest that is possible to identify discrete legal rules which are likely to be, from a legal theoretical standpoint, amenable to simpler computational representation. The text is also important on account of introducing the idea of representational complexity – a measure offering information on the degree to which legal theoretical issues are likely to complicate the task of computationally modeling a given legal rule. They also extend this idea be presenting the factors that influence this measure by applying framework which reflects the stages of rule life-cycle model.

Another paper [11] aims to merge three fields, proposing to use artificial intelligence (AI) techniques to support verification of consistency and completeness within processes in the legal domain, namely the document flow within prosecution and court institutions. The authors propose to build a legal ontology. This ontology should describe the type of information in the documents and then be stored in the meta-data of those documents. Then fuzzy-matching techniques are suggested for use in order to enforce internal consistency. The authors also stress the presented approach as highly value-adding. As we will see in the next sections, our work leads us to similar conclusion and conforms the need for, among other aspects, such ontology of legal information and legal processes.

3 Approach to Building a Generic Legal Ontologies

Below in the Fig.1 **.** we present the outline for the method we decided to use in order to attain our goal, namely to prepare material and develop the generic ontology of norms.

Our study is mainly based on very close and strict observation and lessons learned from previous works with the similar character. We analyzed carefully all the presented methods and related works in the field of legal ontologies building and maintenance as well as selected works connected to specific applications of such knowledge models which approached to the employment of our model as planned. The authors have also a moderate experience in building this type of models related from earlier works and research activities, as for example [12].

The proposed method is made up of 7 steps. We assume that the input for our work is to prepare a set of selected cases. The cases are analyzed independently and mind maps are created upon them. The cases which seem to be mostly diversified are then taken to produce their formalized versions. In our case the formalism used was

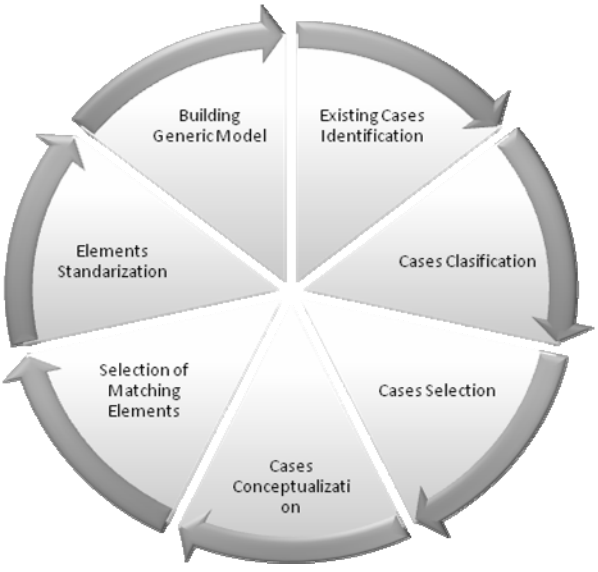


Fig. 1. Method for Building a Generic Ontology of Norms

WSML[3]. As a result we developed 3 ontologies which are better described in next section. In the process of matching elements selection and final agreement of concepts the output is obtained in the form of generic ontology. This preliminary version of the resulting ontology is presented in Section 5.

4 Examples and the Individual Ontologies

During our study we created 3 individual ontologies. All three ontologies are domain specific knowledge models, developed in order to reflect the policy of selected Web portals. Such policies are mainly declared in a semi-formal document referred to as "terms of use" and "privacy policy".

4.1 Amazon.com Shipping and Delivery Case

Amazon.com is a renowned Fortune 500 company based in Seattle started in 1995. The company is the global leader in e-commerce. The dotcom's growth is driven by technological innovations. The company operates worldwide with nearly $15 billion in annual sales in 2007, is one of the iconic companies of the Internet era.

Because of the global character of the enterprise the rules and regulations connected to Shipping & Delivery sphere are of greatest importance to the business. Those rules have to be as simple as possible in order not to distract customers yet they

[3] Web service modeling language. www.wsmo.org. The ontology language used by the WSMF (Web Service Modeling Framework) for semantic web services (SWS). This language is also used as an ontology language by the SUPER platform for SBPM.

also need to be comprehensive and extensive enough to resolve most of common problems on the other hand.

All the related rules are given on Amazon web site[4] and were treated as a main knowledge source in the phase of modeling the ontology for the case[5]. Fig. 2 represents the graphical view of the ontology. We have defined two classes depicting policy Subjects: Clients and Amazon itself. The Provision concept is the central one as it integrates most of others by its attributes of types Action, Subject, Condition and Modality. The Modality concept is derived from the union of AlethicPolicy and DeonticPolicy (introducing the modal operators like obligations, allowances, forbiddances, etc.).

The Condition is a general concept which subsets may be used in formulation of conditional Provisions. The one example of a Condition is the ProductRestriction concept which may be useful in modeling the provision3_WhereDoWeShip as Amazon policy is to ship products anywhere but some types of products cannot be shipped to selected InternationalDestinations.

4.2 MyBlogLog Guidelines Case

MyBlogLog was started in 2004. Since then the portal which renders services of internal web site traffic and content categories popularity has become a part of Yahoo. The site offers the description in the form of "Yahoo! Terms of Service" and "Do's vs. Dont's".

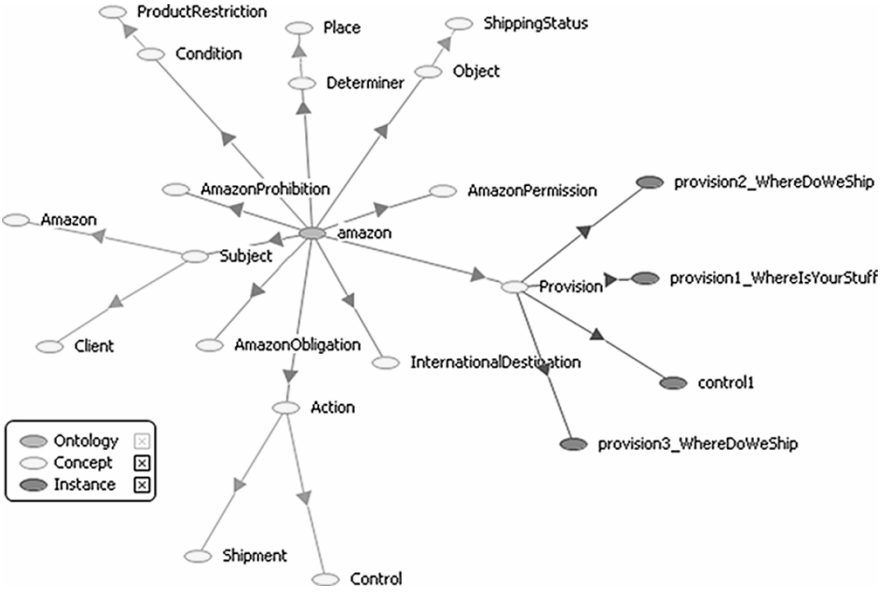


Fig. 2. Amazon case ontology

[4] http://www.amazon.com/gp/help/customer/display.html?ie=UTF8&nodeId=468520
[5] The ontology is available at: http://www.semiramida.info/ontologies/policies/Amazon.wsml

While the business of running this kind of services is immersed in the social aspects of the web and managing the communities gathered round the portal is of greatest importance, the right and easy-to-understand specifications of norms and rules obliged to particular users is the issue of particular attention of owners. As the services of MyBlogLog are widely recognized we assume that the kind of defined norm could be regarded as reference material[6].

In the process of formalizing the rules we availed ourselves of FAQs guidelines comments and terms of service.

In the ontology we defined among others such concepts as:

```
concept MustDo subConceptOf dpo#Obligation
concept CanDo subConceptOf dpo#Permission
concept CannotDo subConceptOf dpo#Prohibition
concept Action
   takesFurtherAction ofType Action
```

The first three are connected to deontic logic per se. Yet, they are defined on the basis of concepts taken from another – upper level – ontology. The Action concept describes any possible deed made by any policy subject.

Any Action may be a MonitoredAction meaning that this is a well specified act described within a given domain characteristic for an Organization. An Action may also be succeeded by further action which is to be taken, so for instance, a User may be warned at first and then banned from the community if necessary.

```
relation User_Can_Do_Action( impliesType User, impliesType UserAction)
relation User_Cannot_Do_Action( impliesType User, impliesType UserAction)
relation User_Can_Upload_Photo( impliesType User, impliesType Action) subRelationOf
User_Can_Do_Action
relationInstance Mauri_Can_Send_Message User_Can_Do_Action (Mauri, SendMessage)
relationInstance Piotr_Can_Join_Community User_Can_Do_Action(Piotr, JoinSurfers)
```

The relation excerpts reveal definitions of three relations responsible for descriptions of expected or forbidden operations done by users. Those relations are then instantiated.

Finally we also define a number of axioms:

```
axiom DontTagMemberWithURL definedBy
!- ?x memberOf User and ?y memberOf TagUser and ?y[hasTagFormat ofType URLTag] and
?y[hasTaggingEntity hasValue ?x].
```

The DontTagMemberWithURL axiom states that this is illegal to put an URL as a tag when tagging users as this is recognized as an inappropriate behavior and is sanctioned with tags removal or further actions as well.

[6] The ontology is available at: http://www.semiramida.info/ontologies/policies/MyBlogLog.wsml

4.3 TechTarget Case

TechTarget is an enterprise that specializes in publishing integrated media that enable IT marketers to reach targeted communities of IT professionals and executives in all phases of the technology decision-making and purchase process. The business won many awards for its innovation and leadership in introducing new technologies for the industry. The important information is that the company host services for more than 1100 prosperous customers with such names as Microsoft, SAP, HP, etc.

The fact of possessing the portfolio of this kind of clients, which are large and leading companies in their markets is crucial for the strict observing their privacy policy with setting the most rigorous rules and standards in this field[7].

In the Fig. 3 the TechTarget case ontology is shown. The creation of this ontology was the most challenging form the presented three cases as the policy corpus is complicated in nature and contains a lot of technical jargon. This led us to the conclusion that modeling these specific policies involves the modeling of domain information, which may be of relevance only to the interpretation of the modeled regulation and the processes making use of or depending on it. We separated all these informational aspects from the concepts relevant for regulation enforcement in a spate ontology module.

A lot of the policy volume is dedicated to information, information typology and related sites or services. This feature of the policy is reflected in the construction of our ontology.

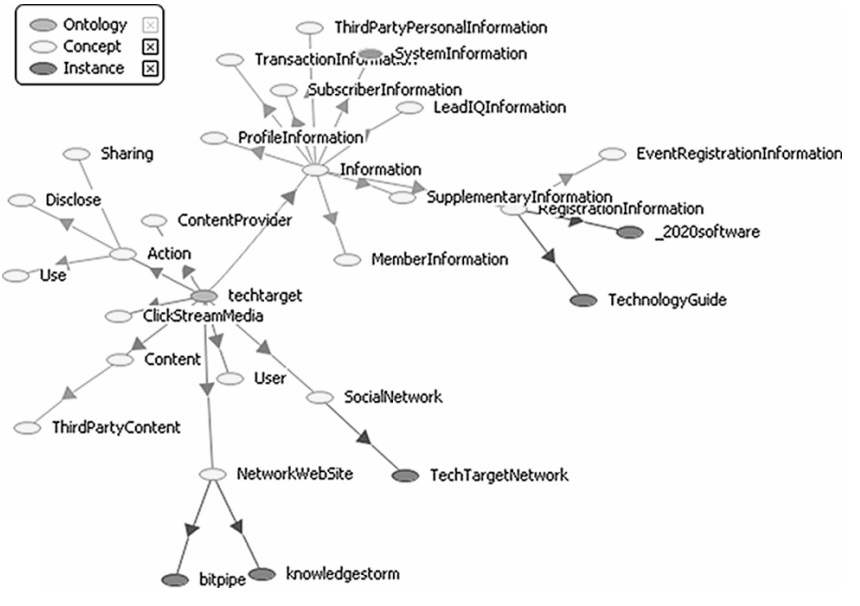


Fig. 3. TechTarget case ontology

[7] The ontology is available at: http://www.semiramida.info/ontologies/policies/TechTarget.wsml

5 A Proposal of a Generic Normative Ontology of Policies for Description of Business Processes Compliance

Fig. 1 represents a simplified preliminary version of the outcome of the procedure of ontology generation as given in Section 3. We assume that the version is preliminary as we would like to extend our studies taking into account a number of related aspects as indicated in the Section 6 as well a greater number of varied regulation cases.

The result generic ontology of policy norms is characterized with the list of features as presented below:

- It only covers User Agreements between corporations providing services and human or legal users. In the current state of work, these are: privacy policies and terms of use.
- It uses a bottom-up approach, using several iterations for building up a generic core ontology for representing legal knowledge. Where each iteration consists of the processing of a given regulation case according to the approach explained in Fig. 1. This approach makes no assumption about the size or complexity of each modeled regulation case.
- It is policy and rule centered, which means that the concepts of policy and rule have been introduced in the ontology although not (always) explicitly mentioned in the modeled regulations. The reason for this is that a background motivation for this work is the integrated modeling of legal aspects of information systems and business processes. Another fact is that an assumption made by this work (based on previous work in [1, 2]) is that policies and business rules can be used for modeling, checking and enforcing actionable and logical aspects of regulations.
- Shows the necessity for a policy and rule infrastructure ontology for implementation of regulation. The concept of policy for modeling decision-making and of rule for action taking fits well with the structure of the modeled cases, as is shown by Fig. 4.

Fig. 4 basically places the concept of agreement at the center of modeled regulations. An agreement is made between several subjects. Three types of subjects have been mentioned by regulations: client, user, thirds party, this of course additionally to the policy owner who defines the regulation. A separate role and organizational ontology such as are built in the context of the SUPER project can be used for further refinement of this aspect [13].

A policy is one kind of an agreement implementation and has a modality which can be deontic (prohibition, permission, obligation). A policy also specifies a condition which is itself modeled using determiners (Time, place, context, etc…). determiners are very vaguely defined in the regulation cases we processed and we estimate that relying on a common upper-level ontology for concepts such as time and place would be very helpful in making the modeling of regulations more generic and enhance interoperability among the latter. A policy is enacted upon an object, which can be any kind of resource, product or information. A policy has usually a further description destined to define more or less precisely how the decision made by the policy can

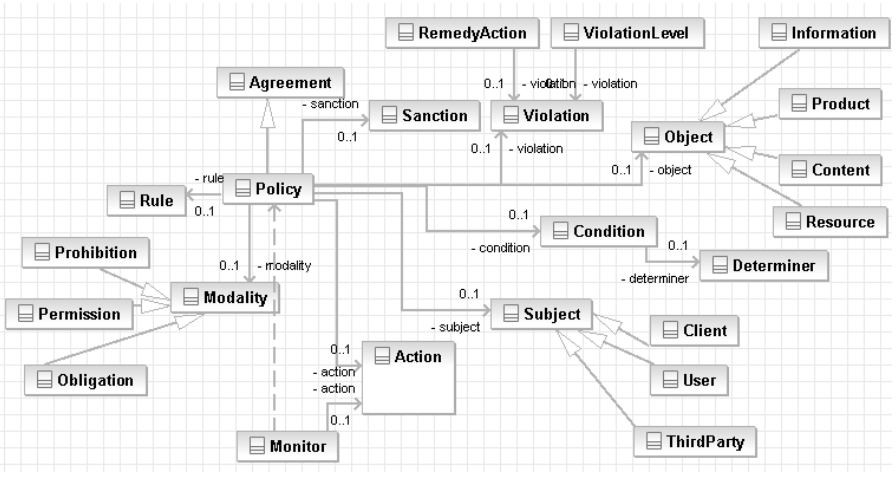


Fig. 4. Reference Generic Regulation Ontology (draft view)

be automated or computed. However, this is not always the case. An example of such rules is given by the MyBlogLog guidelines ontology in Section 4.2.

The coverage of user agreement means that it is hardly possible to model other types of policies which regulate for instance relations between collaborators or third corporate parties.

6 Discussion, Related and Future Work

There are scarcely works in the legal informatics[8] connected directly to the topic of validation frameworks confronting information systems on the one hand and legal regulations on the other. Regardless of this gap we present here few achievements which should be recognized as relevant for future research.

We recognize that the challenge that has most common to do with the compliance control is the normative conflicts resolutions, i.e. approaches adopted in order to resolve conflicts between norms and rules. In [14] authors discuss normative conflicts, their explication and typology, and relate them to the conceptualization of legal knowledge and methods for representing it. Another interesting approach is the differentiation between normative knowledge (legal norms) and world knowledge (abstract description of the world).

Another interesting approach is presented in [15], where the author makes a generalized notion of consistency of norms. Furthermore the contribution is based on a formal model using Rule Logic presented in the text. [16] shows a complete method of legal reasoning with norms derived in the common law environments. The presented approach is useful when facing competing arguments. Those arguments can be

[8] In BPM discipline this problem is much better addressed. For extended descriptions of related works see [1].

harmonized by contextualization with the comprehensive logic theories. The mentioned work can be very productive and influential in further developing of our approach

7 Conclusion

This work discussed and motivated the necessity of providing a generic framework for modeling regulations. We also proposed a bottom-up approach for creating such a framework, and presented the initial ontology resulting from the modeling of three use case regulations in the domain of privacy and term of contract agreements. The goal here is to provide automation by relying on technologies from research on SBPM, in particular the results of the SUPER project on semantic policy modeling. Using this semantic regulation framework, users in communities such as SBPM but also responsible websites can check the validity as well as enforce regulations on the ecosystems governed by the latter.

References

1. El-Kharbili, M., Pulvermüller, E.: A Semantic Framework for Compliance Management in Business Process Management (2009)
2. Kharbili, M.E., Stein, S., Markovic, I., Pulvermüller, E.: Towards Policy-Powered Semantic Enterprise Compliance Management – Discussion Paper. In: CEUR Workshop Proceedings, pp. 16–21 (2008)
3. Despres, S., Szulman, S.: Construction of a Legal Ontology from a European Community Legislative Text, pp. 79–88. IOS Press, Amsterdam (2004)
4. Gangemi, A.: Design patterns for legal ontology construction, `http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-321/paper4.pdf`
5. Visser, P.R.S., Bench-Capon, T.J.M.: Ontologies in the Design of Legal Knowledge Systems (1999), `https://eprints.kfupm.edu.sa/55792/1/55792.pdf`
6. Boer, A., van Engers, T., Winkels, R.: Using ontologies for comparing and harmonizing legislation, pp. 60–69. ACM, New York (2003)
7. Nagy, M., Vargas-Vera, M., Stolarski, P., Motta, E.: DSSim results for OAEI 2008. In: Proceedings of the 3rd International Workshop on Ontology Matching (OM 2008), Karlsruhe, Germany, October 26. CEUR WS, vol. 431 (2008)
8. Bourcier, D., de Rosnay, M.D., Legrand, J.: Methodological perspectives for legal ontologies building: an interdisciplinary experience, pp. 240–241. ACM, New York (2005)
9. Surden, H., Genesereth, M., Logu, B.: Research abstracts 2: Representational complexity in law, pp. 193–194. ACM, New York (2007)
10. Kowalski, R., Sergot, M.: Computer Representation of the Law. In: Proceedings of IJCAI 1985, pp. 1269–1270 (1985)
11. Warnier, M., Brazier, F., Apistola, M., Oskamp, A.: Towards automatic identification of completeness and consistency in digital dossiers, pp. 177–181. ACM, New York (2007)
12. Tomaszewski, T., Stolarski, P.: Legal Framework for eCommerce Tax Analysis. In: Camarinha-Matos, L.M., Picard, W. (eds.) Pervasive Collaborative Networks, IFIP TC 5 WG 5.5 Ninth Working Conference on Virtual Enterprises, Poznan, Poland. Springer, Heidelberg (2008)

13. Filipowska, A., Kaczmarek, M., Starzecka, M., Stolarski, P., Walczak, A.: Semantic Enterprise Description for the Needs of Business Process Automation SemBPM. In: COMPSAC 2008, pp. 987–992 (2008)
14. Elhag, A.A.O., Breuker, J.A., Brouwer, B.W.: On the Formal Analysis of Normative Conflicts. In: van den Herik, H., et al. (eds.) JURIX 1999: The Twelfth Annual Conference, GNI, pp. 35–46 (1999)
15. Hage, J.: Rule consistency. In: van den Herik, H., et al. (eds.) JURIX 1999: The Twelfth Annual Conference, GNI (1999)
16. Bench-Capon, T., Giovann, S.: Using values and theories to resolve disagreement in law. In: JURIX 2000: The 13th Annual Conference, GNI (2000)

Legal Aspects of Functioning of the Social Network Sites

Sylwia Nawrot

Member of Polish Bar Association in Poznań, ul. Marii Konopnickiej 15, 60 – 771 Poznań
syl.nawrot@gmail.com

Abstract. The goal of this paper is to present the legal aspects of functioning of the social network sites. The analysis will be accomplished on the basis of precepts of Polish and European Union law. Firstly, the main legal issues of functioning of the social network sites will be pointed out. Secondly, the special attention to the protection of the personal data issue in functioning of the social network sites will be paid. The theoretical reflections are enriched with case studies from the legal practice of the social network sites.

1 Introduction

The advent and the widespread usage of the Internet, and in particular the Web, have affected almost all of the domains of human existence. The global network has brought a new dimension to the notion of the information society and has determined its development. From a sociological point of view, the Internet / the Web has changed and has defined new standards of communication, interpersonal relations and working of the human societies as a whole.

Nowadays, one of the main forms of the activity in the Web are the social network sites (SNSs). The concept of communication has changed and the SNSs have become a useful tool in this new context of global communication. The SNS can be understood as a web-based service which users construct their own public or semi-public profiles where their personal data and other information related to them are released. The SNSs usually provide a variety of tools (like i.e. instant massages, electronic mailing, photo/video-sharing, blogging etc.) to facilitate its users to interact with each other, to exchange information and to collaborate. The SNSs can be targeted on around a specific social group (i.e. school class, professionals of a certain kind, etc.) or on the exchange of a particular information (i.e. certain interests like music, photography, etc.) [1]. Although these networks vary widely in terms of subject matter, the basic concept remain the same: to provide a channel through which its users can communicate. At the moment, the examples of the most popular SNSs in Europe and in United States are Facebook, MySpace or LinkedIn, and in Poland, Nasza klasa, Grono.net or Fotka.pl.

The importance of the SNSs is best illustrated with the numbers. The last statistics from March, 2008 show that the SNSs are used by 272 million people worldwide which constitute 58% of all users of the Internet and which constitute 21% growth in

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 280–291, 2009.

comparison with the state from February, 2007[1]. Facebook informs about 175 millions of the active users on its site, MySpace about 75 millions, LinkedIn about 35 millions and Nasza klasa about 11 millions[2].

The SNSs are increasingly attracting the attention of the academic and the industry researchers. The phenomenon of the SNSs is in particular the important object of interest to the sociologists in the context of its influence on the shape and on the way of working of the contemporary society. The sociologists, in their research on the SNSs, bring up in particular the issues concerning the identity of the users of the SNSs, their privacy or their social capital.

However, the SNSs are not only the object of research for the sociologists. The characteristic way of functioning of the SNSs is an interesting field of research for law sciences as well.

2 Principal Legal Aspects of Functioning of SNSs

From a legal point of view, there are in particular the following important aspects of functioning of the SNSs: protection of the personality rights, protection of the intellectual property rights and protection of the personal data.

2.1 Protection of Personality Rights

Definition and legal regulation

According to the standpoint presented in the legal doctrine and in the judicial decisions, the catalog of the personality rights does not have a closed character. The different lists of the personality rights it is possible to find in the different legal acts are not exhaustive or completed [2]. From the point of view of functioning of the SNSs, the most important personality rights are: **honor and dignity[3], bodily inviolability and sexual integrity, name or pseudonym, image and correspondence secret**.

In the Polish law system the protection of the personality rights is regulated in particular in the basic act in the Polish legal acts hierarchy – the Constitution of the Republic of Poland dated on April 6, 1997. The Polish Constitution refers to *the inherent and inalienable dignity of the person* as the principal rule of the human and citizen constitutional rights system. This dignity constitute a *source of freedoms and rights of persons and citizens and is inviolable*. Moreover, the Constitution directly provides the *right to legal protection of the private and family life, of the honor and of the good reputation, the right to make decisions about the personal life* and *the freedom and privacy of communication* as well (articles 30, 47 and 49) [6].

Among other Polish regulations on the protection of the personality rights the most important are the norms of the Polish Civil Code according to which *personality rights such as the health, freedom, honor and dignity, freedom of conscience, name or*

[1] *Power to the people social media, Wave 3*, Universal McCann, March 2008, http://www.universalmccann.com/.

[2] Facebook: http://www.facebook.com, MySpace: http://www.myspace.com, LinkedIn: http://www.linkedin.com, Nasza klasa: http://nasza-klasa.pl.

[3] Polish Supreme Court sentence dated on October 8, 1987, number II CR 269/87, published in OSN 1989/4/66.

pseudonym, image, correspondence secret, inviolability of the house, science, artistic or inventive creation, are protected according to the civil law (...)[4].

Independently of the above indicated regulations, the separate legal basis for the image and correspondence secret protection contains the Polish Law on Intellectual Property and the Related Rights Act, dated on February 4, 1994[5]. According to this act, the distribution of the image or of the correspondence generally needs the authorization of the person they belong to.

The protection of the personality rights rules are provided also in the acts of international law which have been ratified by Poland and are applied together with the internal law. These are the following acts:

- the Universal Declaration of Human Rights adopted by the United Nations General Assembly, dated on December 10, 1948[6],
- the International Covenant on Civil and Political Rights adopted by the United Nations General Assembly, dates on December 16, 1966[7],
- the Convention for the Protection of Human Rights and Fundamental Freedoms, adopted under the auspices of the Council of Europe in 1950 in Rome[8].

The similar regulation contains the Charter of Fundamental Rights of the European Union as well[9]. The document was supposed to enter into force upon the signature of all member countries by January 1, 2009. Due to the lack of ratification of the Treaty of Lisbon (and the Charter as one of its parts) by all the member countries of the EU, the Charter does not have the legal force at the moment. Nevertheless this is of little legal consequence, as the law systems of all UE countries are generally similar to each other in respect of the protection of the personality rights. Firstly, because most of the member countries is part of the international acts as above stated. Secondly, the similar regulations are consequence of the law tradition and European heritage in general common to every member country.

The analysis of the above listed legal acts indicate that the terminology used by the legislators as well as the form of formulating the norms are different in different acts. Regardless of that, all of these acts have one common denominator: in all of these acts **it is guaranteed to every person the exclusive right to decide about the fact and the form of use of the personality rights belonging to this person. The rule is that no third person should use in any way the personality rights belonging to the other person** [3]. There are two principal exceptions to this rule:

---

[4] Art. 23 Polish Civil Code dated on April 23, 1964, Dz. U. 64/16/93 dated on May 18, 1964 with amendments.

[5] Polish Law on Intellectual Property Rights and the Related Rights Act dated on February 4, 1994, Dz. U. 06/90/631 with amendments: *„Art. 81.1. The distribution of the image needs the authorization of the person presented."* and *"Art. 82. If the person the correspondence is addressed to, has not expressed other will, the distribution of the correspondence, in twenty years counting from the death of that person, needs the authorization of his/her spouse, and in case of lack of the spouse, the authorization of his/her children, parents or siblings.*

[6] Art. 12 UN Universal Declaration of Human Rights.

[7] Art. 17 UN International Covenant on Civil and Political Rights.

[8] Art. 8 Convention for the Protection of Human Rights and Fundamental Freedoms.

[9] Art. 7 Charter of Fundamental Rights of the European Union.

– the consent of the person the personality rights belong to,
– the special law regulations i.e. the norm according to which it is allowed to publish the name and the image of the persons convicted to certain kind of crimes.

It is important to underline that the consent of the person the personality rights belong to should be clear and detailed, it means it should indicate what personality right and what kind of use it refers to.

In case of infringement of the personality rights, the injured party can sue the other party and claim as follows: the desistance of the illegal actions, the clearance of the consequences of the illegal actions, making the statement in the proper form and content (so called public apology, in most cases in the newspaper or other public and popular means of communication), the compensation, the indemnity or the transfer of the money to the charity[10]. Moreover, according to the international law the injured party may file the complaint to the European Court of Justice with the possibility of demanding the compensation[11]. On the other hand the guarantee to the protection of the personality rights are the penal norms – the Polish Penal Code sanctions the crimes as libel or insult[12] which can be prosecuted by public prosecutor.

SNSs Practice

The SNSs, due to the kind and the quantity of the sensitive personal data distributed on these sites give a particular possibility to its users to take actions that may be qualified as infringement of the personality rights. The important fact in this context is that there is no real verification of the veracity of the data used and published on the SNSs by its users as well as there are usually no moderation systems of the data, the information or the context released on the SNSs by its users. It is important to underline also that the key circumstance in such cases will be the widespread of the SNSs. The SNSs function as the society but in the virtual Web world. The data, the information or the photo released on the SNSs have to be generally considered as announced publicly. There are different situations of the infringement of the personality rights in the practice of the SNSs. The most common examples are the following:

Case 1:

The person A using the SNSs tools (like i.e. instant massages, electronic mailing, blogging etc.) places false, offensive or insulting commentaries related to the person B or comments photos or other information found on the profile of the person B already placed there, such us: "B is a stupid, fat pig" or "B is thief, he has being stealing for years from the company he works for by carrying out products during the night shift", etc.

Case 2:

The person A places in his / her own profile on SNS a data, an information and in particular a photo without the permission of the person B who these data, information or photo are related to (even if they are true). This may occur by placing the photos presenting exclusively the person B and by placing the group photos where the person B appears between other people as well. This may occur in particular by placing the photos showing especially embarrassing or shameful situations for the person B or the situations in which

[10] Art. 24 Polish Civil Code.
[11] Art. 19 Convention for the Protection of Human Rights and Fundamental Freedoms.
[12] Art. 212 and following of the Polish Penal Code dated on June 6, 1997, Dz. U. 97/88/553 dated on September 2, 1997 with amendments.

due to the kind of professional or personal life the person B leads at the moment he or she wouldn't like to be seen in (i.e. during a party, in sexual situations, etc.).

Case 3:
The person A constructs a profile using data such as the name, surname or image of the person B without his / her permission and places a false information about the person B on this profile posing as the person B.

Such the actions like above presented in the cases 1-3 not always but in certain situations can be qualified as the infringements of honor and dignity of the person B, his / her name or pseudonym or his / her image.

There is usually a procedure provided in the rules or regulations of every SNS to be used in the situations like the above stated. According to these procedures, generally as a result of the claim of the injured person and once investigated the case, the SNSs block or remove the illegal content or the profiles constructed illegally. Sometimes there is also a possibility to block or to remove the profile of the user acting illegally[13]. Independently o that there is also a possibility of starting the penal investigation or the civil proceedings as presented above.

2.2 Protection of the Intellectual Property Rights

Definition and Legal Regulation
The protection of the intellectual property rights and the related rights is stipulated in the international law acts as follows:

- World Intellectual Property Organization Copyright Treaty signed in Geneva on December 20, 1996[14],
- World Intellectual Property Organization Copyright Related Rights Treaty signed in Geneva on December 20, 1996[15].

[13] Facebook: www.facebook.com , Nasza klasa: www.nasza-klasa.pl.

[14] World Intellectual Property Organization Copyright Treaty signed in Geneva on 20th of December, 1996, Dz. U. 05/3/12 dated on January 7, 2005. *„Art. 6 Authors of literary and artistic works shall enjoy the exclusive right of authorizing the making available to the public of the original and copies of their works through sale or other transfer of ownership"*, *„Art. 7 Authors (…)shall enjoy the exclusive right of authorizing commercial rental to the public of the originals or copies of their works"* and *„Art. 8 (...) authors of literary and artistic works shall enjoy the exclusive right of authorizing any communication to the public of their works, by wire or wireless means, including the making available to the public of their works in such a way that members of the public may access these works from a place and at a time individually chosen by them."*

[15] World Intellectual Property Organization Copyright Related Rights Treaty signed in Geneva on 20th of December, 1996, Dz. U. 04/41/375 dated on January 7, 2005: *"Art. 5. Independently of a performer's economic rights, and even after the transfer of those rights, the performer shall, as regards his live aural performances or performances fixed in phonograms, have the right to claim to be identified as the performer of his performances, except where omission is dictated by the manner of the use of the performance, and to object to any distortion, mutilation or other modification of his performances that would be prejudicial to his reputation."* and *„Art. 6. Performers shall enjoy the exclusive right of authorizing, as regards their performances: (i) the broadcasting and communication to the public of their unfixed performances except where the performance is already a broadcast performance; and (ii) the fixation of their unfixed performances."*

The similar protection of the intellectual property rights and the related rights provide also the law of the European Union – Directive no 2001/29/WE of the European Parliament and the Council dated on May 22, 2001 on the harmonization of certain aspects of copyright and related rights in the information society[16].

The principal Polish regulation in this field is the Law on Intellectual Property Rights and the Related Rights dated on February 4, 1994[17].

According to the above indicated legal acts, **the intellectual property rights are the rights of the author to the work he created** [13].

According to the law, **the work is every manifestation of the creative activity of a person having elements of originality**, regardless of the form, the way of expressing, the purpose or the value of the work. The works are in particular: photos, musical pieces, films, computer programs, literal pieces, etc. The work is also the artistic interpretation of the other work (so called related rights to the intellectual property rights).

There are **two principal kinds of the intellectual property rights** which are protected by law:

- personal rights: in particular the right to mark the work by the name or pseudonym of the author, the right to inviolability of the content and the form of the work and its reliable using, the right of deciding on the first making public the work to the audience and the right of supervision of the way of using the work;
- material rights: the rights to use and to dispose of the work (i.e. rent, sell, etc.)[18].

In other words, **it is in general exclusively up to the author if he or she wants to make public the work, and if so when and how he or she wants to do it**. Moreover, in the case of making public the work, it is to the author of the work who generally all the remuneration related to such actions belongs.

In case of infringement of the intellectual property rights, the injured party may sue the other party and claim as follows: the desistance of the illegal actions, the giving back the financial profits earned due to the illegal actions, the clearance of the consequences of the illegal actions, making the statement in the proper form and content (public apology), the compensation, the indemnity or the transfer of the money to the charity. Independently of the above stated, the legislator has regulated the penal responsibility as well, with the pecuniary penalty, penalty of restricted liberty or custodial sentence [12].

SNSs Practice

The protection of the intellectual property rights is a very important problem of functioning of the whole Web, not only the SNSs. Nevertheless there are a few aspects of the SNSs that intensify more general debate about the intellectual property rights. These are in particular: lowered barriers to publication on SNSs and the co-existence within the same environment of amateur, semiprofessional and professional users [1]. There are usually no verification or moderation systems both of the users of the SNSs

[16] Directive no 2001/29/WE of the European Parliament and the Council dated on May 22, 2001 on the harmonization of certain aspects of copyright and related rights in the information society, Dz. U. UE L dated on June 22, 2001.

[17] Polish Law on Intellectual Property Rights and the Related Rights dated on February 4, 1994.

[18] According to the Polish Law on Intellectual Property Rights and the Related Rights Act dated on February 4, 1994.

and of the contents released on SNSs by these users. This is why the environment to infringement the rules of the intellectual property rights is very favorable in the SNSs. There are different situations of the infringement of the intellectual property rights or related rights in the practice of the SNSs. The examples may be the following:

Case 4:
The person A publishes on his / her own profile a work which intellectual property rights do not belong to him / her but to the person B while marking or not at the same time this work using the person B name or pseudonym.

Case 5:
The person A publishes on his / her own profile a work which intellectual property rights do not belong to him / her but to the person B and earn money on making available this work to the third persons.

All such the actions like above presented in the cases 4-5 not always but in certain situations can be qualified as an infringement of the intellectual property rights or the related rights.

There is usually a procedure provided in the rules and regulations of every SNSs to be used in the situations like the above stated. According to these procedures generally as a result of the claim of the injured person and once investigated the case, the SNSs block or remove the illegal content. Sometimes there is also a possibility to block or to remove the profile of the user acting illegally[19]. Independently on that there is also a criminal responsibility for such actions or the possibility to start the civil proceedings as stated above.

On the other hand, there is another aspect of the intellectual property rights to the works published by a user in the SNSs (i.e. photos, texts, commentaries, music and all other information shared by its users). The entities running the SNSs are vitally interested in acquire intellectual property rights to such works in the widest way possible. On one hand, they need it to fully administer the SNSs, advertise it, publish information of the site, etc. On the other hand, they could possibly take financial advantage of all of this works and information by making it public, selling, renting, etc.

Case 6:
Facebook has recently changed its terms of service. The old and new terms of service both stated that users give Facebook a license to use content "on or in connection with the Facebook service or the promotion thereof." The new agreement, however, eliminated language saying the license would "automatically expire" if users deleted accounts or removed information. There rose the question if Facebook could exploit the site information to profit. Due to the reaction of its users Facebook has changed again its terms of service to the older version[20].

There are always reservations in the terms of use of the SNSs which regulates the issue of intellectual property rights to the contents published on the SNSs by its users.

---

[19] Facebook: www.facebook.com , Nasza klasa: www.nasza-klasa.pl.

[20] Illinois State University, Daily Vidette: http://media.www.dailyvidette.com/media/storage/paper420/news/2009/02/18/Editorials/Facebook.Terms.Change.Sparks.Debate-3634686.shtml

From the legal point of view such provisions of terms of use of SNSs may be controversial as well especially having into consideration the legally required form and content of such clauses and its validity which is related to it.

3 Protection of the Personal Data

Definition and Legal Regulation

The regulation on the personal data arises as the continuation and the development of the regulation on the personality rights. In many legal systems the personal data is considered as one of the personality rights or a special form of right to privacy [2]. That is why the fundamental regulations on the personality rights such as the Universal Declaration of Human Rights adopted by the United Nations General Assembly, dated on December 10, 1948[21] or the Convention for the Protection of Human Rights and Fundamental Freedoms, adopted under the auspices of the Council of Europe in 1950 in Rome[22] will be applied to the personal data as well. Moreover, the rules for protection of the personal data are provided in the following international and UE legal acts:

- Resolution 45(95): Guidelines for the regulation of computerized personal data files adopted by the United Nations General Assembly, dated on December 14, 1990,
- Recommendation of the Council of Organization for Economic Cooperation and Development (OECD) concerning guidelines governing the protection of privacy and transborder flows of personal data, dated on September 23, 1980,
- Convention of the Council of Europe for the Protection of Individuals with regard to Automatic Processing of Personal Data, dated on January 28, 1981 [7],
- Directive 95/46/EC of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data dated on October 24, 1995,
- Directive 2002/58/EC of the European Parliament and of the Council concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications), dated on July 12, 2002 [8].

The principal Polish regulation on the protection of the personal data consists on the following regulations:

- Art. 51 of the Constitution of the Republic of Poland [6],
- The Personal Data Protection Act dated on August 29, 1997.

Besides there are also other legal acts concerning the protection of the personal data heading towards specific sector like telecommunications, electronic commerce, etc.

Having into considerations all of the above indicated regulations, the personal data may be defined as **any information relating to an identified or identifiable natural person**, when an identifiable person is one who can be identified, directly or

[21] Art. 12 UN Universal Declaration of Human Rights.
[22] Art. 8 Convention for the Protection of Human Rights and Fundamental Freedoms.

indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity[23].

From the legal point of view, the definition of the personal data creates some controversies. Generally it is accepted that not only a few pieces of information together but also just one may constitute the personal data on the condition that is sufficient to identify a person. The Polish law indicates additionally that the information is not sufficient to identify a person if too many costs, time or actions would be needed to identify such a person[24]. More recently, the Data Protection Working Party as an independent advisory body set up under Art. 29 of Directive 95/46/EC, released details guidelines how to interpret the definition of the personal data[25]. Under these guidelines, any information (subjective or objective) where the content, purpose or result can be used to determine the identity of a natural person (by anybody other than the person concerned) should be considered as personal data [9].

Therefore as personal data may be considered the following data types: user/group identity if associated to a natural person's name, device identity that is assigned to a natural person via a contract (i.e. SIM), communication information (i.e. IP address, session details), cryptological belongings (i.e. keys generated by server), billing data (i.e. credit card details) [4]. It is worth to underline that the address IP is generally considered a personal data according to European and Polish legal regulations[26] (with exception of the address IP of the public computer i.e. in the internet cafes). Also the address e-mail may be in many cases considered a personal data having into consideration that in most cases its construction is based on the name and surname of the user. The personal data will be also by all means: name, surname, telephone number and address, especially if combined together[27].

The protection of the personal data means **the obligations of the entities processing the personal data** on the one hand, and **the guarantees for the persons whose data are processed** on the other hand [10].

The obligations of the entities processing the personal data consists on: collecting and processing the personal data only in specific and legal goal and only for a period of time necessary to achieve this goal when the personal data should be deleted or anonymized, protecting the personal data against the illegal access and use of third persons, securing the secret of the processed personal data [5].

The principal guarantee for the persons whose data are processed is the fact that excluding the situations literally stipulated in law, the personal data may be collected and processed only with the clear consent of the person the personal data concern. The other guarantees include in particular: the right to the access to the personal data and the information on it and its form of processing, the right to the modification of

---

[23] Art. 2 a) of the Directive 95/46/EC of the European Parliament and the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Art. 6 of the Polish Law on Intellectual Property Rights and the Related Rights dated on February 4, 1994, Dz. U. 06/90/631 with amendments.

[24] Art 6 of the Personal Data Protection Act dated on August 29, 1997 with amendments.

[25] Art.29 Data Protection Working Party.

[26] Opinion 4/2007 of the Data Protection Working Party created by European Parliament and the European Commission. Similarly: Polish General Inspector on Protection of Personal Data.

[27] Sentence of the European Court of Justice, C-101/2001 dated on November 6, 2003 (Lindqvist), § 24.

the personal data, the right to the withdrawal of the authorization to process the personal data and the right to the opposition against the processing the personal data [5].

In case of infringement of the protection of personal data rights, the injured party can claim his or her rights before administration authority especially created to protect the personal data rights – General Inspector on Protection of Personal Data demanding ceasing the illegal operations or applying the operations necessary to process the personal data legally. The injured party may also fight for his or her rights on the base of personality rights if applicable [11].

SNSs Practice

The regulations on the protection of the personal data have an especially important role in the case of the SNSs. Firstly, because the personal data provided by the majority of the users of SNSs is real. The aim and the idea of existing of the SNSs which is equal to find and to communicate with other people, makes it necessary to provide the real personal data. Secondly, because of the coverage of the SNSs which administer at moment with the personal data of millions of people (in this level the SNSs can probably be compared only with the entities providing telecommunication services).

The issues concerning the functioning of the SNSs are numerous and very different: starting from the technical problems to assure the proper protection of the personal data and ending with the problems of providing the users with the possibility of realization of all of the rights they are guaranteed by law. The latest issues concerning the protection of the personal data in practice of SNSs include the indexation of the personal data of the SNSs users and the time of anonymization of the processed personal data.

Case 7:

Person A has a profile on one of the SNSs where he / she release the personal data, including information and photos. The profile of the person A (like the other profiles of the users of this SNS) is free to be indexed by the search engines like Google or Yahoo. As a result anyone using the search engines like Google or Yahoo will receive the information of the person A, including its personal data, other information or photos.

From a legal point of view, such operations may be in some situations questionable. The European and Polish administration organ assumed recently that such operations are generally legal on the condition that the authorization of processing the personal data of the person whose personal data will be indexed include clearly such operations of indexing[28]. It is worth to say that the SNSs situated in United States have been allowing indexing by search engines for a few years now[29].

Case 8:

Person A create a profile on one of the SNSs using false personal data and uses the created profile to spread the pedophile photos on the SNS. Then the person A removes the profile and all of the data and photos it contains claiming the deleting or anonymization of his or her personal data. The question is if the entity running the SNS should satisfy such request regardless of anything and immediately or should retain the personal data of the person A for the needs of i.e. prosecution of the crime or claims of the injured persons.

---

[28] Opinion 1/2008 of the Data Protection Working Party created by European Parliament and the European Commission.

[29] SearchViews: http://www.searchviews.com/index.php/archives/2007/09/facebook-opens-to-search-indexing.php.

From a legal point of view, on the one hand there is an obligation of the entity running the SNSs to delete the personal data on the claim of the person whose this data belong to. On the other hand, there is a risk of losing the personal data of a person who has committed a crime with probably no other possibilities to localize this person. There is no legal regulation in relation to this issue by now. Nevertheless, a similar act has been adopted in relation to the sector of telecommunications where it is generally allowed to retain the personal data in particular in such cases for a period of not less than 6 months and not more than 2 years[30]. There are voices to regulate these issues in regard to the SNSs and the whole Web.

4 Conclusions

As described above, there are a few serious legal aspects that are important in the functioning of the SNSs. From a legal point of view, the new phenomenon of the SNSs has become a new mean which has changed the way of understanding the issues of protection of the personality rights, protection of the intellectual rights and protection of the personal data.

Independently the above stated, it is necessary to underline that the SNSs, similarly to the whole Web, have become the mean used to commit the so called common crimes. Due to the amount and kind of information introduced in the SNSs by its users, the SNSs are especially important and dangerous mean used by the criminalists in i.e. such a crimes as the crimes against sexual integrity, including pedophilia and the crimes against material goods.

Case 9:
The person A using the profile created on the SNSs release the photos containing the pornographic contents or pedophile contents.

Such actions are qualified as the common crimes and are prosecuted. There have been many examples of such crimes registered by the SNSs, like Facebook[31] which informs about hundreds of profiles deleted due to this problem[32] or Nasza klasa[33].

There is one more thing to make clear in respect to the analyzed issues related to the protection of the personality rights, the protection of the intellectual rights and the protection of the personal data on SNSs. The entities running SNSs as the electronic service providers as far as they do not moderate the contents released by its users on the site, they are generally not responsible for the eventual infringements of the third persons rights[34]. The entities running the SNSs are exclusively responsible for the fulfilling the obligations on protecting the personal data of its users.

---

[30] Directive 2006/24/EC of the European Parliament and of the Council on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC dated on March 15, 2006.

[31] Time U.S.: http://www.time.com/time/nation/article/0,8599,1877276,00.html.

[32] Facebook: http://www.facebook.com.

[33] Nasza klasa: http://nasza-klasa.pl.

[34] Art. 12 o the Polish Providing Electronic Services Act dated on September 9, 2002: *"The person who transmits the data is not responsible for the transmitted contents if: 1) this person is not the initiator of the transmission, 2) this person does not choose the recipient of the transmission and 3) this person does not delete or moderate the contents he or she transmits."*

References

1. Humphreys, S.: The challenges of intellectual property for users of social networking sites: a case study of Ravelry. In: MindTrek 2008: Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era, pp. 125–130. ACM, New York (2008)
2. Safjan, M.: System Prawa Prywatnego, Prawo cywilne – część ogólna, t. 1. C.H. Beck, Warsaw (2007)
3. Gniewek, E.: Kodeks cywilny. Komentarz do artykułów 1-534. C.H. Beck, Warsaw (2004)
4. Becker, A., Arnab, A., Serra, M.: Assessing Privacy Criteria for DRM using EU Privacy Legislation. In: DR 2008 (2008)
5. Barta, J., Fajgielski, P., Markiewicz, R.: Ochrona danych osobowych. In: Komentarz. Wolters Kluwer Business, Cracovia (2007)
6. The Constitution of the Republic of Poland, April 2 (1997), http://www.sejm.gov.pl/prawo/konst/angielski/kon1.htm
7. Carey, P.: Data Protection. In: A practical guide to UK and EU law. Oxford University Press, Oxford (2004)
8. Korff, D.: Data protection laws in the European Union. In: FEDMA (2005)
9. Jay, A., Hamilton, R.: Data Protection Law and Practice. Sweet & Maxwell, London (2003)
10. Korff, D.: EC Study on Implementation of Data Protection Directive - Comparative Summary of National Laws, Cambridge (2002)
11. Smith, G.J.H.: Internet law and Regulation. Sweet & Maxwell, London (2002)
12. Keeling, D.T.: Intellectual Property Rights in EU Law: Free Movement and Competition Law, vol. 1. Oxford University Press, Oxford (2004)
13. Phillips, J.: Butterworths Intellectual Property Law Handbook. Butterworths Law (2007)

A Support System for the Analysis and the Management of Complex Ruling Documents

Marco Bianchi[1], Mauro Draoli[1],
Giorgio Gambosi[3,4], and Giovanni Stilo[2,4]

[1] Italian National Centre for ICT in the Public Administrations (CNIPA), Rome, Italy
marco.bianchi@cnipa.it, draoli@cnipa.it
[2] Dept. of Computer Science, University of L'Aquila, L'Aquila, Italy
g.stilo@gmail.com
[3] Dept. of Mathematics , University of Rome "Tor Vergata", Rome, Italy
gambosi@mat.uniroma2.it
[4] NESTOR - Laboratory, University of Rome "Tor Vergata", Rome, Italy

Abstract. This paper reports the experiment conducted for the development and assessment of a new software tool allowing the automatic discovery of correlations in large legislative frameworks.

The system, named *NavigaNorme*, has been mainly designed to support experts of the Legal domain involved in the simplification of the Italian normative framework for all levels of the Public Administration. In fact, the most relevant functionality in *NavigaNorme* is the identification, given a paragraph in a selected norm, of those paragraphs (in the same norm and in other norms) that should be considered in trying to reduce the number of norms being in force or in drafting a new law.

NavigaNorme relies on a search engine that combines classical Information retrieval techniques with some ad-hoc strategies introduced to increase the precision of the retrieval by exploiting implicit information extracted from the logical structure of Legal and normative texts.

The effectiveness of NavigaNorme has been mainly measured in terms of precision, through an assessment procedure that involved experts of the Legal domain.

Keywords: Normative texts analysis, Information retrieval, Evaluation of search engines.

1 Introduction

Currently, activities related to law and legal issues involve a huge group of professionals, and amount to a multi-billion overall value. People involved in such activities have to deal with a challenging amount of text, information and knowledge, contained in thousands of documents. Especially in continental Civil Law countries the system of the laws is particularly complex: the Italian system of laws, in particular, is composed by more than one hundred thousand different norms.

Thousands of civil servants and legal professionals need every day to access and analyse the existing norms, also in order to propose updates and changes. For all them,

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 292–303, 2009.

dealing with information and knowledge is an essential part of their daily work. They are "knowledge workers" and vulnerable to suffering from information overload. Legal professionals and civil servants need to reduce the significant amount of their time spent to finding, reading, analysing and synthesizing information in order to take decisions, and prepare advice and trials.

Currently, in Italy there is a particular interest in the simplification and rationalization (e.g. merging and updating) of this huge amounts of norms. Information technology is expected to be essential in supporting professionals in the legal domain to afford both the challenge of large simplification and the daily management of a corpus of norms.

CNIPA is the Italian National Center for Informatics in Public Administration, devoted, among others, to supporting the Italian Administration in using ICT in an effective way. As a consequence, the introduction of ICT tools which provide support for the access and possible reorganization a collections of norms is of immediate interest for CNIPA itself.

In this paper, we describe a software to support legal professionals in exploring a complex corpus of norms through the automatic detection of relations between different paragraphs of norms and the navigation of "paths" among them.

The developed tool is named *NavigaNorme*. In this paper we present the main technical characteristics of the tool and, in a detailed way, the results of the experimentation of the system on a selection of 120 norms dealing with a specific sub-domain. The experimentation relies both on a statistical evaluation of the quality of the results and on the subjective evaluation given by the users in terms of effectiveness and precision.

The paper is organized as follows: Section 2 presents some related work. Section 3 contains an brief description of the documents collection used during the experimentation; Section 4 describes the main functionalities of NavigaNorme and focuses on the original aspect of this application. Section 5 presents some implementation details. Section 6 reports the results of the experimentation activity aimed to assess the effectiveness of NavigaNorme application. Finally, and Section 7 draws conclusion and describe some future activities.

2 Related Work

NavigaNorme has been mainly designed for experts of the Legal domain involved in the simplification of normative frameworks. Nowadays experts are supported by searching systems forged when users used to query databases by means of Boolean logic. For this reason search platforms for the Legal domain mainly offer just Boolean search functionalities. It means that, given a set of keywords related by Boolean logic operators (e.g. AND, OR, NOT), these systems return exactly all laws satisfying the specified Boolean expression.

In Italy, *Leggi d'Italia* [13] is the most large database of Italian laws: it offers more than 68.000 laws; *DeJure* [14] offers the widest collection about national, regional, and community legislation, including international instruments ratified, and practice. *Norme In Rete* (NiR) [2] is a public project jointly sponsored by CNIPA and the Italian Ministry of Justice: NiR is a search engine allowing the access to information published on Web sites of the Italian public administrations. All these systems mainly offer search strategies based on the Boolean model [10].

To the best of our knowledge also all most relevant on-line legal search services for lawyers and legal professionals in the United States (e.g. WestLaw [15], LexisNexis [16], Quicklaw [17]) offer search strategies based on the Boolean model.

With respect to the above mentioned search services, *NavigaNorme* allows experts of the Legal domain to identify correlations between paragraphs of normative texts. This is a very innovative functionality not yet provided by existing search services.

An interesting and widely studied approach to the problem of searching and extracting the information needed by the used from a wide set of documents is the one based on semantic search engines, where resources (documents, section, etc.) are semantically annotated with respect to a given domain-dependant suitably rich data model (an ontology): this is expected to provide higher precision and recall to search operations.

Since the development of semantic search engines in the Legal domain is based on the definition of ontologies modelling such domain, many research papers presents ontology models [18]. Anyway, once an ontology has been chosen, all documents of the collection have to be semantically annotated: this is a very time consuming operation, requiring a significant human support. Furthermore, to the best of our knowledge there are not semantic search engines really used in the Legal domain.

On the contrast NavigaNorme is built on a statistical search engine. As a consequence the system is highly scalable, since adding a large number of laws and normative texts to the current collection requires a negligible effort. We believe that the efficacy of *NavigaNorme* proves that nowadays systems based on statistical search engines can fill up the temporal gap until semantic search engines will be adopted.

3 Corpus Description

In our work we focus our interest on a document collection consisting of 120 Italian *normative texts*, simply *norms* in the following, regulating the ICT domain. These documents are structured in XML format in accordance with the specifications issued from CNIPA as a deliverable of the *Nome in Rete* project [1,2]. Each document also contains a large number of meta-data and tagged information, such as kind of rule, history of modifications, etc. More precisely, we consider only those tags describing the structure of the document and tags identifying references across norms. For the sake of simplicity we limit our attention only to the information most likely useful to increase the quality of the retrieval: below, we provide a simple fragment of an XML (simplified) document, as used during our test.

```
<?xml version="1.0" encoding="ISO-8859-15">
<document>
<fileinfo>
<hashcode>211</hashcode>
<intestazione xmlns="http://www.normeinrete.it/nir/2.2/">
<tipoDoc>LEGGE </tipoDoc>
<dataDoc norm="19900807">7 agosto 1990</dataDoc>, n.
<numDoc>241</numDoc>
<titoloDoc>Nuove norme in materia di procedimento
amministrativo e di diritto di accesso ai documenti
```

```
amministrativi.
</titoloDoc>
</intestazione>
</fileinfo>
<doc>
<docid>211-212</docid>
<comma xmlns="http://www.normeinrete.it/nir/2.2/" id="art1-com1">
<num>1.</num>
<corpo> L'attivita' amministrativa persegue i fini determinati
dalla legge ed e' retta da criteri di economicita', di efficacia,
di pubblicia' e di trasparenza secondo le modalita' previste
dalla presente legge e dalle altre disposizioni che disciplinano
singoli procedimenti,  nonche' dai principi dell'ordinamento
comunitario.
</corpo>
</comma>
</doc>
...
</document>
```

Each *norm* is hierarchically structured in *sections*, *paragraphs*, and optionally in *sub-paragraphs*.

Since a paragraph can reference another norm or a part of it, we can classify references as follows:

1. paragraph to sub-paragraph reference;
2. paragraph to paragraph reference;
3. paragraph to section reference;
4. paragraph to normative text reference;

Every reference can point to an element belonging to the same norm (*in-norm-reference*) or to a different one (*cross-norm-reference*). References can also point to elements belonging to norms in the collection or to norms outside of it. This last classification can also be considered as a metric to evaluate the completeness of the corpus itself.

Given the above mentioned definitions, our collection contains 8368 paragraphs and 4608 references (roughly 60% of them pointing to norms outside the collection).

Table 1 and Table 2 report the results of the analysis conducted on in-norm and cross-norm references, respectively.

Table 1. In-norm-reference analysis

Reference type		
Paragraph to sub-paragraph	69	6%
Paragraph to paragraph	668	60%
Paragraph to section	383	34%
Total	1120	

Table 2. Cross-norm-reference analysis

Cross-Reference		
paragraph-sub-paragraph	140	4%
paragraph-paragraph	889	25%
paragraph-section	1083	31%
paragraph-law	1376	39%
Total	3488	

It is worth to underline that these data are derived starting from information extracted by parsing the XML documents. During the analysis of the corpus, we noticed some imprecisions at the annotation level: more precisely, some (logical) references occurring in the text of the norm are not explicitly declared through some XML reference. As a consequence, these data represent a lower-bound for the actual references.

4 Functionality

NavigaNorme is a platform allowing specialists of the Legal domain to identify correlations between paragraphs of normative texts. More precisely, starting from an *input paragraph* selected by the user, *NavigaNorme* returns a list of *related paragraphs* sorted by score: the greater the score value the stronger the expected correlation between corresponding related paragraph and the one provided as input.

Given an input paragraph, *NavigaNorme* can assign scores to related paragraphs on the basis of the following strategies:

- *text similarity* - the score of a paragraph is evaluated on the basis of the text similarity between its content and the content of the input paragraph;
- *in-references* - the score of a paragraph is evaluated on the basis of the presence, in the input paragraph, of references to it;
- *out-references* - the score of a paragraph is evaluated on the basis of the presence, within its content, of references to the input paragraph.

These strategies can be easily set by users on the basis of their information needs. In fact users can enable, disable, or tune a strategy just setting an associated *weight*. Strategies can also be used in combination.

By default only the text similarity strategy is turned-on. In this case the final score is assigned by the search engine used by NavigaNorme (see Section 5 for more details).

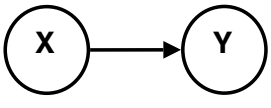
The weight of the other two strategies affects the final score according to the following rules:

- if a related paragraph contains a reference to the input paragraph, its final score is computed adding the *in-reference weight* (specified by the user) to the score automatically assigned by the underlying search engine on the basis of the text

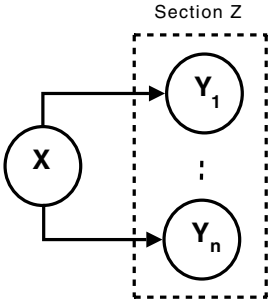
similarity. Notice that negative values of in-reference weight penalize affected paragraphs in the final ranking.
- if the input paragraph contains a reference to a related paragraph, the final score of the latter is computed adding the *out-reference weight* (specified by the user) to the score automatically assigned by the underlying search engine on the basis of the text similarity. Again negative values of out-reference weight penalize affected paragraphs in the final ranking.

The NavigaNorme prototype is mainly characterized by two original working principles.

- The degree of statistical similarity among paragraphs is computed by mean of the DPH [3] retrieval model. Such a model, belonging to the family of probabilistic models for information retrieval based on measuring the divergence from randomness [4], is *parameter free*. This implies that the model has good performance independently of the collection properties, e.g. average length of documents. On other hands, parameter dependent retrieval models can also have better performances, but they need a proper tuning of their parameters, whcih can result in a difficult task. With respect to our work, the choice of DPH as retrieval model is particularly favourable because, in absence of a benchmark, it is not possible to deal with the tuning issue. Furthermore we expect that, when a benchmark shall be available, all results presented in this paper could be boosted adopting (and properly tuning) a parameter dependent retrieval model.
- Another original working principle of NavigaNorme regards its exploitation of the information provided by references across paragraphs in order to improve the retrieval effectiveness. Strategies like this one are widely used in the Information retrieval community: PageRank [5] and HITS [6] are probably the most well-known algorithms of this family, widely applied in the framework of web search. Nevertheless, at the best of our knowledge, nobody has yet tried to apply this kind of strategies to a normative corpus.
  Because of this, starting from the taxonomy of references defined in Section 3, we define the *paragraphs network* (a direct graph) as follows:
  1. Each node of the network represents a single paragraph.
  2. If a paragraph X contains a reference to a paragraph Y, then there exists a direct arc from the node representing X to the one representing Y. Graphically:



  3. if a paragraph X contains a reference to a section Z composed by n paragraphs, $Y_1, ..., Y_n$, then there exist n direct arcs from the node representing X to the nodes representing $Y_1, ..., Y_n$. Graphically:



4. Duplicated references are not considered.

The network of paragraphs built applying the above mentioned rules contains 8368 nodes, 3526 arcs and 5658 (71%) disconnected nodes.

5 Implementation Details

NavigaNorme is built on the Terrier framework [7]. Terrier is an open-source search engine readily deployable on large-scale collections of documents. Furthermore, Terrier implements state-of-the-art indexing and retrieval functionalities and provides a platform for the rapid development of large-scale retrieval applications.

From the Information retrieval point of view, all paragraphs are indexed as different document-unit. Furthermore, since *NavigaNorme* has to retrieve *correlations* between paragraphs, the full text of the input paragraph is submitted to the Terrier as input query on the system in order to retrieve a list of related paragraphs.

During the indexing processing we use the Snowball stemmer [8] for the Italian language and we do not apply any stop-words list. As already mentioned, given an input paragraph, the final score assigned by the default strategy of *NavigaNorme* to each correlated paragraph is equal to the score returned by Terrier.

The in-reference and out-reference strategies are developed as post-processing operations; more precisely, the score, originally computed by Terrier, is modified according to the weights assigned by the user to that strategies.

NavigaNorme is developed in Java and is released as a 3-tiers Web application. The presentation layer is implemented using dynamic web pages (JSP), whereas the persistence layer is encapsulated by the Terrier API. The network of references is stored in a graph data structure managed by means of the Java Universal Network/Graph Framework (JUNG) library [9].

6 Experiments

The main goals of our experimentation were:

1. to establish a range for weights of in-reference and out-reference strategies;
2. to evaluate of the effectiveness of *NavigaNorme*.

In order to establish a range for weights of in-reference and out-reference strategies, we have run a first session of tests, with the aim to verify how these strategies affect the default answer set. More precisely, we defined a procedure to measure differences

in the answer set deriving from a change of the weight associated to the strategy under testing. Both in-reference and out-reference strategies cause variations in the answer set with weights less than 100. Because of this, we establish that their weights should be real values in the interval [0...100], where 0 represents the turniing-off of the strategy .

Regarding the evaluation of the effectiveness of *NavigaNorme*, we noticed that the concept of efficacy should be expressed in terms of how helpful for the legal professional is the set of relations computed by our system. Such an evaluation is usually based on a test reference collection and on an evaluation measure. The test reference collection consists of a corpus of documents, a set of information requests, and a set of relevant documents (provided by specialists) for each information request. The evaluation measure quantifies (for each information request) the *similarity* between the answer set and the set of relevant documents provided by the specialists. This provides an estimation of the quality of the strategy implemented by the system under evaluation [10].

Therefore, given the collection introduced in Section 3, in order to measure the effectiveness of NavigaNorme we need:

- *a statistically significant number of paragraphs to test NavigaNorme*. In fact, since the system effectiveness is known to vary widely across paragraphs, the greater the number of paragraphs used in the experiment the more confident we can be in our conclusions [11]. Since the evaluation process is also a time consuming activity, we selected 20 paragraphs to evaluate the system effectiveness. The list of selected paragraphs is reported in Appendix (Table 5).
- *a team of specialists of the Legal domain*. For this reason we involved in the assessment process three master degree students, each one having a degree in law, and the head of the Office for Legislative Studies of CNIPA.

NavigaNorme has been evaluated on the basis of two different configurations:

- Configuration 1:
  - Out-references Strategy: 0.0
  - In-references Strategy: 1.0
- Configuration 2:
  - Out-references Strategy: 100.0
  - In-references Strategy: 50.0

The first configuration, with all strategies turned off, represents our baseline. It is necessary to measure the gain, in terms of effectiveness, obtained adopting the second configuration in which both in and out-references strategies are turned on.

For each paragraph in Table 5 assessors evaluated the first 40 results of the answer set computed by *NavigaNorme*.

Regarding the evaluation measurement, *precision* and *recall* are the two most basic and frequent measures for effectiveness of systems such as *NavigaNorme* [12].

In this paper we define precision (P) as the ratio between the number of correlations retrieved by the system that are judged valid by the majority of the assessors (relevant correlations), and the total number of correlations retrieved by NavigaNorme (retrieved correlations), that is

$$P = \frac{|\{relevantC\} \cap \{retrievedC\}|}{|\{retrievedC\}|}$$

P can be computed considering different levels of depth. More precisely, $P@K$ denotes the precision computed considering the K most highly scored correlations retrieved .

During the assessment we did not consider the second metric mentioned above, i.e. recall. In fact the recall measure differs from precision by considering, at the denominator, the total number of correlations in the collection instead of the total number of correlations retrieved. Consequently, the recall values can be computed only by having the exact number of correlations in the collection (or, more realistically, an estimation of it) for each paragraph in Table 5. Obviously, these numbers are unknown and we have no elements to estimate them.

Table 3 reports the average precision values evaluated for Configuration 1 and 2 at different levels of depth (i.e. $K = 5$, $K = 10$, $K = 20$, $K = 40$).

Table 3. Average precision values for baseline (i.e. Configuration n.1) and Configuration n.2. The third column highlights the percentage of improvement obtained using Configuration 2 with respect to the baseline.

Precision	Baseline	Conf.2	Improvement
P@5	0.89	0.91	+20%
P@10	0.77	0.81	+17%
P@20	0.66	0.68	+6%
P@40	0.55	0.56	+2%

Table 4. Precision values for configuration 2

ID	P@5	P@10	P@20	P@30	P@40
01	1	0.90	0.85	0.83	0.78
02	1	1	0.85	0.70	0.55
03	1	0.70	0.55	0.37	0.35
04	0.40	0.30	0.20	0.20	0.18
05	0.80	0.50	0.30	0.20	0.15
06	1	0.70	0.35	0.27	0.20
07	1	0.80	0.65	0.47	0.40
08	1	1	0.75	0.73	0.65
09	0.80	0.60	0.45	0.37	0.33
10	1	0.90	0.75	0.70	0.65
11	1	0.90	0.80	0.78	0.75
12	1	0.93	0.90	0.90	0.81
13	0.60	0.60	0.45	0.37	0.30
14	1	0.60	0.40	0.30	0.25
15	1	1	0.85	0.83	0.83
16	0.80	0.80	0.63	0.60	0.58
17	1	1	1	0.90	0.80
18	1	0.90	0.90	0.88	0.87
19	1	1	1	0.93	0.90
20	1	1	1	0.90	0.80

Table 4 provides the precision values of the answer set computed using configuration 2.

It is worth to notice that, with respect to configuration 1, 439 out of the 800 paragraphs retrieved have been judged relevant.

The assessment procedure shows that:

1. An information retrieval approach can be useful to discover relationships among paragraphs. In fact the retrieval performance obtained by the baseline have been considered outstanding by specialists of the Legal domain;
2. Strategies exploiting in and out-references across paragraphs can improve the effectiveness of the statistical retrieval.

7 Conclusions

The subjective evaluation of the system, conducted with four domain experts, reveals that *NavigaNorme* has a precision of more than 55% when computed on the first 40 relations detected for each topic, and of more than 90% when derived for the first 5 relations. This performance has been judged as "very satisfying" by all the experts involved in the research. The ad hoc strategy implemented by the tool has been demonstrated to improve the performance with respect to the baseline. This is particularly encouraging, especially in consideration of the fact that we are considering relations among specific paragraphs of the norms, and not between whole norms. This has been considered by the experts as a "killer" feature, since it allows to immediately find the specific sections of interest in a law.

The implementation of the system has been reasonably quick (whilst the subjective evaluation required a significant human effort). The key points for the success of the project can be summarised as follows:

- the adoption of information retrieval techniques, which significantly reduce the efforts with respect to other widely applied approaches like the semantic one;
- the availability of fine grained structured XML documents reduced development time and made it easier to implement a strategy based on paragraph to paragraph relations;
- the participation of domain experts to the project greatly contributed to design the features of the tool with the purpose of meeting the real expectations of final users.

Let us remind that *NavigaNorme* is a sort of modular framework that allows the experimentation and the evaluation of new search strategies with as less effort as possible. Currently, we are implementing both an ad hoc strategy based on the knowledge of relevant dictionaries of terms, and a more sophisticated strategy based on the semantic annotation of the relations between paragraphs.

NavigaNorme was also developed to assist the jurist during the *drafting* phase of a new law. In fact it is possible to submit the new law to the system to foresee the possible relations with other laws.

References

1. CNIPA: Formato per la rappresentazione elettronica dei provvedimenti normativi tramite il linguaggio di marcatura XML. TR: AIPA-CR-40 (2002)
2. Norme In Rete, http://www.normeinrete.it/
3. Amati, G.: Frequentist and Bayesian Approach to Information Retrieval. In: Lalmas, M., MacFarlane, A., Rüger, S.M., Tombros, A., Tsikrika, T., Yavlinsky, A. (eds.) ECIR 2006. LNCS, vol. 3936, pp. 13–24. Springer, Heidelberg (2006)
4. Amati, G., van Rijsbergen, C.J.: Probabilistic models of information retrieval based on measuring the divergence from randomness. ACM Trans. Inf. 20(4), 357–389 (2002)
5. Brin, S.: The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks and ISDN Systems, 107–117 (1998)
6. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. Journal of the ACM 46(5), 604–632 (1999)
7. Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C., Lioma, C.: Terrier - A High Performance and Scalable Information Retrieval Platform. In: ACM SIGIR 2006 Workshop on Open Source Information Retrieval (OSIR 2006), Seattle, Washington, USA (2006)
8. Stemmer Snowball, http://snowball.tartarus.org
9. The JUNG Web site, http://jung.sourceforge.net
10. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. Addison-Wesley, Reading (1999)
11. Voorhees, E.M.: The Philosophy of Information Retrieval Evaluation. In: Peters, C., Braschler, M., Gonzalo, J., Kluck, M. (eds.) CLEF 2001. LNCS, vol. 2406, pp. 355–370. Springer, Heidelberg (2002)
12. van Rijsbergen, C.J.: Information Retrieval. Butterworth (1979)
13. Leggi dltalia Professionale, http://www.leggiditaliaprofessionale.it
14. Dejure - Giuffr, http://dejure.giuffre.it
15. WestLaw, http://www.westlaw.com
16. LexisNexis, http://www.lexisnexis.com/
17. Quicklaw, http://www.quicklaw.com/
18. Benjamins, R., Casanovas, P., Gangemi, A., Breuker, J.: Law and the Semantic Web - Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications. Springer, Heidelberg (2005)

APPENDIX

Table 5. List of paragraphs used to measure the effectiveness of NavigaNorme. These paragraphs have been selected from two legislative measures of particular interest for CNIPA (i.e L.n. 241/1990 and d.lgs.n.82 7/3/2005).

ID	Legislative measures reference
01	L.n. 241/1990, art. 3-bis, comma 1
02	L.n. 241/1990, art. 4, comma 1
03	L.n. 241/1990, art. 5, comma 1
04	L.n. 241/1990, art. 7, comma 1
05	L.n. 241/1990, art. 8, comma 3
06	L.n. 241/1990, art. 9, comma 1
07	L.n. 241/1990, art. 10, comma 1
08	L.n. 241/1990, art. 14, comma 5
09	L.n. 241/1990, art. 21-bis, comma 1
10	L.n. 241/1990, art. 25, comma 1
11	d.lgs.n.82 7/3/2005, art. 2, comma 1
12	d.lgs.n.82 7/3/2005, art. 7, comma 1
13	d.lgs.n.82 7/3/2005, art. 12, comma 1-bis
14	d.lgs.n.82 7/3/2005, art. 13, comma 1
15	d.lgs.n.82 7/3/2005, art. 15, comma 2
16	d.lgs.n.82 7/3/2005, art. 41, comma 3
17	d.lgs.n.82 7/3/2005, art. 42, comma 1
18	d.lgs.n.82 7/3/2005, art. 52, comma 1
19	d.lgs.n.82 7/3/2005, art. 54, comma 1
20	d.lgs.n.82 7/3/2005, art. 65, comma 1

Legal Advisory System for the Agricultural Tax Law

Tomasz Zurek[1] and Emil Kruk[2]

[1] Institute of Computer Science, Maria Curie-Sklodowska University,
Plac Marii Curie-Sklodowskiej 5, 20-031 Lublin, Poland
zurek@kft.umcs.lublin.pl
[2] Institute of Administration and Public Law, Maria Curie-Sklodowska University,
Plac Marii Curie-Sklodowskiej 5, 20-031 Lublin, Poland
emil.kruk@wp.pl

Abstract. The authors of this study attempted to develop an advisory tool functioning in the scope of the Agricultural Tax Act. The focus of the authors in this study was on presenting the outcome of the efforts connected with building the ontology which would allow for representing individual cases, and dealing with cases not expressly regulated by law. This study will also outline the structure and concept of the system in question.

Keywords: Legal Expert System.

1 Introduction

Legal interpretation is extremely difficult to automate. Years of hard work aimed to create efficient systems supporting legal interpretation have hardly brought this domain out of research laboratories into the common legal practice. The reasons why this proves so difficult are varied, including a great deal of commonsense coming at play, lack of precision and ambiguity of legal provisions, necessity to take the context of a given situation and the aim of the legislator into account, choice of the line of interpretation, and the like.

The authors of this study attempted to develop an advisory tool functioning in the scope of the Agricultural Tax Act [11]. The choice of this Act was inspired by its specificity. The authors' primary emphasis was on the legal act being as deterministic as possible, as it would allow for considerably restricting the interpretation leeway which in the case of other legal acts is very wide. Another reason behind this choice stemmed from the fact that fiscal law calls for grammatical interpretation. Legal acts of this kind significantly facilitate the development of advisory systems, reducing, though not fully eliminating, the impact of interpretation difficulties.

The authors seek to create a tool which would provide the agricultural tax payers and officers with comprehensive advice in the scope of their rights and obligations. The Agricultural Tax Act governs such issues as tax calculation, tax rates, classification of taxpayers and farm land under various taxation classes, tax breaks and reliefs, payment conditions, land class changes, and the like. As the system is entirely based on the Polish statutory law, the Agricultural Tax Act, along with other statutory provisions of a more detailed nature serves as the only source of knowledge. So far, there

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 304 – 309, 2009.

has been no need to refer to any other legal acts although general legal expertise has often proven imperative to properly construe individual provisions.

Agriculture in Poland is not only one of these sectors of economy where the number of employees is still relatively high, but it is also very fragmented (with plenty of relatively small agricultural farms). Therefore, the number of agricultural tax payers is huge. As intended by the authors, the advisory tool, providing legal information on the rights and obligations of the agricultural tax payers, will come in handy not only for the taxpayers but also for the officers dealing with agricultural matters. It can facilitate and speed up the law interpreting process, cutting down the number of frauds. The focus of the authors in this study was on presenting the outcome of the efforts connected with building the ontology which would allow for representing individual cases, and dealing with cases not expressly regulated by law. This study will also outline the structure and concept of the system in question.

2 Drawbacks of Classic Expert System

At the outset of their studies, the authors focused on the possibilities offered by a classic expert system. Despite all the drawbacks discussed in detail by several authors ([3],[4] and others), such systems seemed tailored to expressing the provisions of statutory law (which was also pointed out by some authors [7]). Unfortunately, when put into practice, the classic expert system reveals a range of drawbacks, hindering, if not blocking, its practical application.

In the first place, it should be noted that the knowledge base is constructed for a certain purpose, i.e. the applicability of a specific group of rules, corresponding to a group of legal provisions, is limited to a set of inferences defined in advance. Hence, the provision flexibility is entirely lost, and so is the possibility of it being applied to interpret various cases and to solve various problems.

Cases representation poses another challenge to the classic expert system. In “real” legal practice, the analysed cases often differ, both in terms of cardinal matters and in minor nuances. A professional lawyer is able to fully comprehend and correctly interpret a given case, taking into consideration its specificity. In contrast, the classic expert system, where both interpretation and its outcomes are strictly defined in advance, makes it virtually impossible to reflect such specificity.

The inability to interpret cases not expressly regulated by law constitutes another disadvantage of the classic expert system. To compare, lawyers have a wide array of measures at hand to deal with such cases. Considering the expert system features discussed and the specificity of the legal act, the authors decided to develop a system which would comprise the quality of the classic system, leaving aside at least part of its drawbacks.

3 System Structure

Rules are the major carrier of legal expertise in the system developed by the authors. However, unlike in the classic expert systems, they are “incorporated” into certain elements of ontology, which allows for a case to be described. The ontology thus

forms an interpretation "background". Particular instances of the ontology elements, i.e. input and output elements (conditions and conclusions) of the rules, make it possible to describe specific cases, and to introduce certain semantic aspect into the static knowledge (describing the reality). Apart from the classic legal rules, regulating changes to the legal status (e.g. deontic features), the system also contains more general rules which govern cases not expressly defined in the letter of law.

To implement the system, the JAVA language and the JBOOS Rules engine were used, and the ontology was implemented as the structure of interfaces and classes. The real-life situations were expressed as instances of individual classes. Part of the procedural knowledge (e.g. the mechanisms used for calculating conversion hectares) was defined in the class-specific methods. Finally, legal rules, including the rules governing cases not expressly defined in law, were expressed as the JBOSS Rules engine. The choice of such implementation tools was inspired by the full compatibility of the JBOSS Rules engine with the JAVA language, as well as by the flexibility offered by these tools in representing this specific kind of knowledge.

4 Ontology

Any problem encountered by lawyers is highly specific, and this specificity must be properly accounted for to become interpretable in the context of the existing legal regulations. Several authors have made attempts to create more or less complex ontologies to represent legal acts [1],[2],[8],[12]. In consequence, the authors suggest the use of ontology for expressing the legal aspect of cases analysed. Further details concerning ontology can be found in [13]. It was implemented within the system as a structure comprising interfaces and classes, where an instant case is expressed through individual class instances. For example, if Mr. X is the owner of land in village Y, the description comprises the following class instances:

- Location ("place")
- Land ("land"), class have attribute: Location. Value of the attribute:"place"
- Village ("village Y") class have collection of attributes: Location. Value of the one of them:"place"
- Natural Person ("Mr. X")
- Ownership ("owns"). Attribute Owner, value:"Mr. X", attribute property: "land"

Naturally, each class consists of several attributes, some of which allow for making connections between individual instances. For example, "Location" is one of the attributes of the *Land* class instance, and the *Location* class instance serves as its value.

5 Interpretation of Cases Not Expressly Regulated

The legal theory and practice has given rise to a wide array of methods to deal with cases not expressly regulated by law, some of which were used by the authors. Implementation of one of the basic deontic rules, stating that any actions obligatory are also permitted, received top priority. In general, deontic logic is connected with the

rules of instrumental obligation, and prohibition, and permission. Of these three, the rule of instrumental permission was the only one to be considered relatively unquestionable, and thus was implemented. The authors further considered the possibility to apply the a'contrario interpretation method. The subject of deontic logic is widely discussed i.a. in [6],[9], instrumental reasoning and a'contrario is mentioned in [5][6].

5.1 Principle of Instrumental Permission

The principles of instrumental obligation, prohibition, and permission are among clearly defined legal mechanisms employed to deal with legal cases which are not expressly regulated in the law. As required by the theory of law, all these principles should be reasonably applied, especially within the penal and tax laws where obligations must be defined expressis verbis.

Of the three principles, the principle of instrumental permission is the least questionable, and it is the one to receive our primary attention.

The principle of instrumental permission assumes that the law abiding entity bound by norm X permitting a certain situation, may consider norm Y as binding, provided that it permits a certain act which constitutes a causal necessity for X [6].

Let us assume that there is a certain norm which permits X, and that there is an action which constitutes a causal necessity to satisfy this norm. The law does not expressly state whether this action, serving as the causal necessity to satisfy norm X, is permitted. However, by applying the principle of instrumental permission, we may reinterpret the norm as a new norm Y which permits this action as the causal necessity to satisfy X. Since we have no detailed information on this action, as well as on the conditions it should meet, it is difficult to infer what kind of action it might actually be. This would require, first of all, a good knowledge of norm X, and its connections to other norms, as well as the common-sense understanding of the cause-and-effect connections between individual actions and actual situations. Such knowledge and reasoning still remain a challenge in terms of their implementation in a computer system. Therefore, for the sake of this study, the scope of the principle under analysis will be slightly narrowed down.

Some of the permission norms take the form of a rule. This form allows us to assume that such conditions may be treated as necessary, in terms of causality, to enforce the permission. In light of the above, we may define the principle permitting a certain action A, on condition that there is an permission rule which hinges upon the occurrence of action A. This will narrow down the principle of instrumental permission but, at the same time, serve as a relatively simple means for extending the scope of functioning of the legal expert system, opening way for interpretation of some of the cases not expressly regulated.

5.2 Example

Since we apply an existential quantifier in relation to the rule defined above (rule of instrumental permission), this rule is clearly based on the second order logic, which slightly hinders its implementation. Example of rule of instrumental permission expressed in JBOSS Rules style is presented below (this rule permits first condition in "two-condition" rules):

```
rule "rule of instrumental permission-condition1"
          when
          cnd1 : Action(forbidden == true);
          cnd2 : Action();
          rule: RulePermitting(condition1.class ==
cnd1.class && condition2.class == cnd1.class &&
condition1.performer == condition2.performer);
          then
          cnd1.setPermitted();
          update(cnd1);
          end
```

5.3 A'contrario Interpretation

A'contrario interpretation assumes that if a certain norm prohibits action X, then action Y, being contrary to action X, is permitted [6]. The very specification of the area for comparison (in what sense action X is contradictory to Y) does not pose a huge problem, and such comparison is usually unambiguous. However, in some exceptional situations certain difficulties of various types may pop up – for instance, more than one action is contradictory to X. Implementation of a'contrario interpretation calls for the reference being made to the contradictory actions, while describing the actions within the ontology, in order to ensure that the system will be able to properly match which actions contradict with each other. This also permits the possibility not to define the contradictions for any dubious cases.

6 Conclusions

The authors of this study have attempted to develop an advisory tool functioning in the scope of the Agricultural Tax Act. The principal goal of this is to provide automatic legal advice. Implementation of certain mechanisms which allow for advising on cases not expressly regulated in law is what makes this project exceptional. The system comprises two levels of representation of legal knowledge: the level of ontology and the level of rules. The ontology developed by the authors to allow for representing specific cases serves as the basic representation level, making it possible to describe the strictly legal concepts, as well as the commonsense-based concepts.

Elements of ontology serve as the conditions and conclusions of the rules which form the dynamic part of legal knowledge stored in the system. Apart from the rules which directly reflect the provisions of the legal act, the system also comprises a range of rules of a more general nature. The latter mirror the principles of legal interpretation, including the basic rules of deontic logic, and the rule of instrumental permission.

The elements implemented so far include the ontology and part of the deontic legal principles. The system is well capable of providing correct answers to the cases which clearly fall within the scope of the knowledge already implemented, as well as to certain questions not expressly defined in the provisions.

Future works will focus on implementing further provisions and on developing the module supporting interpretation of cases not expressly regulated in law. The authors envision introducing a distinction between various rules, based i.a. on the results of

studies [7][8], and are also going to focus on the more formal representation of legal knowledge (on base of [5],[9]). This distinction would aim to expand and to crystalise the possibilities related to interpreting some of the cases not expressly regulated by law.

References

1. Benjamins, V.R., Casanovas, P., Breuker, J., Gangemi, A.: Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications. Springer, Heidelberg (2005)
2. Breuker, J., Valente, A., Winkels, R.: Legal Ontologies in Knowledge Engineering and Information Management. Artificial Intelligence and Law 12, 241–277 (2004)
3. Gordon, T.: Some Problems with PROLOG as a Knowledge Representation Language for Legal Expert Systems. In: Arnold, C. (ed.) Yearbook of Law, Computers & Technology, London, pp. 52–67 (1987)
4. Greinke, A.: Legal Expert System: A humanistic critique of Mechanical Legal Interface (1994), http://www.murdoch.edu.au/elaw/issues/v1n4/greinke14.txt
5. Hage, J.C.: Reasoning with Rules: An Essay on Legal Reasoning and Its Underlying Logic. Springer, Heidelberg (1997)
6. Leszczynski, L.: Issues of theory of application of law in Polish Zakamycze (2001)
7. Popple, J.: A Pragmatic Legal Expert System. Dartmouth, Aldershot (1996)
8. Sartor, G., Rubino, R., Rotolo, A.: An OWL Ontology of Fundamental Legal Concepts in: Legal knowledge and information systems. In: JURIX 2006, the Nineteenth Annual Conference, pp. 101–111. IOS Press, Amsterdam (2006)
9. Sartor, G., Shiner, R.A., Rottleuthner, H., Peczenik, A., Pattaro, E.: A Treatise of Legal Philosophy and General Jurisprudence. Springer, Heidelberg (2005)
10. Sartor, G.: Fundamental Legal Concepts A Formal and Teleological Characterisation (2006), http://cadmus.iue.it/dspace/bitstream/1814/4351/1/LAW%202006.11.pdf
11. The Journal of Laws of the Republic of Poland
12. Visser, P.R.S., Bench-Capon, T.J.M.: Ontologies in the Design of Legal Knowledge Systems. Towards a Library of Legal Domain Ontologies (1999), http://citeseer.ist.psu.edu/64358.html
13. Zurek, T.: Knowledge Base Ontology in Legal Expert System. Polish Journal of Enviromental Studies 17(3B), 575–580 (2008)

Identifying the Content Zones of German Court Decisions

Manfred Stede and Florian Kuhn

Applied Computational Linguistics, Dept. of Linguistics
University of Potsdam
Karl-Liebknecht-Str. 24-25, 14476 Golm, Germany

Abstract. A central step in the automatic processing of court decisions is the identification of the various *content zones*, i.e., breaking up the document into functionally independent areas. We assembled a corpus of German court decisions and argue that this genre belongs to the class of *semi-structured text documents*. Currently, we are implementing zone identification by means of a set of recognition rules, following up on our earlier experiences with a different genre (film reviews).

Keywords: court decisions, content zones, document parsing.

1 Introduction

Court decisions are legal documents with a high potential for later retrieval: Many interested parties can benefit from consulting, for instance, older decisions on cases similar to the ones they are involved in. It is, however, difficult to find relevant decisions merely by means of a full-text search, since "similar" cases are rarely characterized by the usage of "identical" words. Accordingly, smart retrieval methods could greatly enhance the accessibility of decisions. One prerequisite, which we are addressing in this paper, is the task of breaking up the court decision into its *content zones*: Identify all relevant portions of information such as the names of plaintiff, defendant, and the court; the date and official number of the decision; the description of facts of the case; the decision and its justification, legal principles and definitions applied, etc. Once this breakup of the overall document is known, it becomes possible to index the decisions with this structured information.

In Section 2, we first provide a characterisation of German court decisions and briefly review earlier work on automatically processing decisions; then we describe the corpus we constructed for our research. Section 3 argues that these court decisions are instances of *semi-structured text documents*, and describes our ongoing work on automatically identifying the content zones (which builds on our earlier research involving the same method but a different genre: film reviews). Finally, Section 4 summarizes the paper and describes our plans for future work with the decision corpus.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 310–315, 2009.

2 German Court Decisions

Since decisions are regularly published on the websites of German courts, it is not difficult to get an overall impression of the structure and regularities in the genre. Restricting ourselves to private law, in our initial informal survey we noticed that the texts have very similar overall structure yet differ in certain details, in particular in the order of information given in the caption, and in wordings used to mark specific portions. The regular presence of these "portions", however, lead us to postulate their role as *content zones:* pieces of information that are constitutive for the genre, show up at the same (or a small number of different) places in the texts, and can be identified either on the grounds of certain keywords or other formal characteristics, or based on their relative position to other content zones. We will show our proposal for an inventory of zones in the next section.

2.1 Earlier Work on Automatic Processing

Early work on identifying portions of legal cases (for purposes of summarization and retrieval) includes the SALOMON project [5]. The authors employed a 'text grammar' to identify the structure of criminal cases in Dutch on the basis of text cues. The parser yielded good results in terms of precision and recall but performed only a partial analysis, i.e., portions of the document were skipped over. In contrast, our goal is to realize a complete analysis of the documents with a hierarchical scheme of content zones.

A more recent related study is that of [3], which deals with summarising judgements rendered by the *UK House of Lords*. The goals were to grant access to judgements to non-experts, and to enable further information retrieval methods. The authors assembled a corpus of 188 Judgements taken from the Website of the UK House of Lords, and manually annotated the *rhetorical role* of each sentence as well as its relevance. The annotation scheme was based on the work on summarizing scientific articles by [6], and the roles used were the following:

- FACT – A recounting of related events or circumstances.
- PROCEEDINGS – A description of legal proceedings.
- BACKGROUND – A direct quotation or citation of source or law material.
- FRAMING – Part of the law lord's argumentation.
- DISPOSAL – Either credits or discredits a claim or previous ruling.
- TEXTUAL – A sentence relating to the structure of the document.
- OTHER – A sentence that does not fit into any of the above categories.

Following the annotation, several machine learning techniques such als NB, SVM and ME were applied to automate rhetorical and relevance classification of the texts. Using state-of-the-art cue phrase information, encouraging results were achieved with SVM and ME, and the authors concluded that the two steps can provide the basis for subsequent summarization algorithms.

For German legal texts, the only relevant work we are aware of is that of [7]. Here, the idea is to extract legal *defintions*, which play a central role in

German jurisprudence and thus are particularly important for retrieval. The authors used a rule-based approach, working with a corpus of some 6000 verdicts of German enviromental law provided by the *Juris* database[1]. The verdicts were first processed by a dependency parser to construct abstract semantic representations of the sentences. These were used to transform the definition's dependency patterns into a set of 33 extraction rules. Different evaluation techniques were used, and the best precision values achieved were slightly above 70%.

2.2 Corpus Creation

When building up our own corpus, we focused on private law but made sure that many legal domains were covered. Also, we took care to include decisions from courts at various levels in the hierarchy and from different regions in Germany. Since courts develop their own habits of structuring and presenting decisions, breadth in the corpus is important so that generalizations as to the presence and rcognizability of content zones can be drawn. We thus collected 40 decisions from 12 different courts, not aiming to achive any "completeness" in legal domains but making sure that no single domain would show up considerably more often than the others. The length of the documents varies between four and 13 pages. Since the great majority of decisions are published in PDF format (with only a few in HTML), we decided to build our extraction rules on plain text. Therefore, the corpus was converted to plain text files with no explicit structural markup.

3 Analyzing Semi-structured Text Documents

Depending on their genre, text documents display different degrees of regularity in terms of both content and structure. For example, the majority of weather reports, cooking recipes, or scientific papers have a highly canonical structure: There is a small set of content zones that collectively characterize the genre, and these appear in a typical linear order, with only little variation. *Semi-structured text documents* are somewhat less regular, and we define them by the following characteristics:

- Texts consist of genre-specific content zones.
- Content zones correspond to layout units (or zones in the *logical document structure*): A single paragraph does not contain more than one zone. (But a zone may be longer than a paragraph.)
- Most zones are obligatory; some are optional.
- The ordering of the zones in texts of the respective genre is not entirely fixed but not arbitrary, either. There are certain regularities: Some zones occur at fixed positions; some occur in the neighbourhood of other zones.
- Some of the content zones can be identified by surface patterns, such as specific keywords or a small number of variants thereof.

[1] www.juris.de

We are interested in defining the tag sets for content zones, and in procedures for identifying them automatically. For the latter purpose, we use a software environment allowing a highly modular approach to document processing.

3.1 The MOTS Workbench

Our implementation is embedded in the MOTS workbench, an environment for text document processing developed at Potsdam University over the past five years. At the heart of MOTS is a highly generic XML standoff format for linguistic data ('PAULA', see [2]), and we have developed a series of converters that map the output of existing analysis tools to and from PAULA. Thus, new modules can be integrated by providing the PAULA converters for in-/output and adding it to the processing pipeline. At present, MOTS contains modules (developed by ourselves or by external parties) for part-of-speech tagging, syntactic parsing, coreference analysis, statistical term weighting, document zoning, and others; target languages are German and English. The first application developed with MOTS was a text summarizer [8].

Most modules in MOTS are genre-independent, but some provide extra functionality for specific text genres only. One of these is a zone identifier for movie reviews, which assigns labels such as DIRECTOR, TITLE, DESCRIBE-STORY, COMMENT-STORY etc. (26 in total) to paragraphs of German film reviews. It is a combination of statistical classifation and matching of manually-written rules; we describe it below for the genre of court decisions. As reported in [1], the average recognition rate was 70% precision and 63% recall, with precision being generally higher for the most frequent zones.

3.2 Content Zones in Court Decisions

From the corpus study (and largely in consent with the German code of civil procedure), we identified five segments that can be further divided into more fine-grained content zones. The five segments always appear in the same order; within the segments there are tendencies for linear order (but no strict sequence). Zones given in square brackets are optional, those with an asterisk can occur multiple times.

- Caption: Court name, Case identifier, Date, Plaintiff, Defendant, Formulae such as "Im Namen des Volkes", "Für Recht erkannt"
- Operative provisions, Summary of judgement: [Consequences for plaintiff], [Consequences for defendant]
- Case description: General description, Plaintiff's view, Plaintiff's proposition, Defendant's view, Defendant's proposition
- Justification: Introductory statement, (Subsumption: sequence of conclusion, definition, and its application)*, (Secondary judgement)*
- Signatures

3.3 Zone Analysis Procedure

Recall that we defined zone breaks to co-occur with paragraph (or line) breaks. Thus the problem of zone identification is to assign a label to each paragraph (or line). For the court decisions, we have the additional constraints of the four larger areas, which in the first step can be identified on the basis of layout and keywords. The remaining problem thus is to make decisions for zones-within-areas: mandatory zones have to be assigned, ordering constraints have to be respected, and ordering preferences should be taken into account.

The implementation is currently under way. As with our film review work (see above), we use a two-step procedure: First assign labels to zones that can be identified reliably; then complete the assignment on the basis of ordering information and the mandatory/optional zone constraints.

For the first phase, we rely on formal features such as length of paragraph (or line), the presence of keywords, and the like. We use LAPIS [4] for writing recognition rules that can contain regular expressions but also offer convenient (and readable) ways of specifying relationships between portions of text. LAPIS is an interactive, web-based tagging system, which makes rule writing and debugging quite easy; in the end, we extract the rules from the system and use the matching mechanism as a Java module that can be called from MOTS. At the end of the first phase, some zones in each of the four regions have already been reliably identified.

In step 2, the yet unassigned paragraphs are being labelled on the basis of ordering constraints and obligatoriness: A zone that according to our "document grammar" has to be present in a certain range of the document can sometimes be assigned only based on these constraints; otherwise, we employ probabilistic information to make the decisions: bigrams give probabilities of zone neighbourhood, and a zone already reliably identified in phase 1 can serve as anchor to make decisions on its neighbours based on the bigram probabilities.

4 Summary and Outlook

We have described our ongoing work on identifying content zones in German court decisions. In comparison to our earlier work with film reviews, we expect better precision and recall, as the court decisions display a more regular structure than the reviews, and many zones can be reliably identified by fairly straightforward rules. Still, the overall recognition process remains a combination of evaluating strict and soft constraints. The inventory of zones we are using has been derived from a corpus study involving 40 decisions from 12 different courts (private law only).

Following the zone identification, our next step is to look in more detail into the *argument structure* of the decision's justification. Building on our earlier work on analyzing the argumentation in newspaper commentaries [9], we develop a level of representation for the argument, and then want to exploit the possibilities of at least partially reconstructing the argument. In comparison to newspaper commentary, the argumentation in justifications is much more transparent, and

individual steps are usually clearly signalled by suitable connectives; on the other hand, the arguments can at times be much more complex than those found in newspaper commentary.

References

1. Bieler, H., Dipper, S., Stede, M.: Identifying formal and functional zones in film reviews. In: Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue (2007)
2. Dipper, S.: XML-based Stand-off Representation and Exploitation of Multi-Level Linguistic Annotation. In: Proc. of Berliner XML-Tage (BXML 2005), pp. 39–50 (2005)
3. Hachey, B., Grover, C.: Extractive summarization of legal texts. Artificial Intelligence and Law 14, 305–345 (2006)
4. Miller, R.C.: Lightweight Structure in Text. PhD thesis, Computer Science Department, School of Computer Science, Carnegie Mellon University (May 2002)
5. Moens, M.-F., Uyttendaele, C., Dumortier, J.: Abstracting of Legal Cases: The SALOMON Experience. In: Proc. of the 6th Int'l Conference on Artificial Intelligence and Law, Melbourne (1996)
6. Teufel, S., Moens, M.: Summarizing Scientific Articles – Experiments with Relevance and Rhetorical Status. Computational Linguistics 28(4) (2002)
7. Walter, S.: Manfred Pinkal. Linguistic support for legal ontology construction. In: Proceedings of ICAIL, pp. 242–243 (2005)
8. Stede, M., Bieler, H., Dipper, S., Suryiawongkul, A.: SUMMaR: Combining linguistics and statistics for text summarization. In: Proceedings of ECAI, Riva del Garda (2006)
9. Stede, M., Sauermann, A.: Linearization of arguments in commentary texts. In: Proceedings of the Workshop on Multidisciplinary Approaches to Discourse, Oslo (2008)

SAW 2009 Workshop Chairs' Message

Dominik Flejter[1], Tomasz Kaczmarek[1], and Marek Kowalkiewicz[2]

[1] Poznan University of Economics
D.Flejter@kie.ue.poznan.pl, T.Kaczmarek@kie.ue.poznan.pl
[2] SAP Research Brisbane
marek.kowalkiewicz@sap.com

In recent years, the Web has moved from a simple one-way communication channel, extending traditional media, to a complex "peer-to-peer" communication space with a blurred author/audience distinction and new ways to create, share, and use knowledge in a social way. This change of paradigm is currently profoundly transforming most areas of our life: our interactions with other people, our relationships, ways of gathering information, ways of developing social norms, opinions, attitudes and even legal aspects, as well as ways of working and doing business. The change also raises a strong need for theoretical, empirical and applied studies related to how people may interact on the Web, how they actually do so, and what new possibilities and challenges are emerging in the social, business and technology dimensions.

Following the two previous events, the call for papers of the 3rd Workshop on Social Aspects of the Web (SAW 2009)[1] covered a wide area of topics related to development and usage of social services in the contemporary World Wide Web, including topics related to people on the social Web (individuals, communities, their collaboration and business activities, as well as connection between on-line and off-line activities), data and content on the social Web (their creation, ordering, dynamics and portability), social software and services (their development, adoption, commercial applications, business models and classification schemes) and social Web mining (mining user-generated content, social graph and activity patterns).

The Program Commitee of the Workshop included 16 researchers from ten countries, specialising in different aspects of on-line social and collaborative tools, social impact of Web developments and socially-driven evolution of Web data, content and on-line services.

Out of nine submissions to SAW 2009 Workshop, three were accepted and presented at the Workshop, held on the second day of BIS 2009 Conference (Tuesday, April 28, 2009). The authors of accepted papers came from four institutions located in four countries. The papers cover significantly different aspects of social Web and approach them from very different positions, underlining interdisciplinary character of the SAW Workshop.

The first paper, authored by Peter Schnitzler, Marius Feldmann, Maximilian Walther and Alexander Schill, presents a systematic approach to social networking platforms evaluation. The authors propose a quantitative framework allowing for classification and comparison of social networking platforms with

[1] http://bis.kie.ae.poznan.pl/12th_bis/wscfp.php?ws=saw2009

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 316–317, 2009.

respect to different aspects of their functionality and offered API. Consecutively, they apply this framework to five social networking platforms, mostly dedicated for enterprise solutions.

The second paper, authored by Mikhail Simonov and Gary Bridgeman, demonstrates how social Web and mobile technologies may be jointly used to limit energy consumption by travellers. The authors approach the topic of energy consumption minimisation from the perspective of social networks and temporal reasoning. In their research they analyse ad-hoc social networks of people that may be dynamically organised through social aggregation, based on their co-presence in specific location, at a given moment of time.

The last paper, authored by Matthieu Tixier and Myriam Lewkowicz, provides an example of how Internet-based tools can help solve real-life societal problems. The paper defines functional requirements for a social on-line platform for communities of family caregivers, based on an in-depth study social support practices and of these caregivers social capital. As a result, the authors demonstrate how traditional social support services, being subject to a number of limitations, can be supplemented by social Web tools.

We would like to thank all authors of submitted papers as well as Program Committee members for making SAW 2009 a success, and to invite you to take part in next edition of SAW Workshop that will be held in May 2010 in Berlin, in conjunction with the 13$^{\text{th}}$ International Conference on Business Information Systems (BIS 2010).

Guideline for Evaluating Social Networks

Peter Schnitzler, Marius Feldmann, Maximilian Walther, and Alexander Schill

Chair for Computer Networks, TU Dresden, Germany
peter@schnitzlers.de,
{marius.feldmann,maximilian.walther,
alexander.schill}@tu-dresden.de

Abstract. It is evident that Web-based Social Network platforms have gained major attention for communication particularly within companies. Nevertheless there exist no real guidelines for selecting an appropriate platform that fits the company's needs yet. In this paper a general guideline for this purpose is developed. A global matrix for evaluating Social Networks is presented. The matrix results in a formula which allows a comprehensive but customized rating for Social Networks. The paper is structured as follows: (1) Categorization of Social Networks; (2) Typical functionalities offered by Social Network platforms; (3) APIs for Social Networks; (4) Criteria for evaluating Social Network Frameworks and White Label Social Networks; (5) Proof of concept for the developed evaluation matrix.

Keywords: Social Networks, matrix of evaluation, white label networks, Social Network frameworks, Guideline.

1 Introduction

Web-based Social Networks are one of the most important phenomena of the so called "Web 2.0". Their success story can be measured by the heavy increase of participants in Social Network platforms and by the number of available platforms itself.

Due to an increase of Social Networks importance, companies and Non-Governmental Organizations (NGOs) have started to take interest in them. They focus not only on the usage of these platforms as a means for marketing or public relations, but also on how they can be used as a tool for internal communication. Therefore, they are particularly intended to enable a more efficient usage of knowledge within the company. This can be achieved by evolving the intranet into a Social Network and developing the company's internal communication within the established platform. The central problem that arises is how to choose an appropriate Social Network platform. In the following section a general guideline will be developed aimed at facilitating an appropriate decision for such a Social Network platform.

There has not been in-depth discussion on how to evaluate Social Networks. As a result, not much related work for this topic is available. Some publications deal with Social Networks in general; however, the topic "Evaluation Social Network Frameworks" has not been covered so far in a publication. Only two larger web-publications by Owyang [7], and Hendrickson [8] are relevant in this area.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 318–329, 2009.

A broad overview of social computing is offered by Parameswaran [1]. She focuses on the different tools in the field of Social Networking that are available and how they change workflows. A basic classification of Social Networks is offered by Richter and Koch [2]. While the description of the different tools is detailed, no scale for evaluation is given. Trust and privacy, which are closely related to the functions of a Social Network are considered by Dwyer und Hiltz [3]. Golbeck's [4] studies focus on the number of users and how they are connected to each other. In a recently published article Reisberger et al. [5] develop a rough classification scheme for Social Networks and identify three distinctive dimensions: Communication, profiling and connections. This approach, which has been used up to now, only offers a basic classification. Even if it provides an insightful overview, it is not sufficiently detailed for the purpose of evaluating frameworks within a certain use case. Moreover, the classification does not offer any guideline regarding performing the actual process of benchmarking. The approach for evaluating Social Networks presented in this paper offers a comprehensive method which can easily be customized according to specific needs.

The first step that has to be taken when thinking about operating a Social Network is a categorization that helps to identify useful networks. Such a categorization is presented in section two.

In a next step the functionalities that should be offered must be defined. In the case of Social Networks this topic deserves special attention since the subject is very broad and the understanding of which functions are necessary for a Social Network can vary to a great extent. Section three presents the most essential functionalities of Social Networks based on large analysis of existing Social Networks. Criteria for evaluating these functionalities are given and a scale for benchmarking is proposed. An important attribute of the Web 2.0 is the existence of an Application Programming Interface (API). Frequently only a combination of different services provides major benefits to the customer; this is also the case with Social Networks. Not only are the possibilities of connecting the network to external services in the focus of interest, but also in particular the connection to further Social Networks. Thus, chapter four puts its emphasis on the APIs of Social Networks. In a first step different kinds of APIs are presented and in the second part important criteria for the evaluation of APIs are derived.

All the previous steps are merged into a matrix of evaluation. This matrix is explained in chapter five. Finally, a proof of concept is presented. It applies the matrix of evaluation to a real world scenario. Five Frameworks have been chosen to demonstrate how the evaluation guideline can be used. The last chapter summarizes the most important results of this paper and identifies future work.

2 Categories of Web-Based Social Network Platforms

First of all the term Social Network has to be defined. Based on the definition of Boyd and Ellison [7] it is defined as the following: “Social Networks are software systems that allow the creation of user profiles, the connection between users and the access to shared resources.”

Based on this definition, different categories of Social Networks are defined. This is a necessary prerequisite for designing a meaningful evaluation matrix for Social

Network frameworks. Most familiar to the broad public are networks which offer a variety of services to the users (*All-in-One Networks*). The second type is formed by networks which only offer a single service such as tagging bookmarks (*Tool Networks*). Although they do not receive the same public attention as All-in-One solutions, such as Facebook or MySpace, they nonetheless do have large numbers of users and are highly relevant since they can be connected to other Social Network services via an API. The third category is formed by Social Networks such as Ning (*Create your own Social Network*), which offer the possibility to create own Social Networks more or less by only choosing a set of options within a web frontend. The Tool Networks can additionally be subdivided into blogging, photo, movie, music, travel/accommodation, bookmarking and meeting tools.

The second aspect is more focused on the user. Firstly, it has to be considered if the network is dedicated to a certain subject. These are networks which, for example, focus on bringing people together while travelling, eg. couchsurfing.org and hospitalityclub.org. The second criterion is whether only certain people are admitted as registered users. For example (by now) Facebook addresses the general public while studiVZ.net is meant for students only.

By using these two dimensions of aspects with their criteria it is possible to gain a first overview of the spectrum of Social Networks that is available.

3 Functionalities of Social Networks

In a next step the focus has to be directed towards a detailed examination of the functionalities offered by the networks. The gathering of the functionalities within this case study took place by collecting information from over thirty networks[1]. As key components of the Networks, functionalities covering the following aspects were identified: Accounting, Application/APIs, Blogs, Privacy, Photos, Guestbook, Groups, Messages, Calendar, Profile, Search, Tagging and Videos. The functionality to connect to other people was not investigated as a separate category since it is a key function and is covered by the discussion of the other categories. For each category particular attention is paid to the possibility of importing and exporting data – in case it is an appropriate function.

3. 1 Functionalities and Their Criteria

For every of the listed functionalities central concerns have been identified and formulated as questions in order to enable a simpler characterization of the functionality:

- Accounting
    - Is a registration necessary?
    - Which information needs to be provided when signing up?
    - Is an OpenID login available?

[1] Asian Avenue, Bebo, Black Planet, Care2, Consumating, Couchsurfing, Cyworld, Dodgeball, Dogster & Catster, Facebook, Flickr, Fotolog, Friendster, Hi5, Hospitalityclub, Hyves, Last.FM, LinkedIn, LiveJournal, lokalisten, meinvz.net, MiGente, Mixi, MyChurch, MySpace, Ning, Orkut, Piczo, schuelervz.net, SecondLife, SixDegrees, studiVZ.net, tribe, Twitter, Windows Live Spaces, Xing, Yahoo! Mash, Yahoo!360, YouTube.

- Applications and APIs (will be considered in detail later in this paper)
  - Which data can be imported and exported?
  - Can developer write customized applications?
- Blogs
  - How similar is the Social Network to a regular blog system?
  - Are key functions for blogging like trackbacks and RSS feeds available?
- Privacy
  - How detailed can the user define the privacy regarding groups of persons?
  - For which data is this possible?
  - Important to consider is that the provider always has access to all data
- Photos
  - Is an upload and download of photos possible?
  - Can persons on pictures be tagged?
- Guestbook
  - Can entries be commented?
  - Can the user include multimedia content?
- Groups
  - What kinds of groups are available?
  - Are there open and closed groups available?
  - Are there special groups like "event centered" groups for lectures?
  - Are there pre-defined groups which were created by the system provider?
- Messages
  - How sophisticated is the messaging system?
  - Can multimedia message be sent and is contact to external persons possible?
- Calendar
  - Can the user create events?
  - Which functions are connected to events?
  - Can events be exported into other programs?
- Profile
  - Can the profile be customized and if yes, to what extent?
  - Are there feeds available which show user activity?
- Search
  - What kinds of searches are available?
  - Which content can be searched for and is it possible to limit the search to certain categories?
  - Are there intelligent searches such as "Persons you might know." Or "Persons that have the same interest and know somebody you know." available?
- Tagging
  - Which information can be tagged by the user?
  - How can the user access tags?
  - Are there tools like automatic tag completion available?

- Videos
    - Can movies be uploaded or integrated into the platform?
    - Can they be downloaded as well?

3.2 Conclusion for Evaluation Functionalities of Social Networks

The presented criteria for functionalities can be used to evaluate the suitability of a Social Network for the specific environment in which the operator is going to use the Social Network. As shown above, a number of functionalities exists that has to be taken into account. Each of these functionalities has again a couple of further options that can be considered. While the criteria can only be briefly described in this paper, a detailed listing is available in [6]. The criteria analysis provided here is used to create the global evaluation matrix.

4 APIs in Social Networks

There are two main fields of application for APIs in Social Networks. The first one is the possibility of applications that can be created by external developers and then be used by the customers. The second type of APIs enables data exchange. This includes user data such as birthdays and address information as well as data about contacts. Especially in an intranet environment APIs play an important role due to the wide variety of applications the customer wants to integrate into the platform.

Applications that may be accessed cover a wide range of variants. Some of them are just simple games while others offer useful additions such as calendar extensions for external synchronization. When evaluating this type of API, a number of important facts have to be considered. The structure and the available wrapper languages are the first aspects that have to be evaluated. Wrapper languages[2] are important since they enable the developer to use his or her preferred coding language and provided customize functions. Even for Representational State Transfer (REST) APIs wrapper languages are useful because they bring the logic provided by the Social Network to the programming language. The second component of many APIs is a special set of tags which can be used to adopt the look and feel of the host Social Network, to trigger certain behavior, such as special rights or as a replacement for variables. OpenSocial is clearly a key player in this area with an API that was designed to work on different systems. Regarding the elaborateness of features, the Facebook API can be considered the most advanced one.

The second important function is the access to user data. Often it is used within the application API, however, this point becomes increasingly important as a standalone feature as well. While the existing APIs only allow the usage of user data within the networks, a new type of API is designed for using the data outside. Facebook with Facebook Connect and Google with Friend Connect are receiving a lot of attention in this field. An already functioning API is offered by "Windows Live "and "Yahoo!

[2] Wrapper languages provide a custom interface for certain programming languages to access for example a REST API from a Social Network.

Address Book API". Their features are quite handy; however, they are not available in the huge Social Networks. The key functionality for this kind of API is the possibility to use own contacts and data on other websites. This works without exporting data but rather by having a real time access - so no manual updating is necessary. While the technical development in this area is starting to leave the beta status, questions of privacy still have to be discussed.

After pointing out the two fundamental possibilities to implement an API, a couple of criteria for evaluation are given which can be divided into two levels. The first level contains all the points regarding the distribution of the API and the second one includes technical aspects.

From a technical point of view, the structure of the API is important since it determines what can be done with the API. Especially if the developer wants to embed the applications into the network, a REST-based API is not helpful but rather tags for the correct look and feel are useful. Furthermore, it is important how the user data can be accessed. The following central questions may be asked: Can data only be read or also be written? Can all types of data be written or only certain fields? Is the access to the API limited concerning transactions?

In addition, the distribution of an API should not be underestimated. There are some network operators which only support their own API; whereas others also do support commands from other APIs. The trend is certainly to support the important APIs from Facebook and Google so that developers can reuse their existing code.

Besides the official APIs provided by the Social Network there are open standards which intend to work independently of the platform. OpenID, FOAF (Friend of a Friend), hCard (HTML vCard) and XFN (XHTML Friends Network) can be considered the most important ones. OpenID enables the user to use one login for several platforms if it is supported by the Social Network. FOAF and hCard are formats for data exchange. While hCard is more like a digital business card, FOAF offers the possibility to export contact data about the user himself, but also about his contacts. Using these data, connecting graphs can be created when analyzing several FOAF files. XFN adds semantic information to links and identifies ties between different sites.

Although the basic functionalities seem more important at first sight when taking a look at Social Networks, the importance of the API should not be underestimated. The API is the only way to extend the functionality of a network as well as to enable the operator to develop customize functions.

5 Matrix of Evaluation

In a next step the described criteria will be applied to the selected frameworks. Each sub-criterion is ranked on a scale from 0-5. 0 indicates that the function does not exist at all and a value of 3 defines a feature that is properly implemented without any special additions. 5 points are assigned when significant extras are offered in addition to the basic functions. This could be for example the possibility to access an API with several different technologies. Afterwards the sub criteria are summed up and an overview of all main categories is created. In the last step the formula is applied.

5.1 Sum

For a first impression on how suitable a network is, the sum of all evaluated points has been taken. This is particularly useful if a framework with a high number of functions is necessary.

5.2 Coefficient of Variation (cov)

When building the sum, the variation of quality that certain criteria have is not reflected. For this purpose the coefficient of variation is very useful. In addition to the sum it indicates if all the functions reach an equal level of quality. A low coefficient of variation means that the quality of all functionality is approximately the same. If a high value is reached, it is likely that some functions are implemented very well while others are poorly implemented.

5.3 Loading

Problems arise when a systems is desired in which certain functionalities are of special importance. In this case it is necessary to integrate the loading. The sum of each group of sub criteria can be matched with a loading, thus allowing the inclusion of individual wishes from customers.

5.4 Formula for Evaluation

Finally, a formula enabling a quick comparison of different frameworks and Social Networks on the basis of the previously presented criteria can be formulated. In addition to the already introduced tools (Coefficient of Variation, Loading and Sum), a new one is added. This is the difference between the value of one criteria and the average of the benchmark of all other criteria. This makes the result more distinct because it gratifies functions that are better implemented than others. The sum of one subgroup can be multiplied with a loading so that user specific interests are included. The next step is to sum up all the criteria groups to one number. This sum is divided by the coefficient of variation. This has the impact that well developed products with an equal quality for all functions score better. If this is not a primary concern for the evaluation this last step might be skipped.

$$\frac{\sum AllCatergories((\frac{\sum SubCriteriaOfOneCriteria}{\varnothing ThisCriteriaFromAllSystems}) \bullet loading))}{CoefficientOfVariation} \quad (1)$$

The result of the calculation gives a quick but solid impression regarding the quality of the evaluated Social Network platform or framework.

6 Evaluation Example

So far the theoretical possibilities for evaluating a Social Network or a Social Network framework have been presented. To prove that these criteria and tools are

suitable, a field test has been applied (more details are described in [6]). The process that was chosen consists of six steps. In the **first step** an intensive search for suitable frameworks has been realized. Over 40 frameworks[3] were found which enable the operation of a Social Network in general. In the **second step** these frameworks were classified by the following categories:

(1) Using an API from an existing network;
(2) Having the customization done by a service provider which offers SaaS products;
(3) Installing the software on an own machine and having full access to it for customization purpose.

The goal of the **third step** has been to reduce the number of frameworks and select the ones that are evaluated in detail. The following important selection features have been identified:

The most important point is the amount of customization possible regarding design and functions. Furthermore, one has to distinguish between the customization that can only be done by the provider and the ones that can be done by the customer or a freely chosen service provider. The next important decision is, if an All-in-One Social Network is required. Regarding changes to the functionality, the API as well as its documentation becomes very important. Other factors, especially important when considering the existence of a product in the future, are the market position of the company as well as which customers are using the framework already. For the evaluation example presented in this paper, the following five points were used to select the final five frameworks:

(1) The possibility to have an installation on your own account;
(2) Should be an "All-in-One"-Solution since they offer the highest amount of functionalities and are more suitable for large evaluations;
(3) The possibility to change the design and add custom code using an API;
(4) A solid and complete documentation;
(5) The Frameworks should already have a certain relevance proven by existing customers or the importance of the company brand.

Due to these criteria the following five networks have been selected:

(1) Peopleaggregator (Broadband Mechanics);
(2) Clearspace Community (Jive Software);

[3] Acquia, Affinity Circles, Alfresco, AlstraSoft, Awareness, Blogtronix, BoonEx, Clonesumating, Community Engine, Converdge, Crowd Factory, Dzoic Handshake, ektron, Elgg, Golightly, GroupSwim, Igloosoftware, Inquira, Insoshi Portal, introNetworks, iScripts SocialWare, KickApss, Kwiqq, Leverage Software, LiquidPlanner, Lithium Technologies, LiveWorld, LovdByLess, MyWorklight, mzinga, Neighborhood America, nNGenera, Omnifuse, OneSite, OpenLink Data Spaces (ODS), PHPizabi, PixPulse, Pringo Networks, Prospero Technologies, SelectMinds, Sharepoint, Small World Labs, Social Platform, SocialEngine, SocialText, Sparte Social Network, Spigit, SuiteTwo, TheSchoolHall, Trampolinesystems, VMIX Media, Web Crossing, Web Scribble Solutions, Webligo.

(3) Community Server (Telligent);
(4) Lotus Connections (IBM);
(5) Ning[4] (Ning).

S**tep four** is the evaluation itself. Each category is split up into subsets of functions as shown in table 1 for applications and APIs.

Table 1. Applications and APIs

	People Aggregator		**Clearspace Community**		**Community Server**		**Lotus Connections**	
API (Import)	3	RSS Widgets FOAF XFN	5	RSS Widgets Blogimport Emailimport	2	Share Membership	0	
API (Export)	3	RSS feed	5	RSS export for all content vcf cards of persons	4	RSS for activities, blogs, bulletin board and comments	4	RSS for activities, blogs, comments and bulletin board
Widgets / Apps	3	Change core code Web Services API	5	Widgets Web Service Client (SOAP, REST, XML_RPC Core API	3	REST API Theme API Community Server API	3	Sametime MS Office Widgets
Import of Addresses	4	hCard Facebook AIM Flickr	0	No options	0	No options	0	No options

Due to the wide range of supported formats and APIs, the SaaS product Ning has been rated with: API (import) 4, API (export) 4, Widgets/Apps 5, Import of Addresses 4.

In **step five** all the subsets are added and inserted into a table. Now the key numbers can be calculated. The sum and the coefficient of variation give a first overview of the networks' relation between each other. In the last and **final sixth** step the formula is used and a single value for each framework is calculated. The formula gives the opportunity to put weight on different criteria. This enables a specific evaluation for the usage in a specific scenario.

[4] Ning has been chosen to get an insight how well SaaS products fit into the testing field and also how well low budget solutions can be used.

Table 2. Evaluation of Social Networks

	Ø	People Aggregator	Clearspace Community	Community Server	Lotus Connections	Ning
APIs	3	3	4	4	2	4
Blog	4	3	4	5	5	5
Privacy	3	3	3	2	1	4
Groups	2	3	3	1	3	2
Search	3	2	4	4	5	2
Content	2	1	4	2	3	2
Messages	2	2	2	4	0	2
Calendar	1	1	0	0	1	4
Tagging	4	3	4	3	5	3
Collaboration	3	2	4	2	5	2
Profile	3	2	2	2	4	5
Sum		26	34	29	34	**35**
Coefficient of variation		**0,345**	0,420	0,569	0,604	0,392
Formula		716	985	518	685	**1310**

The evaluation shows that the functionalities that are available differ heavily. More business centric Social Network Frameworks offer more static functions to connect to other people, while Ning offers a combination of different possibilities. Another important impression is that if less functionality is offered, the single functionalities are better integrated into the system.

Regarding open standards the situation is not really satisfying. Only Community Server offers the OpenID technique. Often, the definition of custom profile fields cannot be accomplished with a graphical tool. The possibility to write a blog is offered by all networks; however, Lotus Connections provides the most distinct features. Really Simple Syndication (RSS) is used for information about new blog entries. Privacy does not appear to be a primary concern of the frameworks. Only People Aggregator enables the user to determine which data should be visible to others. The possibility to restrict the access to certain user groups can only be used by the user of Ning and People Aggregator. The most advanced functions for grouping people are offered by Clearspace Community but only Lotus Connections disposes of event centered groups. Even a simple feature like the search reveals considerable differences. People Aggregator and Ning include no possibility to restrict the search to certain areas, whereas Clearspace Community and Lotus Connections offer intelligent restrictions. Functions to import data and documents are only available on a very limited basis. Sending and receiving messages is a key functionality and offered by all networks except Lotus Connections. The implementation of tags is a good example of how functions can be implemented in depth. While tags in public Social Networks are often only used in a basic manner (e.g. studiVZ), Lotus Connections and Clearspace Community exhibit the wide range of possibilities of this function. Regarding the

API, Clearspace Community offers the greatest variety of functions. While REST and SOAP APIs are common, the possibility to write additional plug-ins is not very common. In summary it can be said that Clearspace Community, Ning, Community Server and Lotus Connections can be recommended, while People Aggregator still exposes too many bugs and unfinished functions.

When comparing the final results of the formula for each network with the evaluation of the single categories, it is proven that the results are very satisfactory.

The proof of concept which was executed in [6] has shown that the matrix of evaluation and the formula provide a good way to compare different Social Networks. The results of the evaluation were confirmed by the use of Clearspace Community for a prototypal network implementation and in-depth analysis of its adequateness.

7 Conclusion

In this paper a global approach for evaluating Social Networks and frameworks for creating them has been provided. In comparison to past approaches this method does not focus on one or two certain topics such as privacy or photos. Although it takes the whole range of functionalities of Social Networks into account, it still offers possibilities to include user specific needs. The criteria for the evaluation were developed on the basis of a large amount of existing Social Networks. As a result, a formula has been developed which can be used to compare Social Networks and their frameworks quickly. The validity of the approach was shown by a proof of concept using five existing frameworks. The evaluation proved that the results are comparable and that user interests can be taken into account easily. In future, the next step will be a deployment in different use cases to gain further knowledge using the formula.

References

1. Parameswaran, M., Whinston, A.: Social Computing: An Overview. Communications of the Association for Information Systems 19 (2007)
2. Koch, M., Richter, A.: Social Software - Status quo und Zukunft, Technischer Bericht Nr. 2007-01, Fakultät für Informatik, Universität der Bundeswehr München (2007), http://www.kooperationssysteme.de/wp-content/uploads/RichterKoch2007.pdf
3. Dwyer, C., Hiltz, S., Passerini, K.: Trust and privacy concern within Social Networking sites: A comparison of Facebook and MySpace. In: Americas Conference on Information Systems (AMCIS), Keystone (2007)
4. Goldbeck, J.: The Dynamics of Web-based Social Networks: Membership, Relationships, and Change. In: Sunbelt XXVIII International Sunbelt Social Network Conference (2008)
5. Reisberger, T., Reisberger, P., Smolnik, S.: Entwicklung eines funktionalen Klassifikationsschemas für Social-Networking-Systeme. In: Meissner, K., Engelien, M. (eds.) Workshop GeNeMe 2008 - Gemeinschaften in Neuen Medien (Veranst.): Virtuelle Organisation und Neue Medien 2008 Workshop GeNeMe 2008 - Gemeinschaften in Neuen Medien, pp. S43–S55 (2008)

6. Schnitzler, P.: Technologische Analysen im Umfeld Sozialer Netzwerke, Diplomathesis, http://nbn-resolving.de/urn:nbn:de:bsz:14-ds-1226399874317-02633
7. Danah, B.D., Ellison, N.: Social Network Sites: Definition, History y, and Scholarship. Journal of Computer-Mediated Communication 13, S.210–S.230 (2008), http://dx.doi.org/10.1111/j.1083-6101.2007.00393.x, DOI 10.1111/j.1083–6101.2007.00393.x
8. Owyang, J.: Forrester Under way to Catalog the White Label Social Networking Space (2008), http://www.web-strategist.com/blog/2008/05/28/forrester-underway-to-catalog-the-white-label-social-networking-space/
9. Hendrickson, M.: 34 More Ways to Build Your Own Social Network (2007), http://www.techcrunch.com/2007/08/14/34-more-ways-to-build-your-own-social-network/

Connected Traveller, Social Web and Energy Efficiency in Mobility

Mikhail Simonov[1] and Gary Bridgeman[2]

[1] ISMB, Via P.C. Boggio 61, 10138 Turin, Italy
[2] ERTICO, Avenue Louise 326, 1050 Bruxelles, Belgium
simonov@ismb.it, g.bridgeman@mail.ertico.com

Abstract. Individual travelling is the most energy consuming day-by-day activity. Since fuel use is a type of consumer behaviour reflecting the interests to maximize some objective function, the human being activities seen in energy terms might be used to create the social aggregations or groups. Energy minimization in mobility conflicts with the objective function being maximized, however the virtual social networking by service-oriented architectures might improve the ecologic latter optimising the overall resources used by the community. Authors propose a method to build real life communities of connected travellers with the context awareness permitting to achieve some cooperative behaviour among the above-said virtual community networked.

Keywords: connected traveller, social network, user modelling, energy efficiency in mobility, time reasoning.

1 Introduction

Human being is social by its nature and consumes energy to achieve its goals. Therefore energy profiles reflect the human behaviour, and represent the lifestyle and social inhabits. Any mobility, undertaken to achieve individual goals in different geographic areas, can be virtualised using the machine-processable event collections, even if the approach adds some complexity, because of the consumable resources, achievable goals, needed energy, and pollution being generated being accounted. The artefacts created by humans and the carrying out individual travelling might have some social meaning, because reflecting the relationship with the community, a certain living standing and a social status. This social group membership might deliver benefits in energy terms, helping to save or share some energy or resources. The social grouping of those manifesting common patterns can be created using the similarity clustering techniques. Since energy use is a type of consumer behaviour reflecting the interests to maximize some goals, e.g. objective functions, the undertaking activities seen in energy terms might be used to create the social aggregations or groups. Humans influence the energy use through their individual or collective purchase decisions, and this can be used to move towards ecologic decisions in mobility. Thus energy demand management analyses the energy-using behaviour of consumers: individuals in households, travellers in mobility and other categories. Energy users are conscious

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 330–335, 2009.

and act in their own interests to maximize an objective function, overlooking the ecologic impact to minimize. The demand management considers the budget optimisation function because the energy is not free, and the ecologic motivation of the traveller gives rise to new aggregation of energy consumers maximizing some social objective function, such as the resource sharing until it contributes to the individual goals. The incomplete information available to guide decisions or market constraints will preclude the actions required that will maximise the individual or collective benefits, however some of the information not available to individual travellers might exist in travel agencies, enabling a global service summing up the partial views, creating the overall picture complemented by necessary elements for better decision-making. The connected traveller platform will build a virtual social aggregation of travellers capitalizing on the full information that is needed to achieve the temporal social cohesion and to minimize the consumable resources while travelling because of the resource sharing between users.

2 Behaviour Modelling

Energy domain researchers [1, 2] have produced several load shaping models through the adoption of the econometric, statistical, engineering and combined approaches. Load shaping of the different categories of users, such as industrial, tertiary, and especially of the residential ones, is a complex task [3], because it is linked to the lifestyle and psychological factors of the users, notably imprecise and subjective factors, while the definition of the standard behaviour of the various types of customers through statistical correlations does not solve the problem, failing to consider the variability of the demand, e.g. a random factor.

The human behaviour can be described by its objective function correlated with the goal's achievements. The set of resources used in travelling includes the material resources, energy, time, information and knowledge. The determination to reach the above-mentioned goals becomes the target, while the degree of the goal achievements, reflected by the objective function, determines the personal satisfaction in attaining the achievements. Any system optimising an objective function has to consider the functional limitations and the resource limitations. The resources are consumed to achieve the said goals, the human being activities fit the model, and the resource limitations reflect the consumption dynamics.

Let us define the Social Group of Connected Travellers (CT hereafter) as an aggregation of the human actors presenting the future common and complementary travelling patterns, with a potential to share some resources, because of the expected similar behaviour between travellers and the co-presence of them. The correlated travelling and energy profiles of the travellers, reflecting the needs and motivations of them contain the local energetic resources being consuming almost simultaneously. Adding a social objective function we optimise the material resources, the amount of energy and time, because the maximised information/knowledge makes possible the savings. The aggregation in social groups of human actors presenting the similar travelling patterns enable resource sharing and objective function minimization, but has no universal solutions, so the local optimum should be looked in real life by assessing the concrete situations, making i-Travel contribution relevant.

Let us assume two travellers – with similar decision motivations - arriving to the same airport almost simultaneously and travelling to the town separately because of the lack of information that can support cooperative behaviour. The proactive i-Travel agent might advise both of them about the possibility to share a vehicle saving some money and pollution: being similarly eco-sensible they might confirm through i-Travel the travel together. Therefore the quantification of the saved emissions becomes possible through i-Travel application. The common but volatile knowledge shared between two individuals creates a new CT entity, a virtual or real group of travellers, dynamically connected through i-Travel representative device. The proposed dynamic CT entity shows the presence of the direct interactions among members and the explicit semantic descriptions labelling their aggregation, which contributes in social cohesion. CT, seen from the social point of view, includes a group of people showing similar patterns in their travelling activities, and the willingness to cooperate in resource's savings. Geometrically, CT is a small collection of labelled points in the 4D time-space representing the future of "being there" and "being now" happening every 30 min. and related to some physical persons. The proposed social network of CT looks like the "maturation of the event space" along the time arrow, calculated in the hypothesis of the neutralized or absent external events, having the capability to change the cause-effect event chain. Travellers declaring their intentions to their global e-Agent populate a graph, enabling the temporal reasoning about their future co-presence, and this makes possible the cooperation. A number of travellers linked through i-Travel representative device become a social network sharing resources, energy and experience. The concept creates a virtualisation of the real individual profiles through a new entity manifesting different energy consumption patterns with better ecologic characteristics.

The CT social network keeps various relationships: we use it to emerge and to show explicitly the existing ones of the particular "connected" society, analysing the patterns and correlations in the digitised streams of expected events. We apply the concept of social network proposed by J. Barnes in 1954 to the travelling domain, in which the interactions among individual travellers consuming fuel/energy in real life conditions are totally absent: all travels are done individually, and the possible cooperation is random only, maybe because of the occasional conversation to others while travelling. The real time interactions are not available, so the social aggregation proposed keeps them apparently invisible and almost virtual as well. The Internet of Things gives a right online paradigm to obtain a virtualised social network between CTs. The inclusion criteria in the CT network is not evident, because of the temporal reasoning, however it can be elicited from the behaviour data since user modelling gives the knowledge, which can be made available by the soft computing methods. The above-defined CT social network is the real world entity of temporal neighbours.

Fuzzy Set theory [4] permits the gradual assessment of the membership of elements in a set, described with the aid of a membership function valued in the real unit interval [0, 1]. Fuzzy Logic [5] deals with reasoning that is approximate rather than precisely deduced from classical predicate logic, becoming a way of processing data by allowing partial set membership rather than crisp set ones, e.g. a problem-solving control system methodology to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information. The model is empirically based rather than on the technical understanding of the system. For example imprecise

terms like "IF (travellers are in_the_same_airport) AND (arrival_time is late_evening) THEN (travel_together_by_taxi)" are descriptive of what must happen. In order to deal with uncertainty, we represent it and reason about classifying the similar categories of the travellers. The system currently needs the rules to be formulated by an expert. Further reading is available in [6].

3 Proposition

The i-Travel project [7] has researched on a service platform for the connected traveller designing a personal travel assistant service for anyone planning and making a trip. The global service is adaptive and delivered through fixed or mobile terminals. A context-aware intelligent agent proactively manages events and/or requests, relying on the service offering from the e-marketplace back-end, on which a community of service and information providers can publish, negotiate and supply their services. Travel services offered according to a traveller's expressed preferences includes journey planning, route guidance, traffic and transport real time information, parking services, ticket booking and payment, as well as immediate news of any problems and the adaptation of a traveller's itinerary. Floating traveller data collection, or anonymous monitoring of individual travellers in order to deduce demand status, detect incidents and to monitor the status of public transport services in real time and over entire networks. A personalized service agent can anticipate a traveller's needs from knowledge of his current context. It pushes services as needed when the traveller arrives at milestones, or when disruptive events occur along the itinerary. The e-Marketplace exposes transport and travel services to be requested and consumed in real time, offering content and service to virtually all travellers. The adaptivity contributes in a global market for travel information and services negotiating the dedicated specialized local contents. The simplest travel might be the "flight" object already managed by airlines, giving the list of passengers and contact details of those expected to arrive in a given place simultaneously. This entity is not available because of the privacy constraints, and it lacks the final destination, the composition of the segments, and the willingness to cooperate, while the complete travel might be known to the travel agency as a "bouquet" of services requested at the booking time.

Let us define the digital travel as a collection of elements (1).

$$T(t_i) = \{ [\text{From}(\text{Place}_0, \text{Time}_0, \text{Mean}_0), \text{To}(\text{Place}_0, \text{Time}_0, \text{Mean}_0), \text{Attributes}(x_0, y_0, z_0, g_0)], \ldots \} \cup \text{Needs}(t_i) \quad (1)$$

The geo-referencing enables us to calculate the future occurrences, while the mean enable us to account the context and decide the possible aggregations. The overall architecture that services the i-Travel concept (Fig. 1) comprises the representative user device, which might be an in-pocket smart-phone with GPS localization and dedicated software running, plus the server side component elaborating the context, user needs, user-related events, the independent environmental events, and delivering the proactive service offerings.

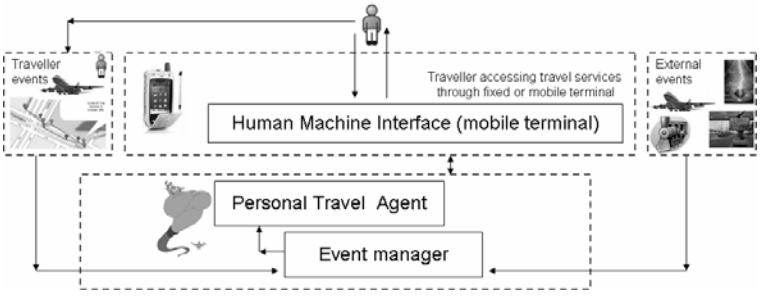


Fig. 1. i-Travel conceptual architecture

To perform the unification we estimate future presence points, calculating them in the pre-trip stage as the event chain of maturations in 4D space-time terms, e.g. all the expected locations and times for each traveller. For each of the future presence points we add the most likely actions, or the behaviours, formalized in the Traveller Domain Ontology as the rules. A similar example is described in [8]. A Fuzzy rule "IF (in-Airport) AND (night-time) THEN (taxi-use)" determines an intended behaviour likely undertaken by travellers. The time-space of the intended presences would be available in advance to calculate the co-presence of the users, making possible the proactive cohesion action in the pre-trip or in-trip stages. The knowledge server detects the overlap and expected co-presence in the same location, catching the chance to offer them to share a taxi, grouping to reduce emissions. Two components enable global connected traveller services: a digital travel accounting the process, and a method to build the dynamic community of CT, aggregating the market at run time.

i-Travel considers a collection of the travel plans by subscribed stakeholders. Any ***being now*** and ***being there*** overlap co-occurring become a challenge for the cohesion to be proposed after the verification of the real travelling conditions, the willingness to cooperate, and the absence/presence of the extraordinary events. The $M_j = \{N_i\}$ a cluster of end user's nodes aggregated together and logically seen as CT Unit becomes the dynamic social group. The research question we have tried to answer is **how** to aggregate nodes N_i - in real life populated by arbitrary elements becoming a physical neighbours because of the travel - logically optimising the cluster M_j in a way to group consumers likely to use the same resources, proposing them the sharing enabling the smart energy in mobility, the new market of connected traveller services.

Typically there are many events populating the event space, requiring the similarity clustering algorithms to process the data in order to find the possible co-presences. From the energy point of view the proposed approach uses a new mathematic model augmenting the information available to the respective users about their co-presence and about the same or similar resources. The new resource consumption will be qualified by aggregations $\mathbf{M_j}$ instead of the sum of consumptions by the single nodes $\mathbf{N_i}$, enabling the minimization of the energy because of the shared consumptions. The representative – virtual - nodes $\mathbf{M_j}$ exemplify the CT communities living shortly, but manifesting the new social identity of the aggregation and the collective behaviour in both social and energy terms.

4 Conclusions

Individuals manifesting similar travelling profiles can form temporarily a social network based on the cohesion criteria expressed by Fuzzy rules qualifying the energy for mobility patterns, becoming a new entity interacting on the travel market. Authors have described the computational method building a local virtual community while optimising the resource sharing and energy savings, obtaining virtually a cooperative colony. There are two enablers: digital travel as an instrument to account for the travel, and a method to build the dynamic community as a cohesion tool. The most important issue is the clustering technique creating the dynamic but optimised aggregation or community. The Fuzzy system becomes criteria for the social cohesion of the newly proposed community, while the aggregated profile represents the collective behaviour. The social aggregation proposed to physical persons is intentionally made virtual because of the temporal (volatile) nature of the possible transactions. However it might become real because of the motivation for grouping coming from the Fuzzy rules describing the objective function rewritten in lifestyle terms "people consistently flying to" instead of the event-related ones like "in BRU airport between 7 and 8 PM". There is no meaningful minimal or average grouping, ensuring the best local resource (fuel) consumption, however it appears the groups bigger than 3-4 units are unlikely cooperative. The further work is the implementation project of the i-Travel framework and its e-Marketplace.

Acknowledgments

The described work is undertaken from within the i-Travel project, co-funded by the European Commission and a consortium of industry and research organizations aiming to deliver key innovations to the travel and mobility market.

References

1. Chan, M.L., Marsh, E.N., Yoon, J.Y., Ackerman, G.B., Stoughton, N.: Simulation-based load synthesis methodology for evaluating load-management programs. IEEE Transactions on Power Apparatus and Systems 4(PAS-100), 1771–1778 (1981)
2. Broehl, J.: An end-use approach to demand forecasting. IEEE Transactions on Power Apparatus and Systems 6(PAS-100), 2714–2718 (1981)
3. EPRI, Electric Power Research Institute: Combining engineering and statistical approaches to estimate end-use load shapes: Methodology and results. Report EA-4310, Palo Alto, California, USA, vol. 2 (1985)
4. Zadeh, L.: Fuzzy Sets, Information and Control, vol. 8, pp. 338–353. Elsevier, Amsterdam (1965)
5. Baldwin, J.: Fuzzy logic and Fuzzy reasoning. In: Mamdani, E., Gaines, B. (eds.) Fuzzy Reasoning and its Applications. Academic Press, London (1981)
6. Halpern, J.: Reasoning about Uncertainty. MIT Press, Cambridge (2003)
7. i-Travel project, `http://www.i-travelproject.com`
8. Qiao, Y., Li, X., Wang, H., Zhong, K.: Real-Time Reasoning Based on Event-Condition-Action Rules. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM-WS 2008. LNCS, vol. 5333, pp. 1–2. Springer, Heidelberg (2008)

Designing Social Support Online Services for Communities of Family Caregivers

Matthieu Tixier and Myriam Lewkowicz

ICD – Tech-CICO – FRE CNRS 2848 – Université de Technologie de Troyes
12, rue Marie Curie, BP 2060
10010 Troyes Cedex, France
{Matthieu.Tixier,Myriam.Lewkowicz}@utt.fr

Abstract. In the recent years, more and more people turned towards the Internet to find support, share their experience and feelings when they live difficult situation such as diseases. Considering this phenomenon, we are working with a community of caregivers of patients suffering from Alzheimer's disease. We aim at providing them an online platform where they can share social support. Such a tool offers to overcome the limits of the current face-to-face practices within the community. Based on a deep study of social support practices and of these caregivers social capital, this paper presents our design approach of innovative information and communication services dedicated to social support.

Keywords: Community of Family Caregivers, Design, Social Support, Social Capital, Healthcare Network, Support Group, Alzheimer's disease.

1 Introduction

Social support is generally defined as an exchange of verbal and non-verbal messages which convey emotions or information and reduce uncertainty and stress [2]. Social support is often provided by relatives. However, the modern life style involves long-distance relationships, the decreasing size of families and the loosening of social ties. People therefore naturally seek other sources of social support outside the family, such as experts (social workers, psychologists, etc.) or peers involved in support groups or contacted via online discussion groups. The members of these groups discuss their experience and talk about the problems with which they are having to cope. All these exchanges lead to setting up communities based on shared experience [10].

Groups of this kind tend to be formed among people in distress because of problems such as serious diseases. They cater for for patients themselves or caregivers seeking for information, emotional support and tangible help [6]. When a seriously incapacitating disease is diagnosed, family caregivers often have no training for this difficult role. They have to understand and apply complex medical instructions, but lack the knowledge and skills enabling them to interact easily and efficiently with the patient. Few channels exist for expressing the distress they feel as family caregivers. This situation is all the more serious since caregivers are completely monopolized by their ailing relative and therefore have very little time to spare for information seeking, making an evolution, with hindsight, of their view of their practices, and social interactions in general.

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 336–347, 2009.

Our ongoing research project focuses on non professional caregivers looking after relatives with highly invalidating conditions such as Alzheimer's disease. We are interested in the emergence of communities of caregivers, the social support practices within these communities and Information and Communication Technologies (ICTs) for assisting and supporting these practices. As new trends about health on the Internet show, it seemed likely that ICTs could be used to enhance these caregivers' social capital and their feeling of membership of a community.

Social support is a complex activity which existed of course long before the internet and also exists outside the internet. The people involved in social support do not all make use of online forums. Although online forums enable people who are familiar with this kind of communication system to benefit from online discussions [13], we became aware that those who do not have the knowledge to use them are excluded from these services. It seemed likely, however, that by observing actual face-to-face social support practices, it would be possible to suggest some innovative ways of designing a system which would make it possible to reach people who would not normally have been users of online discussion forums.

The originality of our approach can be summarized as follows: (1) the communities we are interested in are not communities of patients [10,11,13], but relatives acting as caregivers; (2) a survey was conducted on a real-life caregivers' collective in order to define how to design an efficient tool for providing caregivers with social support. (3) Our design process is based on the translation of typical situations into online functionalities dedicated to social support. In this way, it is hoped to develop a design process yielding a tool which can be readily used even by people with little experience of computer tools.

In this article, after presenting the topic of interest (family caregivers dealing with Alzheimer patients), and suggesting how a ICT tool might help these people to cope with their distress, the results of a review of some existing systems dedicated to online social support are presented. Then, our proposal for designing a web-based system providing family caregivers with social support will be presented. Lastly the next steps in this work will be explored.

2 Family Caregivers for Alzheimer's Patients

Serious neurodegenerative diseases such as Alzheimer's disease (AD) greatly reduce the patients' autonomy as their cognitive abilities gradually decline. The patients become unable to deal with their own day-to-day requirements as the symptoms worsen, and their relatives have to assist them increasingly with even the simplest operations, such as shopping, bathing, and getting dressed.

Patients' relatives find it hard to cope with playing the role of caregiver for which they have not been prepared. As shown by several surveys on French caregivers' situation [9,17], apart from the financial cost of the disease (assistance with housework, adapting the home, etc.), it makes heavy demands on the supporting spouse or relative in terms of time and attention. Caregivers tend to be strongly committed to the support relationship with their ailing relative, whether they are the patient's spouse or child. The

assistance they provide takes up a large part of their time and energy and leaves the caregivers little opportunity of escaping and taking care of themselves.

Due to caregivers' lack of time, providing a web-based system available at all times seems to be a relevant response, since it would provide them with a space where they can share social support with peers. In addition, receiving social support on a daily basis enables people to talk about their problems immediately rather than letting worries accumulate for a long time before they find a sympathetic hear. This aspect is in line with the hypothesis that the benefits of social support provided by peers are mainly short-term benefits [18]. The anonymous communications mediated by the internet with people living elsewhere might motivate people who are reluctant or unable to participate in face-to-face support groups [20]. Online services of this kind would usefully complement the services already proposed by medical healthcare networks, as it would improve caregivers' access to information and help them cope with their distressing situation. In the following section, the healthcare network we are working with and our analysis of the caregivers groups it accompanies are both presented.

2.1 Healthcare Network and Support Groups

"Réseau Pôle Mémoire" (RPM) is a healthcare network which focuses on memory disorders such as AD. The network was launched in 2001 by a group of healthcare professionals in response to the many problems arising in the diagnosis and management of AD patients in the Aube region (N-E France). The main aim of the network is to coordinate the work of the many professionals (neurologists, general practitioners, social workers, psychologists) involved in the care and support of these patients. Secondly, it dispenses care by performing neuropsychological screening activities (cognitive testing and diagnosis) and follow-up with patients and their caregivers. The network is also responsible for informing professionals and the public about memory disorders and their treatment. For this purpose, it organizes training sessions for professionals and family caregivers, and medical symposia for professionals and publishes documents targeting various audiences (such as booklets and web site documents). The RPM also runs monthly support group meetings for spouses and children who are caregivers in order to provide them with a space where they can talk about their experience and discuss their problems.

The spouse caregivers' support group, which has been in existence for about a year, mainly includes caregivers who have attended a session of RPM training at which health professionals explained how they contributed to the care and support of AD patients.

The children caregivers' support group has been in existence for over two years and closer relationships have been formed among its members, although the assistance with which they provide to their parents is of various kinds and the burden they carry differs from one person to another.

The existence of two separate groups corresponds to common practices [18,23]. The two care-giving situations and the relationship with the recipient of the care are very different (in terms of whether the caregivers live with the patient, their age, whether or not they have a job, etc.). The members of the two support groups have no

contact with each other, although there are no hard and fast rules about this point. Each group consists of about fifteen members who attend some or all of the regular meetings held on the first Friday of every month. The afternoon session is intended for spouses who are caregivers and the evening session, for children who are caregivers. The program of these support group sessions is roughly the same each time: the caregivers meet around a table at the RPM office, where they are served with cakes and drinks, which are sometimes provided by the caregivers themselves. The ensuing discussions are led by the network coordinator, who is a psychologist. She gives everybody an opportunity to speak in turn and dispenses accurate medical information. In the following section, it is proposed to use the concept of social capital to analyze these collectives.

2.2 Communities of Family Caregivers

We have used an analytical grid of social capital developed by [8] in order to give an overall picture of the RPM caregivers' community. This grid is based on three dimensions in line with several studies [1,14]. In this frame we adopt a definition of social capital grounded in the work of Pierre Bourdieu ([3] in [14]), where social capital is defined as the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit. However we draw the reader's attention to the point that more does not necessarily imply better. Social capital is a complex notion and its positive or negative impact on a social setting is mainly contextual [19], for example having a lot of information exchanges could discourage newcomers to join the network, or strong ties could be alienating. With this concern in mind, we use the descriptive grid mentioned above which can be summarized as follows:

- The structural/opportunity dimension concerns the relationships existing among community members thanks to the structure of the social network, which provides an opportunity for members to share resources.
- The cognitive/ability dimension focuses on the information and resources that circulate within the network. It also includes the ability of people to make use of the resources available to them via the network.
- The relational/motivational dimension can be described as the strength of the ties formed, the conventions and norms that are shared and negotiated by the community members, and the question of trust.

From the structural point of view, the two RPM caregivers' groups, which include about 30 persons in all, correspond to only a small proportion of the 600 patients currently catered for by the RPM. Some of the support group members occasionally call each other between group meetings, but most of them have no direct or indirect contacts apart from their monthly meeting at the RPM office.

At the cognitive level, the caregivers have a lot to say about their experience with their ailing spouse or parent and about their daily care practices, and they often give each other advice. They express their opinions about institutional issues (nursing homes, insurance policies, etc.) and medical practitioners. Spouses who are caregivers often refer to the training they underwent at the RPM. They mention the knowledge

and the benefits of the hindsight thus gained about their spouse's disease. The treatment available and the latest scientific knowledge about these diseases are frequently discussed during the meetings chaired by the RPM coordinator, who represents an authority on the subject.

One of the main differences between the two groups focuses on the relational aspects of their social capital. Although the experiences and events they talk about are sometimes emotionally quite intense, the level of mutual listening was not found to be very high in the spouse caregivers' group. The many digressions which tend to occur oblige the coordinator to intervene frequently and to manage the caregivers' interventions more than with the other group. The children caregivers' support group seems to be a much more closely knit group. They discuss the crises and the upsetting episodes they have experienced with the other members of the group. They also talk about the friendly events they have enjoyed together (such as the New Year party, members' birthdays, etc.) at their monthly meetings at RPM office. These caregivers show great attachment to the group and some of them are still attending the meetings although their parents or spouses are no longer alive.

In the light of the above findings, it can be said that the RPM caregivers' community, which actually consists of two separate groups, is already emerging within the institutional structure in which these interactions were initiated. The RPM plays an active role: it organizes meetings and support groups, initiates discussions, and creates contacts between caregivers. Developing the opportunities for caregivers to communicate and to discuss their practices together via the online services proposed would therefore help to develop a wider and more active community, which would hopefully result in a "virtuous circle". This is an evolution process as such mentioned by several authors, for instance on the social embeddedness of community tools [8] or the dyads and groups dynamics in social networks [5].

In the following section, a set of web sites providing information and communication services of this kind will be analyzed. These particular platforms were selected either because they explicitly claim to provide social support or because their focus (healthcare, social problems, etc.), their contents and the communication facilities they provide are relevant to develop social support practices.

3 Existing Web-Based Solutions Providing Social Support

Thirteen websites were selected for this study[1]. It was attempted to include equal numbers of research projects (Hutchworld [4], Krebsgemeinshaft [7], CHESS [11]), classical discussion forums (5), and web2.0 generation platforms (5) showing features typical of most social networking systems. Whenever possible a user account was created on these sites. It was thus possible to test normal users' experience of the functionalities, contents and presentation of most of these sites. We started by assuming that online social support services consist essentially of interactions between users. On this basis, we described the functionalities that enable users to have

[1] The complete list of websites and the data set on which this study is based are available at the URL: http://www.orkidees.com/missWiki.

discussions via the platform. We focused in particular on the multidimensional aspects of these mediatised communications [24], i.e., whether they were public/private, synchronous or asynchronous communication systems and whether the messages were intended for someone in particular or not. It was then attempted to group the items identified into larger functional components common to several sites.

27 functional components were identified that were common to the platforms studied. Some functionality can be part of larger functional entities that serve more specific goals. These functional components were then subdivided into "elementary" and "complex" components (written as follows: component_1 x component_2 x component_i). The components that figures in *italics* are common to all of the studied platforms.

Table 1. "Elementary" functional components

Functional components	Description
Assessment	A system that calculates users' activities on the base of specific criteria. Contents are either rated by users, either using a predetermined list of items (which are often strictly positive or qualitative) or users are free to add their own qualitative comments.
Profile	A personal page on which users can present themselves in a variably flexible frame.
Role	Members adopt a specific role (e.g., that of moderator, administrator or honorary member), which can determine their right of access to the platforms' functionalities and the extent to which they are to visible to other users.
Awareness	Enables users to share information about their status or about what they are doing at the moment, or to talk about their moods and feelings.
Anonymity, confidentiality	Systems that enable users to control the level of visibility and the range of information they are willing to share on the platform.
Friendship, relationship	Users can create formal relationships with other users (generally in the friendship mode).
Users' groups, communities	Enables users to constitute groups, and generally to have access to a private space with dedicated functionalities managing the group activity.
Internal messaging system	A private, asynchronous targeted communication system, which is internal to the platform.
Instant messaging	A private, synchronous targeted communication system.
Chat group	A public synchronous communication system.
Forum	A public asynchronous communication system.
Digest	An event handling system which enables the system to display or send by mail to a user digests about her friends' activities, the activities of users belonging to her groups and about the activities taking place on the platform.
Search engine	A search engine gives users transversal access to the contents on the platform.
Bookmark system	Enables users to pick-up and follow interesting discussions, messages and contributions.
Emoticon	Users can express their feelings and moods via emoticons.

Table 2. "Complex" functional components

Functional components	Description
Exchange of advice (Forum x Evaluation)	Enables users to exchange advice, to discuss it and give their opinion about it.
Questions and Answers system (Forum x Evaluation)	Enables users to explicitly send a question to the community. Members' answers are generally assessed.
Stories, experience sharing (Forum x Users' groups, community x Evaluation)	Users can share their experience and tell stories as with a blog system, where other users can react by adding comments and other forms of assessment. Common themes occurring in users' contribution are highlighted with tags or other systems.
Wall, personal guestbook (Profile x Forum)	An asynchronous, public targeted communication system.
Hugs, gestures, gifts (Profile x Internal mail system)	Users can exchange gifts or conventional attention marks (i.e. hugs, tributes) between members and add a short message.
Goal management (Self-monitoring x Hugs, gestures, gifts)	Users can set goals and be given support to help them reach these goals via the platform (e.g., quitting smoking). They can show how they are getting on as time goes by (in terms of days, weeks or months). The other users receive incentives from the platform in order to make them encourage the person to succeed. They can also say that they share the same goals and are beginning the same process.
Self-monitoring (Evaluation x Profile x Awareness)	A system that enables users to regularly give details about a detail that concerns them (such as physiological variables, weight and mood). This enables them to have a picture of how their state is evolving and to share this information in graphic form with other users.
MatchMaking (Profile x Search engines)	A system that proposes to match user with other users with similar interests, goals or experiences. These systems either perform at the user's request or passively on the base of their profile.
Signature *(Role x Profil)*	Enables users to add their signature to their contributions, often with a slogan or an invitation to contact them.
Group Awareness *(Use s' groups, community x Awareness)*	A system that informs users about each other's presence online: a list of connected members, a list of the participants in a discussion, etc.
Notifying contents to administrators *(Internal mail system x Role)*	Enables users to warn administrators about contents, usually because of their litigious nature.

These groups of functionalities which are common to several platforms show the existence of a general trend among complex functional components: several functionalities tend to be combined in order to provide more specific situations than those available in classical discussion forums. Some platforms allow users not only to exchange messages, for example, but also to ask or answer questions, to give or receive advice and to share experience.

The innovations observed on these platforms, which mostly result from research projects or the web 2.0 generation, are not based in our opinion on novel technical solutions. 3D or video contents were quite rarely encountered on the platforms we studied. Emblematic technologies such as AJAX have essentially improved the fluidity of requests transfer and the updating of contents. At the end there were the similar text fields and other checkboxes already used at the start of the internet which support the core of these platforms. Our findings show that the systems studied tend to rely on

combinations of elements (e.g., an asynchronous communication system combined with assessment functionalities), in line with a model for innovation which resembles to Schumpeter's description ([22] in [14]), in order to create functionalities which encourage the users to refer to normal face-to-face communication situations.

The designers of these websites have taken concrete everyday situations as their starting point and *translated* them into innovative functionalities. For example, they have translated experience sharing situations by combining an asynchronous communication tool with a function enabling users to comment on the stories they read via a system of assessment. We have emphasized this idea of translation, which often underlies the so-called "designer's insights", in order to describe some of the principles on which this process is based. The problem therefore consists in identifying the most relevant situations occurring in everyday practice and defining the signs (i.e. the words) which enable users to recognize the situation in question.

4 Design Process

A critical issue involved in developing online communication and cooperation tools is that they should make it possible to perform activities such as social support activities that are already being carried out offline without the mediation or even the existence of computers or the internet.

When attending support groups and meeting caregivers, we detected some recurrences in the everyday practices observed: the way the people performed round-table where they are asked to speak in turn, the way the coordinator began the sessions, and the way the caregivers introduced themselves, talked about their problems and asked questions. The recurrent features of these situations seemed to be particularly relevant to designing social support platforms. It seemed to be worth attempting to *translate* these features and these situations when designing a tool dedicated to social support. The aim was therefore to encourage the users to perform these typical situations online, since they are familiar to them.

The results of our review of existing social support platforms illustrate this idea. One example of an online function encouraging the users to perform, through the system, real communication situations which they are familiar to is the questions-and-answers (Q&A) function: question-and-answer games are implicitly based on common experience. Q&A functionalities are typical for many social support websites and interest many other application domains [15]. By using them instead of more classical internet systems such as forums, users will be given a more intuitive grasp of the communication situation proposed. They will be able to use the general scheme of the situation, the script [21] they are familiar with, to guide their interactions with the system and with other users.

When online communication situations resemble those occurring offline, users are able to intuitively organize their interactions. Some features of the situation naturally differ in the case of computer-mediated communications and users have to renegotiate part of the process, especially when the tool includes new possibilities that do not exist in real-life situations, but the interaction management load will still be lighter for them in the end.

The problem is how to design functionalities on the lines of the scripts underlying the corresponding real-life situations. It was therefore proposed to *translat*e these typical situations and the scripts intuitively recognized by users in designing our platform. To do this, it was necessary to specify this process of *translation*.

The first step in the process of translation involves identifying the words, signs and symbols used in real-life situations in order to use them online by displaying them on the users' screen through the system's interface. This will help users to recognize the situation of communication proposed through the system. This is in keeping with the use of the notion of convention in Human Computer Interface (HCI) design as presented by Norman [16]. In line with him we think signs and symbols on screens relate to conventions, shared between users and designers for example. Conventions serve as incentives to promote interactions between users. In this case, the conventions helped us to select suitable semiotic cues with which to design functionalities stimulating users to perform typical situations of communication through the platform.

Among the typical situations observed, that arising when the members of support groups are *introducing themselves* provides a good example of how conventions are used. Caregivers generally present themselves to newcomers at support group meetings. The first semiotic cues relating to conventions were naturally the words and expressions used by the participants to refer to the situation: "Ils se presentent". We also detected several semiotic cues in the way caregivers describe their own situation using specific signs, words and locutions. They do not speak in terms of AD patients or the ailing person they are caring for, but about their *husband or wife*, *mother* or *father* whose *disease* (i.e. Alzheimer's disease) *was diagnosed X months/years ago*. They mention whether their relative was *living in a nursing home,* giving the name of the institution. They also describe the care with which they provide their ailing relative.

This raises the question about the conventions of introducing oneself at the support group. In order to design *profile pages* which correspond to the reference situation, it is necessary that the text-fields it is proposed to display should match the stages in the scripts followed by caregivers in real-life situations. But it is also essential to use the same signs, words and expressions as caregivers use in real life, in order to guide the users to perform the situation where they are presenting themselves to the community.

Users will therefore be prompted by "Introduce yourself", to state for instance "*my husband* is suffering from *Alzheimer's disease* that was diagnosed *18 months* ago" (information in italic could be chosen by the user among several possibilities on the system).

We then have to address a second level of *translation* that concerns the ability of the system to promote a set of interactions that roughly correspond to those in which the users engage in real life.

Another typical situation we observed at support groups was the "round-table" where caregiver's are invited to speak in turn to give their opinion on a specific topic or recounting their experience. All the caregivers who attend support groups have a rough idea of the script describing the conventional course of this relatively formal situation.

The coordinator initiates round-table discussions by proposing a question or a topic on which participants are invited to express themselves (i.e. nursing homes). She refers more or less directly to the communication situation "Could each of you tell us

about your experience with nursing homes?". The coordinator checks whether she has the group's attention and assent. She then suggests which of the participants should speak first: "Mr/Mrs …, what do you think about…" or "How is your wife doing at the … … nursing home?"

The caregiver speaks about the subject.

Other participants can react.

The coordinator can either press the speaker to give more details, or she can reframe the discussion if digressions occur. The coordinator can cause a digression herself if she notes an interesting point in the caregiver's discourse that she wants to bring to the attention of the entire group.

Once the first caregiver has finished speaking, she can either invite the next participant to speak by giving a sign or tell the coordinator that she has nothing more to say.

The round-table discussion ends when all the participants have spoken or it is time to close the meeting.

In this round-table situation, participants expect to be invited to express themselves on a topic on which all the other participants will also speak in turn. They also expect to obtain reactions and comments from other participants. The "round-table script" can therefore be used to define a functionality that could be implemented in an online social support platform as follows: users could initiate a round-table discussion by sending a message presenting a theme to a group of peers. Several contributions could be collected on the same topic and each participant could comment on peers' contributions in the same way as on a blog. The idea is naturally not to reproduce exactly the same script, according to the new possibilities offered by ICTs, but to promote a series of interactions that will meet users' expectations about the situation. The round-table functionalities adopted will therefore be based on an asynchronous communication device, since caregivers are not able to participate simultaneously because they have little time to spare. The initiator could close the round-table discussion once all the participants have made a contribution.

We have discussed separately each of the principles on which the translation of typical situations into functionalities should be based in order to make this presentation as clear as possible. Conventions and scripts are both necessary to completely specify the functionalities presenting situations which are recognized by users and meet their expectations about how they are going to interact with peers via the online website.

5 Conclusion and Future Perspectives

In this paper, we have presented the communities we are interested in, which are communities of family caregivers looking after relatives with severe diseases, especially AD. After describing their burden and the role ICTs could play to help alleviate their distress by providing them with greater social support, some existing web-based systems of social support were reviewed. This analysis led to the conclusion that there is an interest to design online systems for social support which resemble real-life social support practices more closely. By attending meetings of the support groups run by this network, it was possible to identify situations and scripts promoting

interactions between members. A design framework was then suggested for *translating* these observations into functionalities on the internet social support platform. By implementing this design framework, it is hoped to design a system which will resemble caregivers' actual practices as closely as possible, and therefore provide an easy to use and helpful tool.

The next steps in this study will consist first in integrating the findings based on the interviews carried out by our sociologist colleague. We will be able to extend our reflexion to address questions as privacy with these findings. The combination between the observation of support groups and the findings made in these interviews should yield a detailed picture of existing social support practices in both face-to-face and online situations, as well as providing additional information about family caregivers. Our second objective is to complete our description of the relevant scripts. These descriptions have been based so far on the observations carried out during face-to-face interactions. We now plan to study online social support exchanges (in forums) in order to define a kind of communication contract to which users of social support forums implicitly conform. We will then again attempt to combine this communication contract with the scripts on which face-to-face meetings are based in order to obtain a broader picture of how online interactions could provide efficient social support [12]. Thirdly, having completed our close analysis and definition of social support practices, it is now proposed to develop a web-based platform dedicated to family caregivers, based on the framework described above. We will present this platform to the family caregivers in the RPM, but it will also be available on the Web to caregivers outside this particular healthcare network.

Lastly, this platform will have to be assessed, first as regards its usability, and secondly from the social capital point of view. The concept of social capital will help us to analyse the evolution of family caregivers' communities using an online social support system. It will then be possible to assess our proposal to extend the existing social practices by introducing an online system. One of the issues on which it would be interesting to focus is that of technological tools as catalysts for communities.

Acknowledgments. This research was conducted with the support of Conseil Général de l'Aube in the framework of a UTT strategic programme.

References

1. Adler, P.S., Kwon, S.: Social Capital: Prospects for a New Concept. The Academy of Management Review 27, 17–40 (2002)
2. Barnes, M.K., Duck, S.: Everyday communicative contexts for social support. In: Burleson, B.R., Albrecht, T.L., Sarason, I.G. (eds.) Communication of social support: Messages, interactions, relationships and community, pp. 175–194. Sage, Thousand Oaks (1994)
3. Bourdieu, P.: The forms of capital. In: Richardson, J.G. (ed.) Handbook of theory and research for the sociology of education, pp. 241–258. Greenwood, New York (1986)
4. Cheng, L., Stone, L., Farnham, S., Clark, A.M., Zaner, M.: HutchWorld: Lessons Learned-A Collaborative Project: Fred Hutchsinson Cancer Research Center & Microsoft Research. In: Heudin, J.-C. (ed.) VW 2000. LNCS, vol. 1834, pp. 12–23. Springer, Heidelberg (2000)
5. Coenen, T.: Structural aspects of online social networking systems and their influence on knowledge sharing. In: Proceedings Web Based Communities. San Sebastian, Spain (2006)

6. Gottlieb, B.: Social Support and the Study of Personal Relationships. Journal of Social and Personal Relationships 2, 351–375 (1985)
7. Gustafson, D.H., Hawkins, R.P., Boberg, E.W., McTavish, F., Owens, B., Wise, M.: CHESS: 10 years of research and development in consumer health informatics for broad populations, including the underserved. International Journal of Medical Informatics 65, 169–177 (2002)
8. Huysman, M., Wulf, V.: IT to support knowledge sharing in communities, towards a social capital analysis. Journal of Information Technology 21(1), 40–51 (2006)
9. IFOP: Etude nationale Connaître les aidants et leurs attentes (2008), http://www.aveclesaidants.fr/index.php?rub=alaune&ssrub=enbref&lid=522
10. Josefsson, U.: Patients' Online Communities Experiences of Emergent Swedish Self-help on the Internet. In: Huysman, M.H., Wenger, E., Wulf, V. (eds.) Proceedings of the First Communities and Technologies Conference C&T 2005, pp. 369–390. Springer, Heidelberg (2003)
11. Leimeister, J.M., Krcmar, H.: Acceptance and Utility of a Systematically Designed Virtual Community for Cancer Patients. In: van der Besselaar, P., de Michelis, G., Preece, J., Simone, C. (eds.) Proceedings of the Second Communities and Technologies Conference C&T 2005, pp. 129–149. Springer, Heidelberg (2005)
12. Lewkowicz, M., Marcoccia, M., Atifi, H., Bénel, A., Gaglio, G., Gauducheau, N., Tixier, M.: Online Social Support: Benefits of an Interdisciplinary Approach for Studying and Designing Cooperative Computer-Mediated Solutions. In: Proceedings of the 8th Conference on the Design of Cooperative Systems COOP 2008, pp. 99–110 (2008)
13. Maloney-Krichmar, D., Preece, J.: A Multilevel Analysis of Sociability, Usability, and Community Dynamics in an Online Health Community. ACM Transactions on Computer-Human Interaction 12, 201–232 (2005)
14. Nahapiet, J., Ghoshal, S.: Social Capital, Intellectual Capital, and the Organizational Advantage. The Academy of Management Review 23, 242–266 (1998)
15. Nguyen, D., Thompson, S., Hoile, C.: Hubbub – An innovative customer support forum. In: BIS 2008 Workshop Proceedings, 2nd Workshop on Social Aspects of the Web (SAW 2008), pp. 55–67 (2008)
16. Norman, D.A.: Affordance, conventions, and design. Interactions 6, 38–43 (1999)
17. PIXEL Novartis: Etude PIXEL - L'entourage familial des patients atteints de la maladie d'Alzheimer (2000), http://www.mediathequenovartis.fr/novartis/spip.php?article107
18. Pillemer, K., Suitor, J.J.: Peer Support for Alzheimer's Caregivers: Is it Enough to Make a Difference? Research on Aging 24, 171–192 (2002)
19. Portes, A.: Social Capital: Its Origins and Applications in Modern Sociology. Annual Review of Sociology 24, 1–24 (1998)
20. Salem, D.A., Bogat, G.A., Reid, C.: Mutual help goes online. Journal of Community Psychology 25, 189–207 (1998)
21. Schank, R.C., Abelson, R.: Scripts, Plans, Goals, and Understanding. Erlbaum Assoc., Hillsdale (1977)
22. Schumpeter, J.A.: The theory of economic development: An inquiry into profits, capital, credit, interest and the business cycle. Harvard University Press, Cambridge (1934) (reprinted in 1962)
23. Thomas, P., Hazif-Thomas, C., Lalloué, F., Preux, P.-M.: Dementia patients caregivers quality of life: the PIXEL study. International Journal of Geriatric Psychiatry 21, 50–56 (2006)
24. Xie, B.: Multimodal Computer-Mediated Communication and Social Support among Older Chinese Internet Users. Journal of Computer-Mediated Communication 13, 751–767 (2008)

SDS-SOA 2009 Workshop Chairs' Message

Konstanty Haniewicz[1], Monika Kaczmarek[1], Maciej Zaremba[2], and Dominik Zyskowski[1]

[1] Poznań University of Economics, al. Niepodległości 10, 60-875 Poznań, Poland
{k.haniewicz,m.kaczmarek,d.zyskowski}@kie.ue.poznan.pl
[2] National University of Ireland, Lower Dangan, Galway, Ireland
maciej.zaremba@deri.org

We are very pleased to present the proceedings of the First Workshop on Service Discovery and Selection in SOA Ecosystems (SDS-SOA 2009). The workshop was held on the premises and with support of Poznan University of Economics, Poznan, Poland, on the 29th April 2009. The workshop was organized in conjunction with 12th International Conference on Business Information Systems (BIS 2009).

The main aim of the workshop was to gather researchers and practitioners to present their original work in the topics of service interactions within service ecosystems with a special focus on the interactions such as dynamic service discovery and selection.

Service oriented architectures enable service compositions and facilitate business processes reengineering. What is more, the SOA paradigm also enables a multitude of service providers to offer on service markets loosely coupled and interoperable services at different quality levels. This creates a unique opportunity for businesses to dynamically discover and select, in a cost-effective manner, services that meet their business and quality needs. However, outsourcing of IT services to external vendors causes dependency and introduces new challenges. In particular, this applies to the quality of outsourced services. If a mission-critical service is sourced out to an external vendor, provisioning of the required service and its quality is beyond immediate control of the service customer. Therefore, the success of the companies depends heavily on the ability to discover and select the best services in the given context for the needs of business processes. Viable choice has to take into account both functional and non-functional characteristic of a service.

Three papers from a set of excellent submissions were selected for publication by the Program Committee in a peer review process and reflect truly international interest in Web service interactions and SOA ecosystems. We would like to thank all members of the Program Committee for their thoughtful and detailed reviews of the papers and fruitful collaboration.

The topics of the accepted papers cover a broad spectrum of topics starting from SLA issues, to Web services selection methods.

The first paper focuses on reliable description of service non-functional properties. Namely, André Ludwig and Marek Kowalkiewicz in their paper “Supporting Service Level Agreement Creation with Past Service Behavior Data” propose an interesting approach of utilizing data from past service behavior to enrich and refine definitions

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 348–349, 2009.

of SLAs. This work describes the method of utilizing service profiling mechanism deriving information on services based on monitoring data as a source of information for analysis and prediction of SLA attributes. The obtained service description may be then utilized during service discovery and selection.

The selection of services for the needs of business processes based on service description is the main topic of the second paper. In their paper on "QoS-Aware Peer Services Selection Using Ant Colony Optimisation" Jun Shen and Shuai Yuan present a new approach to Web service selection problem. The authors propose to apply Ant Colony Optimization algorithm to identify the optimal configuration of services. Also, a study of algorithm's efficiency is presented giving reader a great opportunity for observing how hive-wise optimization can be applied in the domain of Web services.

The selection algorithm operating on a service description available in service registries like e.g. UDDI was the main topic of the third paper. Colin Atkinson, Philipp Bostan, Gergana Deneva and Marcus Schumacher present in their paper titled "Towards High Integrity UDDI Systems" an extension of UDDI architecture. The paper explains the failure of public UDDI registries by showing three main shortcomings of that technology. Authors propose appropriate extensions in order to address the lack of automated management of UBR content, lack of QoS monitoring of registered Web services and over simplicity of search process within UDDI. They show enhanced UDDI architecture and present an implementation of it based on well known and publicly available tools.

Towards High Integrity UDDI Systems

Colin Atkinson, Philipp Bostan, Gergana Deneva, and Marcus Schumacher

Institute of Computer Science, University of Mannheim
68131 Mannheim, Germany
{atkinson,bostan,deneva,schumacher}@informatik.uni-mannheim.de

Abstract. The basic idea of a ubiquitous service market where services can be published by their providers and discovered by potential users is – besides increased interoperability – one of the driving forces behind the SOA vision. However, the failure of the famous UDDI Business Registry demonstrated that service-orientation per se does not solve the problems related to the operation of such service registries – a problem well-known from the component-based development community. On the contrary, since services are highly volatile it is even more difficult to effectively manage a repository. In this paper we discuss the challenges associated with the operation of service brokering solutions with a clear focus on automation and integrity of the overall system as well as the offered search capabilities. To address the identified challenges we propose an extended UDDI architecture.

Keywords: UDDI, Service Discovery, Web Services Quality.

1 Introduction

Over the last few years SOA has become the dominant approach for designing and implementing distributed enterprise computing systems as well as business processes. Today, the main question facing enterprise system developers is not whether to use SOA but how. The success of service-oriented architectures to date stems primarily from its effective support for service interoperability rather than its promotion of service discovery or reuse. Two of the three fundamental SOA standards (WSDL, SOAP) are essentially focused on the former aspect, while the third one (UDDI) is responsible for the latter.

Despite its ubiquity, one key ingredient of the SOA vision has failed to live up to expectations and can be viewed as something of an "Achilles Heel" – this is the service brokerage model which is supposed to allow service consumers to find services as well as service providers and to provide the basic foundation of service marketplaces. As the well known SOA triangle in Figure 1 depicts, the core standard that supports this aspect of SOA is UDDI (Universal Description, Discovery and Integration). Although service registries are often used intensely within enterprise boundaries, so far all attempts to set up publicly accessible UDDI-based service brokers have not been very successful. The most well known example is the so-called UDDI Business Registry (UBR), operated jointly by SAP, IBM and Microsoft. This public available registry was finally shut-down in early 2006 because it contained hardly any usable content [1].

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 350–361, 2009.

While the design principles of the standard Web service brokerage model [2] presented in Figure 1 are basically sound, we believe that the current UDDI brokerage model includes three inherent weaknesses that make it difficult to set up an effective Web service registry solution. These issues also relate to the driving ideas behind SOA – the automation of service discovery, integration and execution.

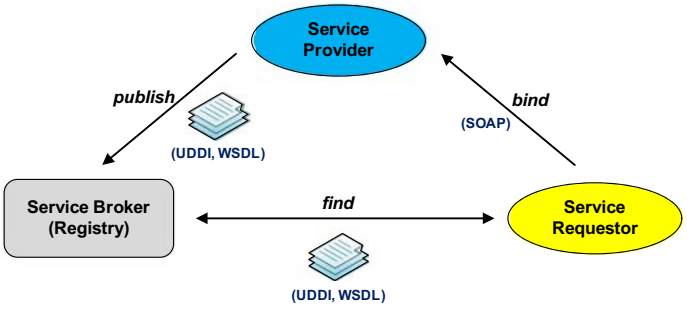


Fig. 1. Standard Web Service Brokerage Model [2]

The first weakness stems from the fact that all information about services in a UDDI registry is supposed to be provided manually. More specifically, to register a service with a UDDI registry, a human has to input a collection of metadata for the so called white, yellow and green pages of the UDDI standard.

The second weakness relates to the maintenance and data integrity of services once they have been registered with the system. Again this is a function that human service providers are expected to perform for the full lifetime of the service. They have to keep information for registered services up-to-date and they have to manually remove registered services once they have been shut down. In 2008, Al-Masri and Mahmoud recognized in their investigation of current Web Services on the World Wide Web [3] that many of the Web service descriptions that can be found on the web point to services that are no more active. As has also already been observed with component repositories by Poulin in [4], once a collection grows beyond a certain size it is no longer manually maintainable nor manually browsable.

This leads directly to the third weakness – the limited search capabilities offered by UDDI that are restricted to browsing the registry or using the inquiry API [5] which supports only simple keyword-based searches on UDDI entities. As demonstrated by the directory-based approaches of most early web search engines, this approach was quickly overtaken by the "crawl and search" approach of today's search engines.

The net result of these weaknesses is that the overall quality of the search results delivered by public UDDI registries is very poor. Over time, such registries tend to be overwhelmed by services that offer only poor performance, are no longer maintained, or are even no longer available. As a result, potential service users are forced to apply elaborate "trial and error" strategies for testing the discovered services in a real usage scenario to assess if they fulfill all expected requirements.

In this paper we introduce an extended UDDI server architecture that addresses these weaknesses and that significantly enhances the apparent integrity of the stored information. The remainder of this paper is structured as follows. In section 2 we first

introduce our approach for an extended UDDI architecture. In sections 3, 4 and 5 we then give detailed information about the enhancements and the mechanisms that address the weaknesses that have been identified. Section 6 discusses related work and we provide a final conclusion and an outlook on future work in section 7.

2 Extended UDDI Architecture

To address the introduced weaknesses we extended the basic UDDI architecture, its inquiry API and its respective interfaces to provide enhanced support for automated registration, higher data integrity and more sophisticated search capabilities. The primary goal of the approach and its implementation which are presented in this paper is to remove the weaknesses while still conforming to the UDDI specification and its recommended programming API and to be fully transparent to users.

Our reference implementation is based on *jUDDI* [6] which offers an open-source implementation of the UDDI specification, including a registry and respective UDDI APIs, and is part of the Web services project at the Apache Software Foundation. Our extended UDDI server architecture is not a node in the sense of common UDDI registries since it is not used to replicate its data within a UDDI ring, as the original concept proposes [5]. Moreover, the presented approach is intended to be a vehicle for supporting public stand-alone and enterprise UDDI servers.

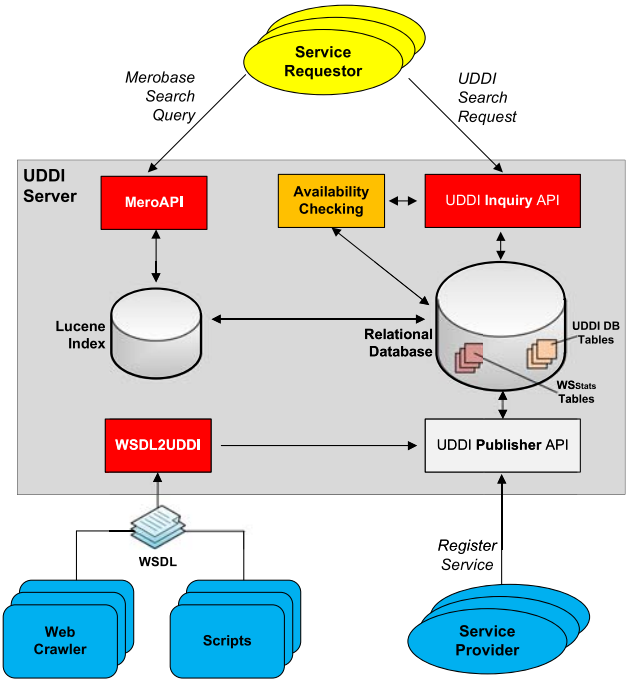


Fig. 2. Enhanced UDDI Architecture

The main extensions that we propose to the UDDI architecture are presented in Figure 2. These mainly consist of a richer UDDI data model supporting availability and quality measures, additional processing components together with database-oriented support for quality assessment and enhanced search capabilities. As also depicted in Figure 2, our enhanced UDDI architecture offers two interfaces for the registration of services. The first interface is represented by the well-known UDDI publisher API that allows service providers to register their services manually, using either programmatic solutions like UDDI4J or JAXR, or web-based tools like the Novell NSure UDDI Client, for example. The second interface that we additionally added provides automated registration of Web services. This may be realized using web crawlers or scripts (e.g. to iterate over an existing enterprise UDDI registry) that find and register multiple Web services sequentially. With the automated approach the only information that is available for registering a service is contained in its WSDL description. Thus, the registration process requires an automated mapping of the WSDL data structure to a UDDI-conformant data structure, which is realized by the *WSDL2UDDI* component of our architecture. After this step of pre-processing a service description and its respective data can be added to the service registry as before using the common UDDI publisher API.

The most important of our extensions facilitates the automated quality assessment of Web services that is realized by the *AvailabilityChecking* component and additional database support. This extension is responsible for evaluating registered services at periodic time intervals to assess information related to availability and other quality measures. This mechanism requires a separate database table (*WSStats*) that keeps track of the registered services and stores the information needed to assess their quality. Each time a service is added over the publisher interface a new record is inserted into the respective table, while binding templates are also replicated in the UDDI database table and are used as the primary key. The availability and quality assessment is based on load tests which are periodically executed on all registered services to determine when (or whether) services are available and to collect statistical values that provide a long term view of the Web service's performance. Several of the determined quality values are stored in the database and are also replicated in Lucene[1] [15] that represents another form of database using an inverted index where each service and its properties are represented by a single document.

To support more sophisticated search capabilities, our extended UDDI architecture generally offers two kinds of search queries. First, the UDDI inquiry API can be used in the usual way, supported by automatic filtering of services that are currently not available. This has been achieved by enhancing the implementation of the (*jUDDI*) inquiry API with access to the database that keeps track of availability and statistics information for registered services. Second, our architecture offers an interface to the embedded Lucene index over the Merobase Search Engine API (*MeroAPI*). This API evolved from the Merobase component search engine [7] and provides advanced forms of queries over Web services and other forms of components. Advanced query types include interface-based searches which match services to queries based on their syntactic form, and also test-driven searches which match services based on their semantics (i.e. their behavior).

[1] A framework for document-based indexing and text-based searches.

In the following three sections we will at first explain how our UDDI architecture enhancements help to overcome the three weaknesses that we identified in section 1 as well as how these are realized in detail. This includes (a) automated crawling of the World Wide Web or a respective environment to add potential Web services to the service registry, (b) a higher degree of automated maintenance and continuous inspection of registered services and (c) enhanced service discovery capabilities with new forms of queries for higher-precision service retrieval.

3 Automated Registration of Web Services

As also shown in [3,8,9], today's internet offers many public Web services that may be reused in service-oriented application development. To allow users to find these publicly available services or services within the boundaries of an enterprise, the respective environment first has to be crawled in an automated way. Related details are described in another of our papers in [10]. Second, when services are found, they have to be registered with the UDDI server. In this automated discovery mode, the only available information about a service is contained in its WSDL description. Thus, a mapping to the UDDI data model is required for registration in a UDDI server environment. Since both are specifications within the basic Web services architecture [2] it would be natural to assume that they are compatible and that there is an intuitive and straight-forward way of transforming the service descriptions of WSDL into an equal description for UDDI. However, this is not the case. There are a number of issues that arise when correlating WSDL and UDDI that needed to be resolved by an OASIS technical note [11], which defines a consistent mapping between the data models of both specifications as shown in Figure 3.

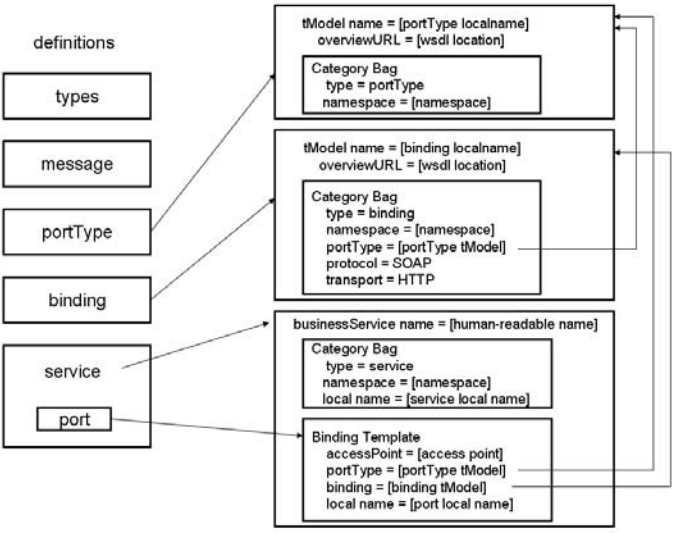


Fig. 3. WSDL to UDDI mapping [11]

OASIS defines some additional *tModels* to represent WSDL concepts in the UDDI structure including a *WSDL Entity Type tModel*, an *XML Namespace tModel*, a *WSDL portType Reference Model*, a *SOAP Protocol tModel* and finally a *HTTP Protocol tModel*. One of the main issues in this case is that to define new *tModels* in UDDI. The specification of an *overviewURL* within the data model is essential to point to the real "*interface*" – the WSDL description. However, if all WSDL elements are contained in the same description document and the WSDL schema does not allow ID attributes on WSDL elements, a mechanism which references single elements within the WSDL specification is needed. This issue has been resolved by OASIS with a pointer concept, called the *XPointer* syntax, which is described in [11]. Using this syntax it is possible to present each WSDL element by a *tModel* (as also shown in Figure 3) with the requirement that each *tModel* has the name of the corresponding local WSDL element and the namespace is saved as a so called *categoryBag* in the respective *tModel*. These *tModels* also contain the *overviewURL* which points to the WSDL element in the specification document. The *categoryBag* itself must contain a *keyedReference* with a *tModelKey* of the *WSDL Entity Type Category System* and a *keyValue* of "*portType*", "*binding*", "*service*", etc.. These *tModels* should contain other *keyedReferences* as well, for example to identify the corresponding *portType* within the *tModel* of the binding element, which can be viewed as a specification of the binding element. Each of these WSDL elements finally gets its own *tModel* within the UDDI data model using several specific parameters to cope with the individual requirements of the WSDL elements. For example, the binding element should contain another *keyedReference* to a *tModel* of the *UDDI Types Category* with the value of "*wsdlSpec*" for backward compatibility purposes, or there should exist references to the protocol and transport information e.g. of type "*soap:binding*" and "*http:binding*". The complete description of the mapping from the WSDL elements to the UDDI *tModels* is contained in [11].

4 Quality Assessment of Web Services

To solve the problem of data integrity and to support higher quality retrieval of Web services, our enhanced UDDI architecture introduces a mechanism to obtain availability information and statistics based on periodically repeated load tests of the registered Web services. These load tests are initiated by a timer service on a daily basis. To this end we use the leading Web service testing tool *soapUI* [12] that offers a framework for the testing of Web services.

The *soapUI* framework offers a variety of modes for functional and load testing of Web services. To achieve more realistic result values for performance evaluation, we have chosen load tests which are performed simulating multiple client requests in multiple threads which are executed concurrently. The simultaneous execution of only a couple of requests can change the response time of a Web service significantly. This approximates a real-world usage scenario where the number of concurrent requests can vary over time. Information that is provided by the executed load tests includes the response time, transactions per second and the number of errors that occurred. The result data that is collected from the load tests is stored in our *WSStats* database table over time. In the service discovery process, this information source is then used to

find the most suitable Web services based on the quality measures. The response time is viewed as an important benchmark. But since the network performance may vary depending on the location of the physical machine that executes the load tests, the determined values should be used only comparatively within the boundaries of the maintained service registry using a fixed testing environment.

Many approaches that consider the quality of Web services do not use information gained from testing to establish whether the service under consideration is active. In our approach, however, we use this information to designate the Web service as being "*active*" or not. This information is used to filter currently unavailable services from search results. Also, a potential service user probably won't be interested in the long term availability information if the service has not been running in the recent short term. Therefore, another mechanism in our approach additionally monitors if a service has failed to respond for a given period of time. If this is the case, the service is assigned the value "*quarantine*". There are two possible events that can change this state of a service. Either it can be accessed again with one of the load tests and is assigned the value "*available*" or the specified quarantine time period for the considered service passes by without a response and it is then assigned the value "*unavailable*". Similar to the quarantine time period, a time period for unavailable services (which needs to be much longer than the quarantine time period) is specified since over a certain threshold the probability that a service will become available again is very low and the service is finally removed from the registry.

Besides the already introduced filter mechanism, our approach introduces an "*availability*" classification schema for Web services that is determined from values stored in the database and values of the current load test. To determine the *availability* value of a certain Web service we compare the number of effective *running days* to the number of *totally monitored days*. In the same way we evaluate the *reliability* of a Web service by comparing the *number of response messages* to the *total number of request messages* that have been sent to the considered Web service.

The classification concept for Web services that we apply based on the evaluation of the availability value consists of the three levels which we characterize as "*excellent*", "*good*", and "*basic*". These are based on a scale that can be defined and used arbitrarily for value assignment. For example, services are classified as *excellent* if their availability value is higher than 95%, as *good* if the value is between 85% and 95% and as *basic* with an availability value below 85%. Using this classification, service requestors are offered clear decision criteria to choose appropriate services depending on their individual requirements. Other approaches that apply some kind of ranking value calculated as a compound function of many factors are often unclear to service requestors and users. These calculations also often contain attributes, e.g. interoperability characteristics, which are not crucial for service integration at the client's side.

All of the properties described in this section for characterizing the quality of a Web service which are collected in the process of quality assessment are stored in the *WSStats* database table and can be summarized as follows:

- the average, minimum and maximum response time (in ms),
- throughput (in tps),
- availability (in %),
- reliability (in %),

- active (available, quarantine, unavailable) and
- classification (excellent, good, basic).

5 Enhancing Search Capabilities

As mentioned in the previous section, the main advantage of our approach and its availability checking mechanism is that services that are currently not available are not considered as potential services in the process of service discovery. Using our extended UDDI architecture, the service requestor can choose between the common UDDI inquiry API and another interface that uses the Merobase Search Engine API (called Merobase API in the following) to search for appropriate services in sophisticated ways based on functional and non-functional properties.

We therefore have modified the implementation of the *jUDDI* inquiry API to access the quality information stored in the *WSStats* database tables for pre-filtering services that are currently not available or under quarantine. The quality information is also replicated in the Lucene index to support advanced queries that use the Merobase API to provide rapid, high-value results to a service requestor's search query. Moreover, the consistency of various data stores (UDDI database, quality database and the Lucene index) is maintained by updating the latter every time the former two are changed. Using common primary keys in each of the data stores, all the distributed information that is associated with one of the registered services can be returned in a query result or aggregated in a result set.

Using our Merobase API for enhanced Web service searches offers the service requestor multiple enhanced semantic query types in comparison to the simple UDDI key-word based search mechanisms. This is made possible by parsing the WSDL descriptions during the registration process and storing data in the Lucene index that is related to signatures of the Web service's operations. Furthermore we store other information in Lucene (e.g. service name, operation names, etc.) and additional information that can be gathered during the crawling process (e.g. provider host, last-updated, etc.). Service requestors can not only specify constraints on searches, e.g. "*excellent*" quality, they can also apply queries that specify the whole signature they expect related to their requirements. The following query is formulated in the Merobase Query Language (MQL) [7] to search for a calculator Web service with a certain interface (i.e. set of operations) and excellent quality:

```
Calculator (
  add(int, int):int;
  sub(int, int):int;
)
type:service
quality:excellent
```

In response to this query, all registered Web services that exactly conform to this specification are returned. In addition, an extension of our infrastructure as depicted in Figure 4 is able to offer test-driven searches where searches are driven by test cases that need to be fulfilled by semantically matching components. It therefore uses a user-specified test case to evaluate each service found through the regular

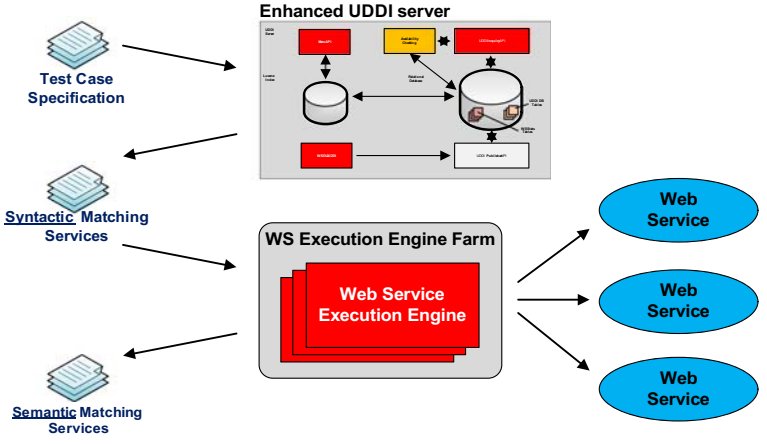


Fig. 4. Test-driven Web service searches

interface-based search mechanism by execution of the service and finally returns only those services that additionally have fulfilled the specified test case semantics.

6 Related Work

Besides the public UDDI Business Registry, several approaches have been applied in the last few years to broker services over the internet, based on Web service portals and search engines as also summarized in [3,8,9]. However, most of these approaches do not provide availability (and enhanced quality) information about the registered or offered Web services.

One of the more recent web sites that provides additional information gained from monitoring registered Web services is the service search engine seekda.com [13], introduced in the year 2008. With their approach, the internet is crawled for potential services to populate the index. In response to key-word based search queries it shows search results with the services' activity information and response time statistics. However, it offers a relatively weak search facility similar to the capabilities of the standard UDDI approach. Services which are currently not available or even no more available are not removed from the search space although they could have been inactive for months. Even though their downtime is shown in the provided statistics information, these services could appear on top of the list of search results.

In another work, Zhou et al. present an approach called UDDI eXtension (UX) that "facilitates requesters to discover services with good qualities" [14]. According to the concept of UX, the network is regarded as a unity of different domains (mainly organizations' domains – a university's domain, a company's domain, etc.), each of which holds a local UDDI registry that collects the local services of the domain. The UX-server serves as an intermediary between host, service requestor and the local registry. Furthermore it routes queries to the local registry to search for related results and starts a federated service search over other domain's UDDI registries if the num-

ber of result items is too low. As feedback, the UX-server receives QoS reports from service requestors that contain quality information like response time, reliability, costs, timestamp, and report number. This information is stored in a local database and is used as the basis for determining quality metrics for each service with an average or a weighted function.

Another approach for an extended UDDI architecture is presented in [16] with UDDIe, which supports three basic enhancements as follows. First, a "leasing" concept for services that allows those to be registered for a limited period of time to solve some of the data integrity problems; second additional attributes in the UDDI data model based on QoS information; and third, an extension of the find operation supporting "qualifier-based" search with numeric or logical UDDI queries. The basic UDDIe approach is used in the context of a grid computing framework where a QoS broker receives client requests enhanced with QoS attributes and contacts the UDDIe server. The obtained QoS data is used in a selection algorithm to return the best services based on a weighted average value. However, the QoS broker is not part of the UDDIe implementation.

Another approach that provides a conceptual model for quality factors and quality management factors within the Web services architecture is the Quality Model for Web services [17] introduced by OASIS. It defines two different kinds of quality factors divided into performance and stability factors. The former consist of response time and maximum throughput and the latter are represented by availability, successability and accessibility. This model also includes different actor roles – consumer, developer, provider, QoS broker, and assurer. The focus of the model is the QoS broker, whose role is to provide the most objective measurement and criteria for service qualities. The QoS broker is often viewed as a fourth actor in the middle of the well-known SOA-triangle presented in Figure 1.

Another piece of work that addresses the problem of Web service quality is the Web Service Relevancy Function (WsRF) introduced in [18] by Al-Masri and Mahmoud and applied in their Web Service Repository Builder architecture introduced in [19]. Their architecture introduces a Web Service QoS Manager that mediates between the UDDI architecture and clients. It retrieves and returns Web service information, and verifies quality in the sense of a QoS broker. To this end, they propose to use a link from UDDI information stored in a *tModel* that points to an external resource that stores quality metrics.

As illustrated in Figure 5 which shows the conceptual architecture from a 2002 W3C specification [2], QoS is regarded as an additional layer parallel to the main technology stack of Web services. We believe that the essential part of useful QoS functionality can be maintained in the discovery layer within UDDI as indicated in Figure 5. The extended UDDI architecture that we have presented in this paper can be enhanced along the lines of the Quality Model for Web services to act as a QoS broker. This provides two benefits: the first enables UDDI to accomplish its task and deliver the right services to requestors; the second shields the SOA architecture from further complexity associated with additional actors.

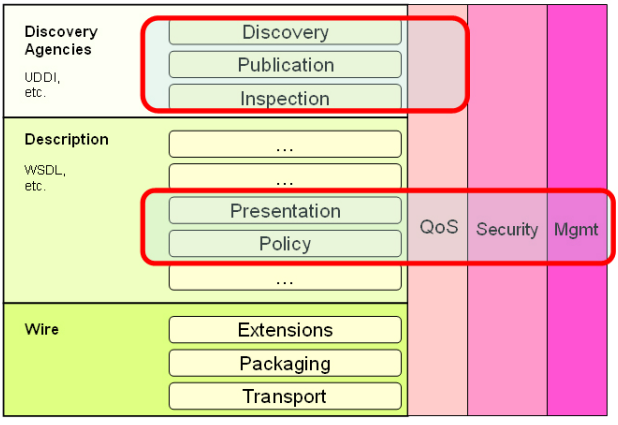


Fig. 5. The revised Web services stack

7 Conclusion and Outlook on Future Work

In this paper we have introduced an extended UDDI server architecture that supports automated Web service registration, availability and quality assessment. Our approach furthermore supports enhanced search capabilities for better Web service brokerage solutions and our mechanism for automated registration of Web services reduces the effort that service providers have to perform in registering their services. The main drawback of this approach is that information about service providers that can be manually added using the UDDI publisher API or any other approach taken by web-based portals may be missing. To overcome this problem we propose to add standardized annotations to WSDL descriptions that contain the information that normally is entered by hand. The use of a parser in the automated registration approach would overcome this drawback to deliver full information about services and providers conforming to UDDI and our approach.

As an extension of our availability checking and quality evaluation mechanisms, we are planning to extend our approach with automated periodic tests based on test case input that has been delivered by test-driven searches using our Web service execution engine which is part of the Merobase search engine infrastructure. We are able to collect and reuse any set of service parameters that has been used for testing a service with the Web Service execution engine. Furthermore, we are planning to implement a user feedback service for confirmation and assessment of successful Web service tests for this purpose. Storing service parameter data for successful test cases related to a certain service then allows our availability checking mechanism to randomly choose test data sets for executing functional as well as load tests.

Although the architecture presented in this paper is relatively simple, we believe it can have a major impact on the usability of service brokerage solutions in general and UDDI in particular. We are currently in the process of consolidating and testing our prototype implementation and plan to make the technology available in the near future in association with the Merobase search engine.

References

1. Hummel, O., Atkinson, C.: Using the Web as a Reuse Repository. In: Morisio, M. (ed.) ICSR 2006. LNCS, vol. 4039, pp. 298–311. Springer, Heidelberg (2006)
2. Web Services Architecture (W3C), http://www.w3.org/TR/2002/WD-ws-arch-20021114/
3. Al-Masri, E., Mahmoud, Q.H.: Investigating Web Services on the World Wide Web. In: 17th International Conference on World Wide Web, pp. 795–804. ACM, New York (2008)
4. Poulin, J.: Populating Software Repositories: Incentives and Domain-Specific Software. Journal of Systems and Software 30(3), 187–199 (1995)
5. UDDI Version 3.0.2 (OASIS), http://www.oasis-open.org/committees/uddi-spec/doc/spec//v3/uddi-v3.0.2-0041019.htm
6. Apache jUDDI, http://ws.apache.org/juddi
7. Merobase Component Search Engine, http://www.merobase.com
8. Hagemann, S., Letz, C., Vossen, G.: Web Service Discovery - Reality Check 2.0. In: 3rd International Conference on Next Generation Web Services Practices, pp. 113–118. IEEE Computer Society, Washington (2007)
9. Fan, J., Kambhampati, S.: A Snapshot of Public Web Services. SIGMOD Rec. 34(1), 24–32 (2005)
10. Atkinson, C., Bostan, P., Hummel, O., Stoll, D.: A Practical Approach to Web Service Discovery and Retrieval. In: 7th IEEE International Conference on Web Services, pp. 241–248. IEEE Computer Society, Washington (2007)
11. Using WSDL in a UDDI Registry (OASIS), http://www.oasis-open.org/committees/uddi-spec/doc/tn/uddi-spec-tc-tn-wsdl-v2.htm
12. soapUI – The Web Service Testing Tool, http://www.soapui.org
13. seekda! – Web Services Search Engine, http://www.seekda.com
14. Zhou, C., Chia, L.T., Silverajan, B., Lee, B.S.: UX- An Architecture Providing QoS-Aware and Federated Support for UDDI. In: 3rd IEEE International Conference on Web Services, pp. 171–176. CSREA Press (2003)
15. Lucene, http://lucene.apache.org
16. Universal Description, Discovery and Integration – Extension (UDDIe), http://www.wesc.ac.uk/projects/uddie/uddie/index.htm
17. Quality Model for Web Services v2.0 (OASIS), http://www.oasis-open.org/committees/download.php/15910/WSQM-ver-2.0.doc
18. Al-Masri, E., Mahmoud, Q.H.: Discovering the Best Web Service. In: 16th International Conference on World Wide Web, pp. 1257–1258. ACM, New York (2007)
19. Al-Masri, E., Mahmoud, Q.H.: A Framework for Efficient Discovery of Web Services across Heterogeneous Registries. In: 4th IEEE Consumer Communication and Networking Conference, pp. 415–419. IEEE Computer Society, Washington (2007)

QoS-Aware Peer Services Selection Using Ant Colony Optimisation

Jun Shen and Shuai Yuan

School of Information Systems and Technology
University of Wollongong, Australia
Wollongong 2522, Australia
jshen@uow.edu.au, shuai.yuan@hotmail.com

Abstract. Web services coordinated by computational peers can be aggregated to create composite workflows that provide streamlined functionality for human users or other systems. One of the most critical challenges introduced by Peer-to-Peer (P2P) based Web services is represented by Quality of Service (QoS)-driven services composition. Since many available Peers provide overlapping or identical functionalities, though with different QoS, selections need to be quickly made to determine which peers are suitable to participate in an expected composite service. The main contribution of this paper is a heuristic approach which effectively and adaptively finds appropriate service peers for a service workflow composition, and also some uncertainties in the real ad-hoc scenarios are considered by a proper re-planning scheme. We propose to adopt Ant Colony Optimisation (ACO) to tackle the QoS-aware Peers' composition problem in both static and dynamic situations, as ACO represents a more scalable choice, and is suitable to handle and balance generic QoS attributes by pheromones. The proposed approach is able to improve the selection performances in various service composition structures, and also can adaptively handle unexpected events. We present experimental results to illustrate the efficiency and feasibility of the proposed method.

Keywords: P2P, QoS, ACO, service selection, composition.

1 Introduction

Web services are autonomous software systems identified by URIs which can be advertised, located, and accessed through messages encoded according to XML-based standards (e.g., SOAP, WSDL, and UDDI [6]) and transmitted using Internet protocols [18]. In decentralised network, Peer-to-Peer [19] (P2P) based service computing inherits the foundational service-oriented features (e.g. service protocols, service discovery mechanism and QoS awareness, etc.), and therefore becomes more flexible and scalable, due to the capability of fully-distributed computing. Widely accepted and standardised Service-Oriented Architecture (SOA) makes it possible to realise larger scale computing paradigm like SaaS, so that P2P-based service systems, which depend upon peers' cooperation and share on services resources, can bring profound benefits and profits for decentralised service network (e.g. mobile commerce application and Location Based Services [14]). In fully-distributed P2P based service

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 362–374, 2009.

systems, service composition is very inherent to occur as service peers are expected to work together, normally represented by a business process or scientific or engineering workflow, through cooperation/coordination to achieve desirable global goals. JXTA and BPEL [19] also allow peers to cooperate and automate business processes and reengineer their workflow structure, so as to rationally compose and make use of all resources in decentralised environment. In addition, P2P-based composite application is able to increase efficiency and reduce costs, as they are highly reusable and easily restructurable. The fast composition is essentially required to replan a service composition during the execution, sometimes because the actual QoS deviates from the estimated schedule and this may cause constraint violation, or sometimes even simply because some services might not be available on the fly. In this case, the re-composition time will influence the overall service response time, thus it should be kept as minimum as possible.

The scalability of decentralised service composition is a research challenge because the evaluation of all possible service peers can lead to a combinatorial explosion of solutions. In compositions involving large group of peers, an exhaustive search could be very time consuming. Moreover, it is particularly important for the QoS aware composition process itself to be fast. Especially for the synchronised or interactive systems, long delays may be extremely unacceptable. For example, the user of a ticket booking system might not want to wait for a long time while the system searches for candidate services offering flight tickets with the lowest booking fare. While peer composition is viewed as the ability to combine existing peers together in order to generate new functionalities or services, it usually becomes extremely hard to conduct when handling a large number of services. This is because that, to find the solution of such a QoS–aware composition problem is NP-hard. Although previous research efforts have looked at many pragmatic ways to effectively deal with the composition of Web services into executable workflows (e.g. [17, 20]), a remaining concern is how to optimise the service composition in terms of making system more effective. Thus, it is necessary to have an effective approach to handle the composition issue quickly, and also, adaptively.

In this paper we propose to adopt ACO (Ant Colony Optimisation) to enhance performance of service composition for P2P workflow and prove it is suitable to be used in QoS-aware service composition. The remainder of this paper is organised as follows. After related work is presented in Section 2, Section 3 details the proposed approach, i.e. the ACO based method for QoS-aware composition in P2P workflow. Section 4 reports and discusses the experiment results obtained in the simulations, and finally, Section 5 concludes the paper.

2 Related Work

QoS–aware service selection for composition attracts many interesting applications and trials of various methods and search strategies, mostly based on operation research (OR) or artificial intelligence (AI). In general, most of the accepted approaches for effective service composition are focusing on how to find the optimal/close-optimal combinations with the minimum cost (e.g. time, computation resource constraints), and they are usually dependent upon Integer Programming or

Genetic Algorithms, or mixed up with some other optimisation strategies. Recently ACO has been widely used for the problems such as allocating jobs in robot networks and searching the shortest path, but has rarely been considered for service selection issue. Here we would firstly discuss about those typical approaches used for service composition, and briefly compare with our proposed ACO solution.

The Integer programming (IP) solutions with regard to dynamically finding the best service combination have been proposed in some recent papers ([1], [20]). These works consider linearity of the constraints and the objective functions, and find the best combination of the concrete services. From the QoS point of view, they're conducted at run time. Zeng et al. [20] essentially focus on the cost, response time, availability and reliability attributes, where logarithmic reductions are used for the multiplicative aggregation functions, and the model is claimed to be extensible with respect to other similarly behaving attributes. The need to deal with more general constraint criteria, such as service dependencies or user preferences, is strengthened in the work presented by Aggarwal et al. [1]. However, they do not explain how the optimisation problem could be solved. In this paper, we adopt the ACO instead, as any kind of constraint could be handled herein, and we also provide some empirical data to assess the performance of our method versus the (linear) integer programming solution in the service composition setting, especially in terms of comparison of the computation time with the same number of tasks. An alternative to find a close-optimal solution could be to use nonlinear integer programming techniques, but the maturity of the available tools is questionable, and the application of these methods in our setting requires further investigation. A survey of some nonlinear techniques is contained in a paper by Grossmann [9].

Genetic Algorithms (GA) is viewed as a distinguishing heuristic method for seeking a close-optimal combination. With regard to service composition, Canfora et al. [2] utilised GA and focused on the situation where there may be more than one service candidate that provides identical functionality but has different QoS, to determine which service candidates should participate in a required composite service. When compared to widely-used linear integer programming methods, their genetic algorithm deals with non-linear functions while it scales well with an increase in the number of tasks. Technically, ACO is also qualified to deal with that situation adaptively, as it uses updated information from time to time to approach to the optimal result. A randomised heuristic method based on the general principles of genetic search strategy was described by Chockalingam et al. [4] to solve the mapping problem, in which parallel tasks are assigned to a multiprocessor to minimise the execution time. Similarly to a P2P-based service system, all peers are required to cooperate and work together to accomplish a designed composite task, but we would consider using ACO to save more time. The work conducted by Cao et al. [3] is a recent development on the application of GA to the service selection problem in the context of Web service composition. A service selection model defines a business process as a composition of many service agents. Each service agent corresponds to a set of multiple Web services, which are provided by different service providers to perform a specific task. Nevertheless, taking into account the uncertainty and self-adaptive capability, GA may be not as good as ACO to deal with dynamic situations or unexpected events (e.g. when agent breaks down or its quality is changed during iterations). Instead, ACO can take more advantage of heuristic information and adapt itself

to any changing circumstance quickly and easily, and also can facilitate to formulate the selection problem based on the nature of a service application effectively.

A lot of researches regarding the QoS in Web services focus on the development of QoS ontology languages and vocabularies, as well as the identification of various QoS metrics [11] and their measurements with respect to semantic Web services. For example, papers [10] and [13] emphasised a definition of QoS aspects and metrics. In [10], all of the possible quality requirements were introduced and divided into several categories, including runtime related, transaction support related, configuration management and cost related, and security related QoS. Both of the papers shortly present their definitions and possible determinants. In our previous work [15, 16, 19], we focused on P2P-based service selection and the extensible description of non-functional properties via OWL-S and WSMO. In [19], we presented a first sketch of QoS-aware service selection, however, with special attention to the extraction of the ontological description of services and the design of the selection process with OWL-S. With regard to the selection process, the prototype presented in that paper has the limitation in terms of dealing with multi-specifications, because it only considers "ResponseTime" as the selection criteria, by which the selection is not quite realistic for effective services composition. Therefore in [15, 16], we extended the description of non-functional properties via modelling-driven WSMO specification, and presented an algorithm for Peer coordinator to automatically identify the best peers through unifying qualities and properties. Nevertheless, based upon our earlier works, the main aim of this paper is to propose a quick and effective ACO-based approach for P2P-based QoS-aware service composition.

3 The Methodology

Ant Colony Optimisation is a well-established optimisation technique based on the principle that real ants are able to find the shortest paths between their nest and a food source [8]. This mechanism works on the basis of pheromones, some kind of biochemical scent, which is left behind by the ants. Other ants are attracted by these pheromones and always walk in the direction with the highest pheromone concentration. This natural behaviour was first adopted as an optimisation technique by Dorigo, Colorni, and Maniezzo [5, 7]. Their Ant System provides the foundation of the optimisation method presented here. This behaviour is the basis for a cooperative interaction, and this manner can be applied not only to solve discrete optimisation problems but also to solve both static and dynamic combinational optimisation problems [12].

In this work an artificial ant is an agent which moves from node to node on a composition graph. It chooses the nodes to move by using a probabilistic function of both trail accumulated on links, and a heuristic value, which was chosen here to be a function of the number of services within a workflow. By a probability, artificial ants prefer nodes which are connected by links with relatively more pheromone trail. The following are the three heuristic hints from natural ant behaviours that we have translated to our artificial ant colony: 1. the preference for paths with a high pheromone level; 2. the higher rate of growth of the amount of pheromone on ideal paths; 3. the path-mediated communication among ants.

For the purpose of selection processes, ResponseTime, Cost, Availability and Reputation are utilised often as QoS factors, and they can be described as a peer's non-functional properties with WSMO. Assume a composite service "*l*", we define that *RT(l)*, *C(l)*, *A(l)* and *R(l)* are the respective normalised QoS factors (ResponseTime, Cost, Availability and Reputation) in the interval [0,1]. These normalised QoS attributions can be computed by applying the rules described in Table 1, which is based on a simple way of normalisation. The composite service "*l*" contains "*n*" atomic Web services, and we assume that the number of artificial ants is "*m*" at each iteration (after an iteration, ideally every ant can find a combination of possible candidate peers for composite service "*l*").

Table 1. Normalising QoS Attributions

Response Time *(RT)*	$RT(l) = \frac{ReponseTime(l)}{\sum_{i=1}^{m} ReponseTime(l)}; \quad ReponseTime(l) = \sum_{i=1}^{n} ReponseTime(ws_i)$
Cost *(C)*	$C(l) = \frac{Cost(l)}{\sum_{i=1}^{m} Cost(l)}; \quad Cost(l) = \sum_{i=1}^{n} Cost(ws_i)$
Availability *(A)*	$A(l) = \frac{Availability(l)}{\sum_{i=1}^{m} Availability(l)}; \quad Availability(l) = \prod_{i=1}^{n} Availability(ws_i)$
Reputation *(R)*	$R(l) = \frac{Reputation(l)}{\sum_{i=1}^{m} Reputation(l)}; \quad Reputation(l) = \prod_{i=1}^{n} Reputation(ws_i)$

We define a cost function for evaluating QoS of a service composition *l* as follows:

$$Q(l) = \frac{w_1 \cdot RT(l) + w_2 \cdot C(l)}{w_3 \cdot A(l) + w_4 \cdot R(l)} \tag{1}$$

Where: w_1, w_2, w_3 and w_4 are weights which indicate the importance of the QoS factors for service integrator (or user). Q(*l*) represents cost function of composite service *l*, and the objective is to minimise the cost function value of composite service "l". The reason why we define the function in this way is that we need to differentiate the preferences of QoS properties, such as ResponseTime (is preferred as low as possible) and Availability (is preferred as high as possible).

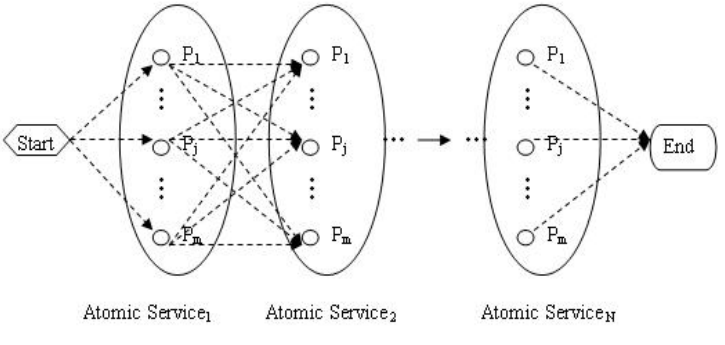


Fig. 1. Composition graph for services workflow

In Figure 1, we may assume that there are 'm' candidate peers $\{P_1, .., P_m\}$ and 'N' atomic services {Atomic Service$_1$, …, Atomic Service$_N$} in a service composition process. That is to say, there are 'N' groups of peers, and each group is $\{P_1, .., P_m\}$. Clearly, the goal is to find right peers from each group for each atomic service. Hence, by applying the ACO based approach, *m* artificial ants are initially placed on Atomic Service$_1$'s nodes (from P_1 to P_m). Ideally, the more ants, the better, but that would consume much longer time for iteration. Hence, in a balanced way, we normally set the number of ants is the number of peers. For any step *i*, ants move to newer possible nodes (the peer group for Atomic Service$_i$) until it reaches the end node (the peer group for Atomic Service$_N$). When all the ants have completed a path, the ant that made the path with the lowest cost function value *Q* would modify the links belonging to its path by adding an amount of pheromone trail. In each iteration, each ant generates a feasible composite service by choosing the nodes according to a probabilistic state transition rule.

For the selection of a node, ant uses heuristic factor as well as the pheromone factor. The heuristic factor denoted by η is a heuristic desirability about an ant moving from a current node to another, and the pheromone factor denoted by τ is an indication of how many ants have visited the link. The probability of selecting a next node (from Atomic Service$_i$ to Atomic Service$_j$) is given by:

$$P_{ij}^{k} = \begin{cases} \dfrac{\tau_{ij}^{\alpha} \cdot \eta_{ij}^{\beta}}{\sum_{h \in H} \tau_{ih}^{\alpha} \cdot \eta_{ih}^{\beta}} & if \ \rightarrow \ i \in H \\ 0 \rightarrow Otherwise \end{cases} \tag{2}$$

where P^k_{ij} is the probability with which ant *k* chooses a next node to go. *H* is the set of nodes that have not been passed by any ant for Atomic Service$_j$, and *h* is a node (i.e. a peer in the group for Atomic Service$_j$) which has not been visited yet by ant *k*. The parameters α and β control the relative importance of the pheromone versus the heuristic information η_{ij}, which is the heuristic desirability of the ant on node *i* moves to node *j*, is given by:

$$\eta_{ij} = \frac{1}{Nj} \tag{3}$$

where N_j is the number of Web services between Atomic Service$_j$ and Atomic Service$_n$. Actually, for service peer composition, η can be viewed as a heuristic stimulus which can keep ants moving from source towards destination covering all atomic services.

The whole path's pheromone update is applied at the end of each iteration by only one ant, which can be the iteration-best or the best-so-far, and the update formula is defined as follows:

$$\tau_{ij} = \begin{cases} (1-\rho) \cdot \tau_{ij} + \rho \cdot \Delta\tau, if \ (i, j) \in L \\ \tau_{ij} \rightarrow Otherwise \end{cases} \tag{4}$$

where $\rho \in (0, 1]$ is a parameter that controls the speed of evaporation of pheromone, and $\Delta\tau = 1/Q(L)$, here L is the current best path with the lowest Q value.

To facilitate the peers' service composition and deal with dynamical situations, the proposed ACO algorithm for service composition is presented as follows. Algorithm 1 is designed to quickly find the close-optimal combination for a composite service, and Algorithm 2 is responsible for monitoring uncertainties and then generating replacement solution for the remaining part of composition in case one or more peer(s) becomes unavailable. In Algorithm 1, each atomic service on a peer is regarded as a virtual node of a potential path that might be found by ants. After each iteration, a possible path would be found by ants, and then the combination QoS is calculated and recorded, afterwards the visited nodes' pheromones would be updated. Finally, the nodes with more pheromones would be more possibly consisted of an eventual path, i.e. a close-optimal combination would be generated. For dynamical changes and uncertainties during execution, the Algorithm 2 is set as a supplement to reasonably cope with peer's breakdown or if there are any other needs.

Algorithm 1. ACO-based Approach for Service Peer Selection

Step 1. Initialisation:
An initial population of ant colony individuals Ant_k, k=1, 2,…, m (m can be the number of candidate nodes), is initialised in this step. N is the number of atomic Web services. Format the composite nodes list of each ant to provide the following steps to record the ant composition history. Set a small amount of pheromone on each node as initial condition for each link. Set iteration counter I=1, and initialise the maximum iteration number $I_{max.}$ (When I reaches I_{max}, Stop.)
Step 2. Starting travel:
for ant k=1 to m **do**
  Place Ant_k on candidate node of the first atomic service.
end for
Step 3. Searching for next node:
Repeat until the composition list is full
for k=1 to m **do**
  if ant Ant_k is already reached the final node of composition
    Then Break and **do** Ant_{k+1}.
  else
    Choose the candidate nodes and calculate the probability P^k given in formula (2).
  end if
  if the ant Ant_k is stuck at a dead node or reach the maximum transport action number N.
    Then move the ant back to the start node, and clear the current composition list of Ant_k.
  end if
end for
Step 4. Calculate the composite service's QoS value for each ant's journey:
for k=1 to m **do**
  Compute the Q value according to Q(l) in formula (1) based on the visited nodes
end for
Step 5. Update the pheromone factor:
For each path's nodes update the pheromone value according to formula (4)
Find the best composition path and update the best composition history.
Step 6. Check to stop criterion:
if ($I<I_{max}$)
  Then record the best composition list, and **Goto** Step 2.
else
  Return the best composite service and stop.
end if

Algorithm 2. Monitoring and Replanning

```
While (1)
    if any breakdown/events happen
        Put all incomplete atomic Web services into an arraylist L;
        Delete the dead peers and refresh the peer group;
        Do Algorithm 1 for replanning the left composition;
    end if
end while
```

4 Experimental Results and Evaluation

In this section, we evaluate performance and feasibility of the proposed ACO method. To verify the proposed approach, experiments are conducted based on a few different scenarios. Essentially, we conduct a few experiments to prove the proposed method: 1. apply ACO in a general composite workflow, and compare ACO with optimisation method in static environments, in terms of computation time; 2. test the re-planning performance by randomly setting a few peers dead in dynamical environments.

In simulation, we established the environment with Matlab, and all experiments were conducted on a PC with Intel Pentium4 3.0GHz, 2MB; 800 MHz FSB; DDR2 1GB @ 667MHz; and MATLAB 2008a. All QoS data (i.e. ResponseTime, Cost, Availability and Reputation) of service peers are randomly generated in Matlab.

4.1 Quick Composition and Comparison

As shown in Figure 2, a general workflow, which contains structures of split and flow for service composition, are depicted, with 7 atomic Web services (AWSs) involved. As a matter of fact, for peer selection process, now we do not consider about differences between simple sequential composition pattern and other mixed patterns, as our emphasis is on selecting an appropriate peer for each atomic service amongst service peers. Therefore, in this work, the selection method we proposed can be extensible for service selection in all sorts of composition patterns, even though different composition patterns may impact the calculation of cost functions.

In this experiment, we assume that there are 10 service candidate peers, and utilise the proposed method to select appropriate service peers which are able to conduct those 7 atomic Web services and provide better overall service quality. The major goal of this method is to identify better peers with pheromones and probabilities for atomic services. In the selection process, there is little difference between a purely

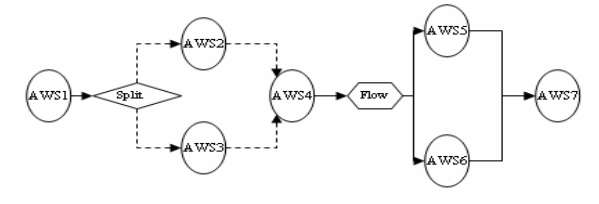


Fig. 2. A general composition workflow

sequential workflow and the mixed one, while using ACO method to search for proper candidate peers. Ants will exploit their own heuristics at join or split nodes.

Figure 3 shows the nature of ACO regarding best composition value and average value at each iteration for this experiment. These composition values are based on calculated Q values. At every iteration, the change of best value indicates that ants found a better path of composition than the previous iteration. As shown in left part of Figure 3, the best composition values were getting better and better from the beginning to the 9^{th} iteration, and afterwards there was no changes happened since ants could not found better solution. In the right part of Figure 3, the situation of average composition values changed obviously during all the 20 iterations. Here, the average value is the mean value of all possible compositions found by ants at each iteration, and the changes of average value are caused by the probabilities of ants' choices and the amount of pheromones. However, the trend of the best composition value is led well by heuristic information, as shown in the left part of Figure 3.

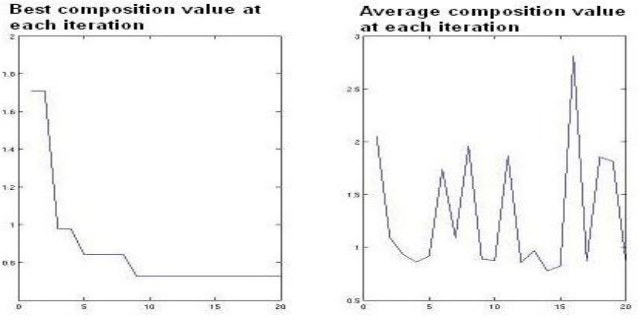


Fig. 3. The Q values in ACO iterations

Table 2. Calculated Q Values of Peers for Atomic Services without ACO Approach

	AWS 1	AWS 2	AWS 3	AWS 4	AWS 5	AWS 6	AWS 7
Peer 1	0.873	0.398	0.780	**0.008**	0.455	0.950	**0.144**
Peer 2	**0.142**	0.413	0.554	Inf	0.869	0.630	0.394
Peer 3	0.682	0.963	Inf	0.677	0.219	0.378	0.259
Peer 4	Inf	**0.075**	0.751	0.217	0.466	0.471	0.338
Peer 5	0.781	0.839	0.532	0.054	0.387	0.887	0.853
Peer 6	0.946	0.248	**0.009**	0.029	Inf	**0.130**	Inf
Peer 7	0.212	0.729	0.683	0.236	0.437	0.442	0.145
Peer 8	Inf	0.695	0.040	0.292	0.372	0.658	0.798
Peer 9	0.990	0.229	0.153	0.229	0.449	Inf	0.162
Peer 10	0.458	0.333	0.830	Inf	**0.193**	0.419	0.465

Table 3. Selection Results with Q Values by ACO Method

AWS1	AWS2	AWS3	AWS4	AWS5	AWS6	AWS7
Peer 2 (0.142)	Peer 4 (0.075)	Peer 6 (0.009)	Peer1 (0.008)	Peer 10 (0.193)	Peer 6 (0.130)	Peer 1 (0.144)

Table 2 shows each service peer's Q value that is calculated based on formulas in Table 1, without ACO approach, and each data represents the relationship (Q value) between a candidate peer and an atomic service rather than a composition service. The infinite (Inf) value of a peer means that the service peer can not fulfil the corresponding atomic service, and Q values of selected peers are calculated based on a set of random data which are representing relevant peers' non-functional properties. Table 3 then lists the results from using ACO method. From the two tables, we can see that the combination results by ACO method are exactly the same selections with the best Q value for atomic services in Table 2. It suggests that ACO based approach can also generate the results without completely knowing Q values of each peer. Actually, exhaustively calculating all peers' Q values in this way (e.g. Table 2) is a simple optimisation method to select service peers for service composition, as it can easily find the best peers combination after knowing each peer's Q values for atomic services. Hence, the major cost of computation by the optimisation method is the calculation on all possible combinations, and when the number of atomic services and peers increases, the time cost would greatly increase as well. Given the computation cost, ACO based approach does not need to calculate all service candidates' Q values beforehand, since the calculation is only needed to conduct when a service peer is chosen with a certain probability. In other words, ACO approach can save much more computation time than the optimisation method, particularly for a large number of possible combinations.

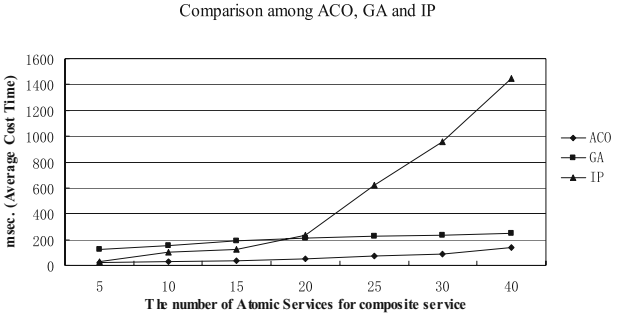


Fig. 4. Comparison of ACO, GA and IP

As shown in Figure 4, ACO is the least time consumer through all the experiments. When the number of concrete services is small, say, less than 15, IP outperforms GA. For about 20 atomic services the performances of ACO and IP tend to be the same. Then with more atomic services, while the ACO and GA are able to keep their computation time performance almost constant, this is not the case for IP based optimisation method, where we see an exponential growth due to the corresponding increment of the number of variables needed to represent the problem. In this experiment, we assumed a set of workflows consisting of 5 to 40 atomic services, and the same numbers of service peers are involved in the selection for each atomic service. The experimental settings of GA were based on [2], where GA is with crossover probability of 0.7, a mutation probability of 0.01 and the number of generation is 100. The parameters of the ACO in the simulations were set as: α, β are 1 and 5, respectively, ρ (speed of evaporation) is 0.1, and 20 iterations (as few changes occur after 20 runs in the test) for each trial.

4.2 Effective Re-planning

In reality, composite services are often conducted in scenarios that are changing dynamically and unexpectedly, due to real environments' uncertainties. In this case, peers need to take actions quickly for the re-planning and/or re-adjustment if any events occur. For example, some of the selected service peers for a composite service might crash down during execution, and then the expected workflow would be stuck forever until a replacement was given. Hence, it is necessary for the ad-hoc execution of a whole workflow to be affiliated with a replanning scheme, i.e., the ability to adaptively find a replacement in order to resume the remaining incomplete tasks. Figure 5 depicts 6 scenarios: two with "5 atomic Web services and 20 peers", two with "10 atomic Web services and 30 peers" and two with "15 atomic Web services and 40 peers". During the composite service execution, we assume some service peers to be dead unexpectedly, e.g., 1, 2 and 4 service peers can be suddenly out of order in each scenario. In the simulation, consumed time for replanning vary from case to case, the maximum replanning time is 45 (msec.), and the minimum is 7 (msec). "F" means that there is no available peer(s) for the required atomic services. The amount of consumed replanning time is actually dependent on the dead peer's virtual location in workflow process where a number of atomic services have not completed when the event occurred.

Based on all experimental results, we conclude that the ACO-based approach leads to effective and adaptive QoS-aware composition with less computation time and reasonable quality of solution.

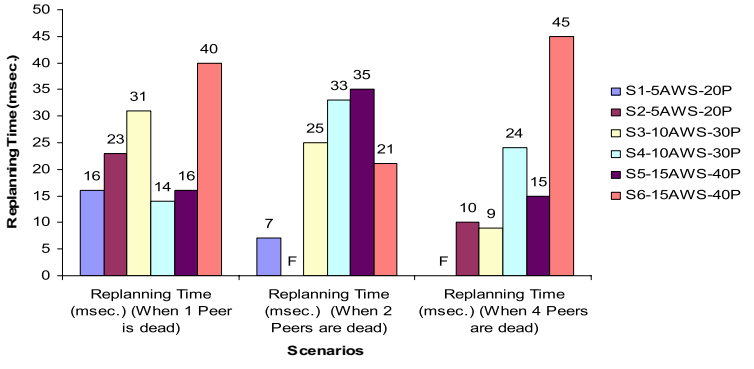


Fig. 5. Replanning Performance of ACO in Dynamical Environments

5 Conclusion and Future Work

In this paper, the ACO-based approach is proposed to tackle the QoS-aware peers' composition problem, in terms of determining a set of peers to be bound to atomic services. Those services are contained in an orchestration, and candidate peers are heuristically selected based on QoS aspects for the orchestration by the means of ACO. With our experimental results, ACO allows a P2P based e-service system to handle QoS attributes with non–linear aggregation functions adaptively and

efficiently, and it is a balanced practice to apply ACO to quickly finding the close-optimal solution. The main contribution of this paper is a heuristic approach to effectively determine and select appropriate service peers for composition, and also the uncertainties in the real service execution scenarios are considered with a proper re-planning scheme.

In the future work, we would investigate some probabilistic models, e.g. Partially Observable Markov Decision Process (POMDP), to cope with more complicated uncertainties within actual service compositions and executions, and also we would implement an essential P2P based service prototype and testify the approach to see how sound it can suit the real applications.

References

[1] Aggarwal, R., Verma, K., Miller, J., Milnor, W.: Constraint driven Web service composition in METEOR-S. In: Proceedings of the 2004 IEEE International Conference on Services Computing, pp. 23–30. IEEE Computer Society, Los Alamitos (2004)

[2] Canfora, G., Penta, M.D., Esposito, R., Villani, M.L.: An approach for QoS-aware service composition based on genetic algorithms. In: Proceedings of the 2005 conference on Genetic and evolutionary computation, New York, USA, pp. 1069–1075 (2005)

[3] Cao, L., Li, M., Cao, J.: Using genetic algorithm to implement cost-driven Web service selection. Multiagent and Grid Systems 3(1), 9–17 (2007)

[4] Chockalingam, T., Arunkumar, S.: Genetic algorithm based heuristics for the mapping problem. Computers and Operations Research 22(1), 55–64 (1995)

[5] Colorni, A., Dorigo, M., Maniezzo, V.: Distributed Optimisation by Ant Colonies. In: Proceedings of the European Conference on Artificial Life, Paris, France, pp. 134–142. Elsevier Publishing, Amsterdam (1991)

[6] Curbera, F., et al.: Unraveling the Web Services: An Introduction to SOAP, WSDL, and UDDI. IEEE Internet Computing 6(2), 86–93 (2002)

[7] Dorigo, M., Maniezzo, V., Colorni, A.: The Ant System: Optimization by a Colony of Cooperating Agents. IEEE Transactions on Systems, Man and Cybernetics - Part B 26(1), 1–13 (1996)

[8] Goss, S., Aron, S., Deneubourg, J.-L., Pasteels, J.M.: Self-Organized Shortcuts in the Argentine Ant. Naturwissenschaften 76(12), 579–581 (1989)

[9] Grossmann, I.: Review of nonlinear mixed-integer and disjunctive programming techniques. Optimization and Engineering 3(3), 227–252 (2002)

[10] Lee, K., Jeon, J., Lee, W., Jeong, S., Park, S.: QoS for Web services: Requirements and Possible Approaches. W3C Working Group Note 25 (2003), http://www.w3c.or.kr/kr-office/TR/2003/ws-qos/

[11] Liu, Y.T., Ngu, A.H.H., Zeng, L.Z.: QoS computation and policing in dynamic Web service selection. In: Proceedings of International Conference on World Wide Web, pp. 165–176. IEEE CS Press, New York (2004)

[12] Lorpunmanee, S., Sap, M.N., Abdullah, A.H., Chompoo-inwai, C.: An Ant Colony Optimization for Dynamic Job Scheduling in Grid Environment. International Journal of Computer and Information Science and Engineering 1(4), 207–214 (2007)

[13] Ran, S.: A model for Web services Discovery with QoS. ACM SIGecom Exchanges 4(1), 1–10

[14] Shen, J., Krishna, A., Yuan, S., Cai, K., Qin, Y.M.: A Pragmatic GIS-Oriented Ontology for Location Based Services. In: The 19th Australian Software Engineering Conference (ASWEC 2008), Perth, Australia, pp. 562–569. IEEE Computer Society Press, Los Alamitos (2008)

[15] Shen, J., Yuan, S.: Adaptive E-Service Selection in P2P-based Workflow with Multiple Property Specifications. In: Ting, I., Wu, H. (eds.) Book Web Mining Applications in E-commerce & E-services, pp. 153–168. Springer, Berlin (2009)

[16] Shen, J., Yuan, S.: Modelling Quality and Spatial Characteristics for Autonomous e-Service Peers. In: The 20th International Conference on Advanced Information Systems Engineering (CAiSE 2008), Forum, Montpellier, France, June 2008, vol. 344, pp. 49–52. CEUR-WS (2008) ISSN: 1613-0073

[17] Vanrompay, Y., Rigole, P., Berbers, Y.: Genetic algorithm-based optimization of service composition and deployment. In: Proceedings of the 3rd international workshop on Services integration in pervasive environments, pp. 13–17 (2008)

[18] Web Services Architecture Requirements Working Group (2004), http://www.w3.org/TR/wsa-reqs

[19] Yuan, S., Shen, J.: Mining E-Services in P2P-based Workflow Enactments. special issue Web Mining Applications in E-commerce and E-services of Online Information Review 32(2), 163–178 (2008)

[20] Zeng, L., Benatallah, B., Ngu, A.H.H., Dumas, M., Kalagnanam, J., Chang, H.: QoS-aware middleware for Web services composition. IEEE Transactions on Software Engineering 30(5), 311–327 (2004)

Supporting Service Level Agreement Creation with Past Service Behavior Data

André Ludwig[1] and Marek Kowalkiewicz[2]

[1] University of Leipzig, Grimmaische Str. 12, 04109 Leipzig, Germany
ludwig@wifa.uni-leipzig.de
[2] SAP Research, 52 Merivale Street, South Brisbane, QLD 4101, Australia
marek.kowalkiewicz@sap.com

Abstract. Service Level Agreement (SLA) is the key instrument for formalizing contractual relationships between service providers and their customers. Besides the contractual application, SLAs can be used by service providers to plan, control, and monitor their service management activities. In order to support the definition of service level objectives in SLAs, past service behavior should be considered. Dynamic service profiles (DSP) are integrated information sources that encapsulate historical service execution data and aggregate it in a way that it can be used for predicting future service behavior. In this paper we describe a solution for applying DSPs to elements in SLA templates to support decision making in SLA creation and negotiation. Based on an example we show how DSPs, derived from past service behavior, are linked with SLA service level objectives.

Keywords: Service Level Agreement, Dynamic Service Profiling, Contracting, Service behavior.

1 Introduction

Service-oriented computing (SOC) has emerged as the most promising design paradigm for next-generation distributed information systems. The vision that goes along with SOC is that once standards have been established and become widely adopted by service providers and requesters, a globally available infrastructure for hosting and accessing services will be created [1]. This infrastructure will allow service providers to offer multiple services with individually adapted service capabilities to their changing customers that can dynamically and on-demand bind these services into their own applications.

However, exchanging services between service providers and requesters causes dependency and imponderability. If a service is sourced from an external service provider, the provisioning of the service is outside the influence area of the service requester. To prevent this, a contractual basis that defines the interface between service requesters and providers is required. The key concept for contracting between service providers and their customers is service level agreement (SLA). A SLA is a document that formally describes a provided service and responsibilities and guarantees of the involved parties. It usually consists of organizational elements (involved

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 375–385, 2009.

parties, contract duration etc.), service-related elements (functional and non-functional properties), and management-related elements (formal consequences of SLA non-compliance, SLA modification procedures etc.). Of these elements in particular non-functional properties are important factors to distinguish different SLA offerings. They refer to the quality of a service provision but also to security and financial aspects of a service. Quality of service usually addresses the performance under which a service is provided and is typically described by parameters such as availability, response time, and throughput. Security refers to the confidentiality and non-repudiation of a service by authenticating service participants, encrypting messages, and providing access control. Financial aspects of a service provision regard the costs that need to be paid for invoking a service.

In order to support the definition of service level objectives of non-functional properties in SLAs, past service behavior should be considered. Service level objectives of non-functional properties in SLAs should be defined in a way that they can really be provided by the service. Historical data that informs how service level objectives of previous SLAs were fulfilled during prior service executions must be collected and aggregated. Thus, conflicts between contracting partners that interrupt service provisioning and require extra resources to solve them can be reduced or avoided from the beginning. One way of taking historical data into account is by using Dynamic Service Profiles (DSP). Dynamic service profiles are integrated and preprocessed information sources that encapsulate historical service execution data and aggregate it in a way that it can be used for predicting future service behavior. They allow service requesters to gain more insight into the quality of services they acquire, offering more information than is usually given by service providers in service description documents. They can therefore be helpful in assessing the risk that the service provider will not fulfill the SLA obligations, or they can help in agreeing to such SLA values that are most likely to be achieved.

In this paper we describe a solution for supporting SLA creation and negotiation by connecting past service behavior aggregated in dynamic service profiles to SLA elements. The approach uses the mechanisms of the COmposite Sla MAnagement (COSMA) [2] approach for the connection of SLA elements with DSPs. After the conceptual model we show on an example how dynamic service profiles, derived from past service behavior, are linked with SLA service level objectives.

The paper is organized as follows: section 2 explains how SLA elements can be connected with dynamic service profiles utilizing the mechanisms of the COSMA approach. Section 3 summarizes the concept of dynamic service profiling and presents their structural outline. Section 4 presents related work and section 5 concludes the paper and gives a brief outlook to next steps.

2 Connecting SLA Elements with Dynamic Service Profiles

For the connection of SLA elements with DSP we utilize the mechanisms of the COmposite Sla MAnagement (COSMA) [2] approach which provides a general solution for managing SLAs in composite services. The central idea behind the COSMA approach is the integration of contractual information encapsulated in SLA documents as well as SLA management data into one composite SLA management document

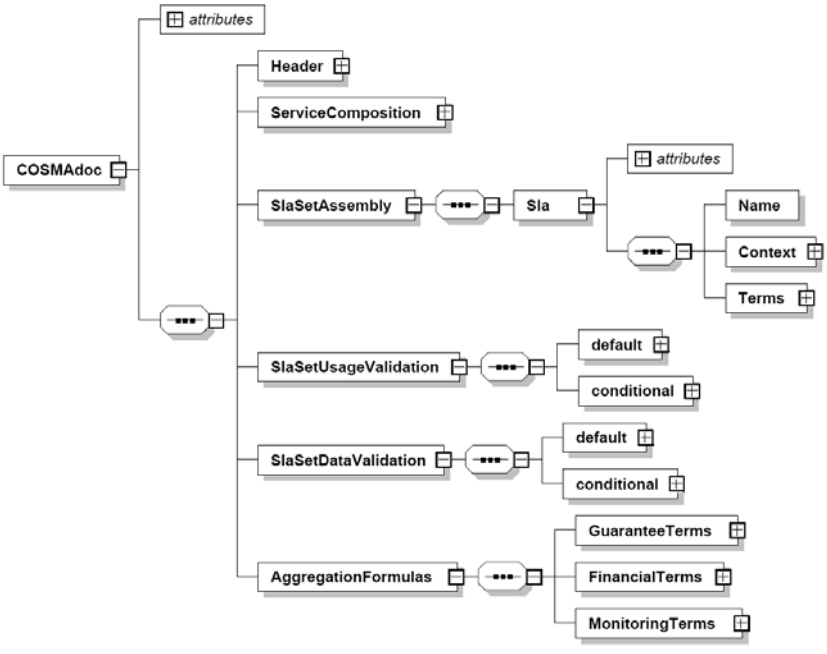


Fig. 1. COSMAdoc information model

(COSMAdoc). At the top-level, COSMAdoc consists of the core sections: Header, ServiceComposition, SlaSetAssembly, SlaSetUsageValidation, SlaSetDataValidation, and AggregationFormulas (Fig. 1). Following the separation of concerns paradigm, contractual information in COSMAdoc is encapsulated in SLA elements of the SlaSetAssembly section. Since they carry only contractual data, they can be exposed to involved parties without publishing COSMAdoc instance-internal management information. The COSMAdoc SLA model is based on the WS-Agreement SLA model [3] for bilateral agreements. It comprises of the sections Name, Context, Terms as defined in WS-Agreement. This means the SlaSetAssembly of COSMAdoc can be used to provide the SLAs to be connected with DSPs.

For the connection of SLA elements with SLA management data, COSMA defines the SlaSetUsageValidation and the SlaSetDataValidation sections. They are used to define specific requirements and constraints on the SLA elements of the SlaSetAssembly. These requirements and constraints regard either the usage or the content data of the involved SLA documents. For our purposes the SlaSetDataValidation section is relevant since DSPs contain data ranges and values rather than usage information (i.e. negotiability, necessity etc. of SLA elements). The SlaSetDataValidation section provides means to explicitly enforce, validate, and check the data values of the involved SLAs by defining predicates on them, i.e. setMaxValue, setValueRange etc. The technique used to identify elements of the SLA is the definition of pointers, i.e. setMaxValue(Pointer, Value). Predicates can be defined as default or depending on a condition.

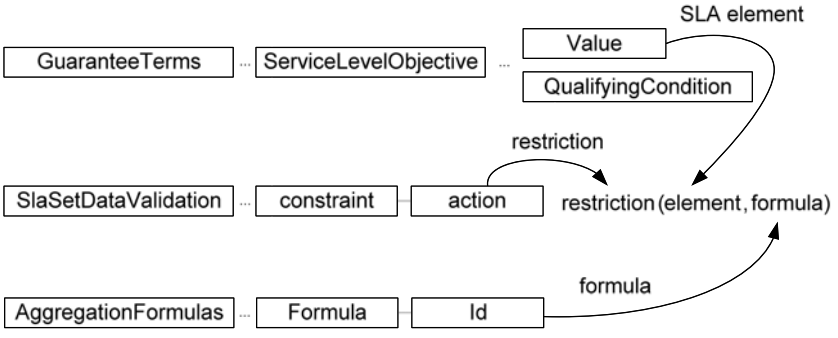


Fig. 2. Connecting SLA elements, data validation predicates and aggregation formulas

Since COSMA primarily addresses composite SLA management it supports that data validation restrictions on composite SLAs can be defined depending on contents of atomic SLAs. Therefore, predicates on composite SLA elements can refer to aggregation formulas that calculate the value from other atomic SLA elements. Thus, a predicate with a link to an aggregation formula connects and restricts values of SLAs, i.e. defined in GuaranteeTerms, with aggregation formulas defined in the AggregationFormulas section, i.e. setMaxValue(Pointer, FormulaId). Fig. 2 shows the general mechanism of connecting SLA elements with predicates that take their assignment value from aggregation formulas stored in the AggregationFormulas section.

For the connection of SLA elements with DSPs the same data validation restriction mechanisms can be applied. SLA elements of SLAs encapsulated in the SlaSetAssembly sections are connected with predicates defined in the SlaSetDataValidation section. These predicates recommend – rather than restrict – a certain fixed value, a value range, a maximum value or a minimum value. Therefore, we defined an additional set of predicates for the SlaSetDataValidation section that recommend certain data values retrieved from DSPs. They are summarized in Table 1. They can be used in the same syntax as existing predicates defined in COSMA using pointers for referring to SLA elements of the SlaSetAssembly and adding DSP values to the predicate, i.e. recommendValue(Pointer, Value). An example of a predicate that defines a minimal response time value recommendation of 500 milliseconds retrieved from DSP is shown below:

```
<constraint action="recommendMinValue(//Sla[@SlaId='1']
  /.../ServiceParameter[@Name='ResponseTime']/
  ServiceLevelObjective/Value,500)"/>
```

Table 1. Predicates that support DSP recommendations onto SLA elements

Predicate	Description
recommendValue()	assigns a pre-defined value to an element of a SLA
recommendMaxValue()	assigns a pre-defined maximum value to an element of a SLA
recommendMinValue()	assigns a pre-defined minimum value to an element of a SLA
recommendValueRange()	assigns a pre-defined range of values to an element of a SLA

Since DSPs are constantly being updated and processed (as it is explained in section 3), recommendation predicates must refer to dynamic service profiles directly rather than specifying a fixed value. For this reason, we added a new section to the COSMAdoc information model that stores dynamically changing DSPs.

Similar to connecting SLA elements with predicates that refer to aggregation formulas, SLA elements can be connected with the newly defined predicates which in turn refer to updated DSP values (Fig. 3). The DynamicServiceProfiles section consists of DSP elements which are structured as outlined in section 3.

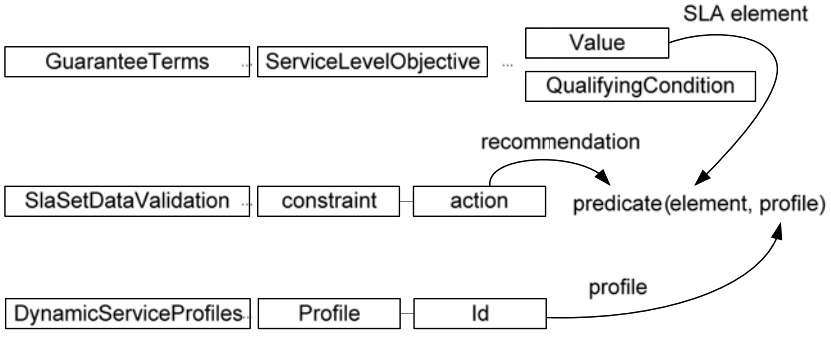


Fig. 3. Connecting SLA elements, data recommendation predicates, DSPs

An example of a predicate that defines a maximum availability value recommendation whereas the dynamically changing value is retrieved from a DSP defined in the DynamicServiceProfiles section (i.e. ProfileId='789') is as follows:

```
<constraint action="recommendMinValue(//Sla[@SlaId='1']
  /.../ServiceParameter[@Name='Availability']/
  ServiceLevelObjective/Value,//DynamicServiceProfiles
  /.../Value[@ProfileId='789'])"/>
```

3 Dynamic Service Profiles

Dynamic service profiles are data entities similar to service descriptions, containing Quality of Service (QoS) information about services. Contrary to service descriptions, information in dynamic service profiles is computed, based on historical execution data. Therefore, the QoS values in dynamic service profiles can change after each known execution of a service. Depending on needs, the dynamic service profiles can be computed based on long-term historical data or just recent information. It is also possible that the approach be mixed, for instance putting more stress on recent execution data, but not disregarding long-term historical data. One of the most promising applications of dynamic service profiles is SLA management, where they can be used to recommend SLA values, following an assumption that historical execution data can be a good indicator of future performance of services and composite service.

Creating dynamic service profiles involves discovery and computation of values of QoS parameters of services. This process should have minimal overhead but should

also be able to achieve sufficient trust by users as well as providers. When creating service profiling mechanisms to be connected with COSMA we followed and extended the aspects outlined by Liu et al. [4], i.e. fair and open QoS computation, preference-oriented service ranking, and using an extensible QoS model. We divide information about service execution into three main categories:

1. *static information*: values of service properties that do not change over time, such as name of the service, provided by the service provider;
2. *semi-static information*: values of service properties that may change over time, such as quality of service, price – this information changes periodically, but not very often (the choice here is purely up to the domain experts);
3. *dynamic information:* values of service properties that may be (and usually are) different for every execution of the service. It relates mainly to the network related Quality of Service.

Another categorization groups the considered properties into two categories, namely:

1. *simple properties:* values of service properties that can be monitored on an individual level. Such properties may include latency time, execution cost and so on;
2. *derived properties*: values where additional manipulation is needed (done by Service Profiling System). Such properties may include reliability, availability and so on.

Service profiles, as an up-to-date description of a service, include dynamically changing parameters of a (composed) service as well as static and semi-static information. A service profile includes both simple and derived properties. Both types of properties can be linked to SLAs and provide recommended constraints in COSMA.

Composite service profiles are aggregations of atomic service profiles. Description of a composite service profile is very similar to a service profile because it treats a composite service like an atomic service (just as sub-processes in process management can be treated in a similar way as atomic activities). That is why the structure of its profile does not differ significantly from the profile of an atomic service. However, the values of some parameters are computed as statistical measures on the basis of characteristics of atomic services included in the composed service.

During calculation of profile of atomic or composite service, a few data sources are taken into account:

1. *service repository*, where a service description provided by a service provider is stored;
2. *monitoring data*, passed in the form of execution logs as well as workflow events from the execution engines;
3. *data from SLA documents*, data coming from binding Service level agreement, that stores information about contracted QoS values that are later compared against real values coming from service execution. In consequence, when analyzing historical performance, it is possible to check to what extent the agreement between provider and consumer is fulfilled;
4. *user feedback*, data coming from end users of a service;
5. *other sources,* (e.g. third parties, authentications centers etc.).

The service profiling system uses various algorithms in order to compute values of parameters included in the profile. Often statistical methods are used, for example identifying mean, maximum, or minimum value. However, a service profile is not to be stored in any repository, but should be dynamically computed when a need for it occurs as each time different needs should be taken into account. Therefore, the recommended values in COSMAdoc might be recomputed and refreshed each time the COSMAdoc is accessed (if needed, the values from the time when the contract was settled can also be stored together with the document, to keep them for reference).

The excerpt of a service profile schema is presented below:

```
<?xml version="1.0" encoding="UTF-8"?>
  <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
    <xs:element name="ProfileData">
      <xs:complexType>
        <xs:element name="ServiceProfile" use="required">
          <xs:complexType>
            <xs:element name="BasicData" use="required">
              <xs:complexType>
                <xs:sequence>
                  <xs:element name = "WS-ID" type =
                     "xs:string" use="required"/>
                  <xs:element name = "WS-Price" type =
                     "xs:float" use="required"/>
                  <xs:element name = "WS-MinPrice" type =
                     "xs:float" use="required"/>
                  <xs:element name = "WS-MaxPrice" type =
                     "xs:float" use="required"/>
                  <xs:element name = "WS-
                     ExecutionDuration" type = "xs:float"
                     use="required"/>
                  <xs:element name = "WS-
                  ExecutionDurationFulfilment" type =
                  "xs:float" use="required"/>
                 <xs:element name = "WS-
                    MinExecutionDuration" type =
                    "xs:positiveInteger" use="required"/>
                  …
                </xs:sequence>
              </xs:complexType>
            </xs:element>
            …
          </xs:complexType>
        </xs:element>
      </xs:complexType>
  </xs:element>
</xs:schema>
```

Users may need a particular instance of a service only once in a given point of time, or they may need to use the service a few times in a given time period. Therefore, the horizon of the computation must be taken into account. In the first case, short-time forecast of service behavior (a short-term-behavior of the service) is important, and in the second case more attention should be paid to the long-term-behavior

of the service, taking into account historical (older) data. That is why service profiling system calculates the weighted mean value of the parameter using historical execution data in the following manner:

$$ED_j = \sum_{i=1}^{n} x_i * ED_i \quad (1)$$

Where:

- ED_j: execution duration for atomic service j
- I: 1 … n; n – number of service instances
- ED_i: value of execution duration from execution of the i-th instance of a service implementation
- x_i: the weight for the i-th instance of service implementation; weights sum up to 1 and depend on the horizon of the prognosis and of the moment of the observation

Execution duration for composite service is different. Two separate values might be computed:

1. On the basis of composite service execution data – using the same procedure as for atomic services.
2. On the basis of atomic services execution data of services included in e.g. the BPEL plan generated during service composition process. First the average values for each atomic service included in the composition are computed, then the plan is analyzed, the critical path is identified and the hypothetical value is computed as the sum of execution duration of services being on the critical path (e.g. "AND" – the highest value is chosen, "OR", "XOR" – average execution duration is taken).

Values of other parameters may be computed as a ratio e.g. for accessibility, which means the probability of a successful invocation of service. For atomic services it is computed by dividing number of successful invocations by all attempts to invoke the service in a given period of time.

$$A_j = \frac{m}{n} \quad (2)$$

Where:

- m: the number of all successful invocations of the service implementation in the given time period
- n: the number of all attempts (successful or not) to invoke the service implementation in the given time period.

Similarly to the execution duration, two separate values may be computed to estimate the value of accessibility parameter for composite service:

1. On the basis of composite service execution data - the same procedure as for atomic services.
2. On the basis of atomic services execution data of service implementations included in the BPEL plan generated during service composition. First the

accessibility values for each atomic service included in the composition are computed, then the plan is analyzed, the hypothetical value is computed as the product of atomic service accessibility value (e.g. "AND" – the product is taken, "OR", "XOR" – we assume that the probability of execution of each branch is the same (e.g. if there are two of them – 0,5, if there are three – 0,33333 etc.) etc.).

An entirely different algorithm is used to compute synthetic indicator value used in order to compare services and create their rankings. We propose a mathematical model of QoS computation based on multiple criteria analysis – deriving the synthetic indicator. The services are compared according to a few characteristics listed below. Each characteristic is assigned a weight, reflecting user's preferences. However, it is out of scope of this article to present all the algorithms and techniques used in order to create service profiles.

4 Related Work

As far as we know, there are no proposals that deal with connecting SLA elements with dynamic service profiles to support creation and negotiation of SLAs. Previously, in [7] we proposed a functional architecture for an adaptive management of quality of service aware service compositions and presented our architectural view onto a system that supports various execution strategies based on dynamic selection and negotiation of services, contracting based on service level agreements, service enactment with flexible support for exception handling, monitoring of service level objectives, and profiling of execution data. Dynamic service profiling and SLA management was not considered in an integrated fashion in this paper. Besides that, related work comprises related work in SLA lifecycle management in service-oriented computing environments and related work in dynamic service profiling.

In the area of SLA management, numerous approaches that provide extensive SLA language formalizations and management frameworks that can be applied to define contextual, functional and non-functional elements of SLAs are available. The most prominent examples of them are WSLA [5], WS-Agreement [3], and WSOL [6]. They are applicable for bilateral SLA management requirements. However, none of these approaches addresses how past service behavior can be considered in the definition of service level objectives in SLAs. An approach which allows the restriction of SLA elements by usage and data predicates – in particular to support the management of composite SLAs – was proposed by the COSMA approach. In our approach we exploit the general mechanisms of COSMA and apply them to the connection of DSPs to SLAs.

There are a number of initiatives aiming at computation of values of QoS parameters on the basis of data collected from different sources. Some researchers focus more on the sources of data, like e.g. service providers (service registries), users, verification centres, monitoring mechanisms [8, 9, 10], whereas others focus more on the algorithms and methods used to compute the values of both atomic and composite services [4, 11, 12, 13, 14]. There exist a few platforms that use and operate on such additional information like e.g. Web services filtering system [15]. However, to our best knowledge, the profiling system like the one presented in this article has not been yet developed.

5 Conclusions and Outlook

The connection of SLAs with past service behavior encapsulated in DSPs is needed, especially to support the creation and negotiation of SLAs and precautionary avoid non-SLA-conformant service behavior and its accompanying conflicts between contracting parties. Thus, in this paper we described a solution for connecting DSPs and SLAs by utilizing the mechanisms of the COSMA approach. We listed new predicates that need to be defined in COSMA for the connection and outlined the concept of dynamic service profiles for these purposes. Having developed the general mechanism to connect SLA and DSP, in future works we will develop an advanced lifecycle management approach for tasks such as: SLA monitoring and parallel building/maintaining of real-time DSPs, SLA renegotiations using these real-time DSPs, etc. Finally, the application of the presented solution on a number of services will lead to a widespread evaluation and assessment of the approach and will result in a broader range of practical experiences.

References

[1] Papazoglou, M.P.: Web Services: Principles and Technology. Prentice Hall, Essex (2007)
[2] Ludwig, A., Franczyk, B.: COSMA - An Approach for Managing SLAs in Composite Services. In: Bouguettaya, A., Krueger, I., Margaria, T. (eds.) ICSOC 2008. LNCS, vol. 5364, pp. 626–632. Springer, Heidelberg (2008)
[3] Andrieux, A., Czajkowski, K., Dan, A., Keahey, K., Ludwig, H., Nakata, T., Pruyne, J., Rofrano, J., Tuecke, S., Xu, M.: Web Service Agreement Specification (WS-Agreement), http://www.gridforum.org/documents/GFD.107.pdf
[4] Liu, Y., Ngu, A.H.H., Zeng, L.: QoS computation and Policing in Dynamic Web Service Selection. In: Proceedings of the 13th international WWW conference. ACM Press, New York (2004)
[5] Keller, A., Ludwig, H.: The WSLA Framework: Specifying and Monitoring Service Level Agreements for Web Services. J. of Network and Systems Management 11, 57–81 (2003)
[6] Tosic, V., Patel, K., Pagurek, B.: WSOL - Web Service Offerings Language. In: Bussler, C.J., McIlraith, S.A., Orlowska, M.E., Pernici, B., Yang, J. (eds.) CAiSE 2002 and WES 2002. LNCS, vol. 2512, pp. 57–67. Springer, Heidelberg (2002)
[7] Momotko, M., Gajewski, M., Ludwig, A., Kowalczyk, R., Kowalkiewicz, M., Zhang, J.Y.: Towards Adaptive Management of QoS-Aware Service Compositions – Functional Architecture. In: Dan, A., Lamersdorf, W. (eds.) ICSOC 2006. LNCS, vol. 4294, pp. 637–649. Springer, Heidelberg (2006)
[8] Casati, F., Castellanos, M., Dayal, U., Shan, M.-C.: Probabilistic, context-sensitive and goal-oriented service selection. In: ICSOC 2004. ACM Press, New York (2004)
[9] Ran, S.: A model for web services discovery with QoS. ACM SIGecom Exchanges 4(1), 1–10 (2003)
[10] Sheth, A., Cardoso, J., Miller, J., Kochut, K.: QoS for service-oriented middleware. In: Proceedings of the 6th World Multinconference on Systemics, Cybernetics and Informatics (SCI 2002), July 2002, pp. 528–534 (2002)
[11] Menasce, D.A.: QoS Issues in Web Services. IEEE Internet Computing 6(6) (2002)

[12] Cardoso, J., Sheth, A., Miller, J., Arnold, J., Kochut, K.: Quality of service for workflows and web service processes. Web Semantics: Science, Services and Agents on the World Wide Web 1(3), 281–308 (2004)
[13] Abramowicz, W., Kaczmarek, M., Zyskowski, D.: Duality in Web Services Reliability. In: The Proceedings of International Conference on Internet and Web Applications and Services (ICIW 2006). IEEE, Guadeloupe (2006)
[14] Cardoso, J.: Quality of Service and Semantic Composition of Workflows. PhD thesis, Universitry of Georgia (2002)
[15] Abramowicz, W., Godlewska, A., Gwizdała, J., Kaczmarek, M., Zyskowski, D.: Application-oriented Web Services Filtering. In: Proceedings of the International Conference on Next Generation Web Services Practices (NWeSP 2005), pp. 63–68. IEEE, Los Alamitos (August 2005)

Author Index