

Tochtermann, Klaus (Ed.); Maurer, Hermann (Ed.)

Proceedings — Published Version

I-KNOW '07: 7th international conference on knowledge management

Suggested Citation: Tochtermann, Klaus (Ed.); Maurer, Hermann (Ed.) (2007) : I-KNOW '07: 7th international conference on knowledge management, Verlag der Technischen Universität Graz, Graz

This Version is available at:

<https://hdl.handle.net/10419/44445>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

I-KNOW '07

7th International Conference on Knowledge Management

Preface	1
K. Tochtermann, H. Maurer	
Social Media Maps of Online Communities	5
M. Smith	
Community Equity – How to Implement and Measure a Social Capital System in Enterprise Communities	6
P. Reiser	
Collaborative Knowledge Visualization: Researching and Augmenting Visual Practices in Knowledge Work	7
M. Eppler	

Scientific Track

Task and Process Exploration and Modelling

- A Guideline for Modelling and Supporting Information Access Processes** 8

J. Huster, C. Hesselmann, A. Becks

- Semantically Integrating Heterogeneous Content: Applying Social Tagging as a Knowledge Management Tool for Process Model Development and Usage** 16

M. Prilla, T. Herrmann

- Semantic Based Knowledge Flow System in European Home Textile: A Process Oriented Approach with PROMOTE** 25

R. Woitsch, D. Karagiannis, H.-G. Fill, V. Blazevic

Knowledge Sharing using Social Software

- Experiences with Introducing and Using Social Software in a Distributed Project Team** 33

R. Slagter, L. Efimova

- Sharing Digital Resources and Metadata for Open and Flexible Knowledge Management Systems** 41

M. Memmel, R. Schirru

- Web 2.0 as Platform for Inclusive Universal Access in Cooperative Learning and Knowledge Sharing** 49

T. Pitner, M. Derntl, T. Hampel, R. Motschnig-Pitrik

Knowledge Management and Web 2.0

- Knowledge Management in the Web 2.0 – Semantic Positioning as an Integrated Approach** 57

P. Erren

- Design of Personalized Knowledge Management in Web 2.0 Network** 65

M. Ivanova, E. Vasilyeva, I. Beleviciute

- Towards an “Enterprise n+1” through the Use of Web 2.0 Design Patterns Enriched by Semantic Web Infrastructure** 71

P. Reiser, H. Story, F. Novak, A. Blumauer

Skills

- Making Expert Knowledge of Adaptations of E-Learning Material Available with Patterns** 79

B. Zimmermann, C. Rensing, R. Steinmetz

- Using Taxonomies to Support the Macro Design Process for the Production of Web Based Trainings** 87

A. Aqqal, C. Rensing, R. Steinmetz

Conceptual Data Model and Meta Framework: Modelling ICT Competence Profiles R. Povalej, P. Weiß	95
<i>Collaboration</i>	
Analyzing Wiki-based Networks to Improve Knowledge Processes in Organizations C. Müller, B. Meuthrath	103
Management, Interchange and Reproducible Execution of Sample Preparation Knowledge in Collaborative Research Scenarios. C. H. P. Durst, F. R. Ihmig, S. G. Shirley, H. Zimmermann	111
Flexible Office: Assignment of Office Space to Enhance Knowledge Work Productivity R. Maier, S. Thalmann, F. Bayer, M. Krüger, H. Nitz, A. Sandow	118
<i>Requirements</i>	
Retaining Knowledge Management Maturity Models: An Ontology-based Approach M. Hefke, F. Kleiner, A. Storckenmaier, O. Paulzen	128
Work Process Oriented Introduction of Knowledge Management: Reconsidering the Guidelines for SME T. Herrmann, G. Brandt-Herrmann, I. Jahnke	136
Towards Semantic based Requirements Engineering T. Riechert, K. Lauenroth, J. Lehmann, S. Auer	144
<i>Tools</i>	
Social Network Analysis for Innovation and Coordination R. Balancieri, R. Cuel, R. Carlos Dos Santos Pacheco	152
Information Exploration via Pen, Brush and Text Marker H. L. Trinkaus	160
ExpertFinder: Collaborative Expertise Localization W.-O. Huijsen, S. Driessen, D. Slijp	168
<i>Knowledge Discovery</i>	
An Orchestration Model for Knowledge Management Tools in SMEs C. Ploder, K. Fink	176
Knowledge Discovery Techniques Applied to Knowledge Management in Universities M. Oprea, I. Tudor, A. Tanasescu	184

Utilizing Lifecycle Information for Knowledge Document Management and Retrieval	192
L. Lehmann, T. Hildebrandt, C. Rensing, R. Steinmetz	
<i>Knowledge Governance</i>	
Governing Knowledge Risks – Design and Results of an Empirical Study	200
F. Bayer, R. Maier	
Knowledge Fusion: From Management to Mobilization	209
P. Keen, M. Tan	
<i>Reflections</i>	
Knowledge Management from an Ancient Chinese Point of View, or the Knowledge Flow through an Organisation	217
W. Riegler	
Emerging Data Mining Applications: Advantages and Threats	225
N. Kulathuramaiyer, H. Maurer	
<i>Knowledge Sharing</i>	
Front End Decision Making: How Initiators Decide to Reveal Their Knowledge about Radical Opportunities	233
D. Globocnik, S. Vorbach	
A Dynamic Factorial Model of Knowledge Exchange and Sharing – An Exploratory Perspective	245
H. . Lehmann, F. Lehner	
<i>Practice</i>	
Practices of Knowledge Management in Companies: A Turkey Survey	253
F. Ölçer	
Becoming Knowledge Focused: Developing a Practice for Managing Knowledge in International Organizations	263
U. Ugbor, C. Chanduvi-Suarez	

Integrating Working and Learning in Business

Introduction	271
S. N. Lindstaedt, T. Ley, A. Ulbrich	
Workplace Process Integrated Learning and Knowledge Organization	273
M. Stoll	
Deploying YAWL for Workflows in Workplace-embedded Learning	281
E. Godehardt, M. Doehring, A. Faatz, M. Goertz	
An Approach for On-Demand E-Learning to Support Knowledge Work	289
S. Thalmann, F. Enoksson	
CAM in Process Execution	297
M. Wolpers, E. Duval, G. Martin	
An Experiment on Task Performance Forecasting Based on the Experience of Different Tasks	305
J. Olivella	
A Semantic-based Integrated Solution to Personnel and Learning Needs	313
S. Colucci, T. Di Noia, E. Di Sciascio, F. M. Donini, A. Ragone	
Wikis as a Technology Fostering Knowledge Maturing: What We Can Learn from Wikipedia	321
S. Braun, A. Schmidt	

Knowledge Organisation and Semantic Technologies

Introduction	330
J. Liegl, M. Lux, V. Pammer, P. Scheir, M. Granitzer, S. N. Lindstaedt	
Incremental Approach to Error Explanations in Ontologies	332
P. Kremen, Z. Kouba	
AUTOMS-F: A Java Framework for Synthesizing Ontology Mapping Methods	340
A. Valarakos, V. Spiliopoulos, K. Kotis, G. Vouros	
O'CoP, an Ontology Dedicated to Communities of Practice	348
A. Tifous, A. El Ghali, A. Giboin, R. Dieng-Kuntz	
Conceptual Clustering of Social Bookmarking Sites	356
M. Grahl, A. Hotho, G. Stumme	
Integration and Semantic Enrichment of Explicit Knowledge through a Multimedia, Multi-source, Metadata-based Knowledge Artefact Repository	365
F. Mödritscher, R. Hoffmann, W. Klieber	

Knowledge Visualization and Knowledge Discovery

Introduction	373
V. Sabol, W. Kienreich, M. Granitzer, K. Andrews	
Printf in 4D User Interfaces	376
A. Oberbichler, L. Böszörményi	
Spatiotemporal Knowledge Visualization and Discovery in Dynamic Social Networks	384
R. Klamma, Y. Cao, M. Spaniol, Y. Leng	
GlobeMash: A Mashup for Accessing GLOBE	392
J. Klerkx, E. Duval	
Sky-Metaphor Visualisation for Self-Organising Maps	400
K. Latif, R. Mayer	
On the Generalized Box-Drawing of Trees	408
B. Stein, F. Benteler	
Usability of Diagrams for Group Knowledge Work: Toward an Analytic Description	416
S. Bresciani, M. Eppler	
Text Mining for Indication of Changes in Long-Term Market Trends	424
P. Kroha, T. Reichel, B. Krellner	
Service Oriented Information Supply Model for Knowledge Workers	432
B. Zaka, H. Maurer	
An Interactive Visualization Model for Competence Management: an Integrative Approach	440
L. Pareto, U. Lundh Snis	
Visualizing Semantic Structures to Support Reading	448
Y. Qiu, A. Elsayed	
Visual Tools Decipher Historic Artefacts Documentation	456
J.-Y. Blaise, I. Dudek	
Visualization of Rule Bases - The Overall Structure	464
V. Zacharias	

Short Papers

A Semantic-based Search Engine for Professional Knowledge S. Colucci, T. Di Noia, E. Di Sciascio, F. M. Donini, A. Ragone, M. Trizio	472
Benefits of an SOA-Implementation for KM - Illustrated by a Case Study W. Krass	476
Conception of Knowledge Management Supported by Information Technologies I. Beleviciute	480
Intellectual Capital Report Controlling W. Halb, M. Kraudinger	485
Knowledge Management Systems: Towards a Failure Theory B. Ababneh, J. S. . Edwards	489
Reconsidering Relationships for Knowledge Representation K. Weller, I. Peters	493
Tacit Knowledge Management (T-KM): The Hidden Agenda of Knowledge Management F. Lehner	497
The Implications of Expert Systems in Knowledge Management I. Hauer, C. Muntean	503

I-KNOW '07

7th International Conference on Knowledge Management

Preface

In today's information society, highly innovative knowledge technologies are widely accepted as key enabling technologies to support organizations in treating knowledge as a valuable asset. Organisations are aware that their survival and future success depend on their knowledge about the latest developments in the field and on the solutions they realise to effectively support the management of knowledge.

I-KNOW '07 is dedicated to the latest scientific trends in IT-based knowledge management. With more than 500 attendees per year I-KNOW could be established as the premier application-oriented conference on this topic in Europe. In 2007, Know-Center, Austria's competence center for knowledge management, organizes I-KNOW for the seventh time. But for the first time I-KNOW is part of the conference series TRIPLE-I. TRIPLE-I combines the three conferences I-KNOW, I-MEDIA and I-SEMANTICS reflecting the increasing importance and convergence of knowledge management, new media technologies and semantic technologies. This unique concept aims at bridging the gaps between the various communities and their technology fields.

One key element of I-KNOW is to function as a platform for networking. Community building is fostered in special interactive events designed around particular themes. Deliberately long breaks throughout the conference and social events in the evenings provide excellent opportunities for meeting and networking with researchers or practitioners from all over the world.

Furthermore, I-KNOW offers its attendees selected contributions reviewed by an international expert committee. This year, 56 full papers and eight short papers have been selected for publication in the conference proceedings of I-KNOW '07.

The I-KNOW program is structured as follows: In thematically focused events community building is supported. Experts and the interested audience are brought together for an in-depth discourse in a thematically focused area. These events include a workshop of the leading EC-funded projects dealing with the integration of learning and working, and for example special tracks on *Knowledge Visualization and Discovery* as well as *Knowledge Organisation and Semantic Technologies*.

These special events are complemented by presentations covering current trends and latest developments in knowledge management. The presentations cover amongst others the following areas:

- Knowledge Sharing
- Collaboration
- Task and Process Exploration and Modeling
- Knowledge Discovery
- Knowledge Governance

An event such as the I-KNOW '07 requires active support at different levels: We are grateful to our three invited keynote speakers, *Martin Eppler* (University Lugano, Switzerland), *Peter Reiser* (SUN Microsystems, Switzerland) and *Marc Smith* (Microsoft Research, USA) for sharing with our attendees their ideas about the future development of knowledge management, new media technologies and semantic technologies. Many thanks go to all authors who submitted their papers and of course to the program committee for their careful reviews. The contributions which were selected by our program committee are published in these proceedings. Revised versions of the best of these contributions will also appear in our electronic journal on universal knowledge management J.UCS (www.jucs.org). J.UCS supports the open access initiative for scientific literature and thus ensures the knowledge transfer towards the community.

We would like to thank our sponsors: insiders Wissensbasierte Systeme, punkt.netSercives, Sun Microsystems, UMA Information Technology, edicos, Pars Group and Go International – an initiative of the Austrian Federal Economic Chamber und Federal Ministry of Economics and Labour. Special thanks also go to Dana Kaiser for preparing and ensuring the high quality the conference proceedings. Finally, we also would like to thank the core group at Know-Center who is responsible for the entire successful organization of the event. This core group includes Anke Beckmann, Gisela Dösinger, Patrick Höfler, Alexander Stocker and Anita Wutte – Thank you for your continuous high motivation in organizing I-KNOW!

We are convinced that I-KNOW '07 will be a terrific event and we hope that the conference in Graz will provide you with new ideas for your research and with new opportunities for partnerships with other research groups.

Sincerely yours,



Klaus Tochtermann



Hermann Maurer

Graz, August 2007

Program Committee I-KNOW '07

- Albert Dietrich, University of Graz, Austria
- Back Andrea, University of St. Gallen, Switzerland
- Bonifacio Matteo, DISA - University of Trento, Italy
- Böhm Karsten, University of Applied Sciences Kufstein, Austria
- Dreher Heinz, Curtin University of Technology, Australia
- Duval Erik, University of Leuven, Belgium
- Eppler Martin, University of Lugano, Switzerland
- Filipe Joaquim, School of Technology of Setubal, Portugal
- Hicks David, Aalborg University Esbjerg, Denmark
- Kiryakov Atanas, Ontotex Lab, Sirma Group, Bulgaria
- Lehner Franz, University of Passau, Germany
- Li Sheng-Tun, National Cheng-Kung University, Taiwan
- Maier Ronald, Leopold-Franzens-Universität Innsbruck, Austria
- Meersman Robert, Free University of Brussels, Belgium
- Naeve Ambjörn, Royal Institute of Technology, Stockholm, Sweden
- Nejdil Wolfgang, L3S and University of Hannover, Germany
- Pitner Tomas, Masaryk University of Brno, Czech Republic
- Reinhardt Rüdiger, Management Center Innsbruck, Austria
- Rittberger Marc, DIPF - German Institute for International Educational Research, Germany
- Sabater Ramón, University of Murcia, Spain
- Schneider Kurt, University of Hannover, Germany
- Schneider Ursula, University of Graz, Austria
- Studer Rudi, University of Karlsruhe, Germany
- Thomas Jim, Pacific Northwest National Laboratory, USA
- Tolksdorf Robert, Free University of Berlin, Germany
- Tomek Ivan, Acadia University, Nova Scotia, Canada
- Tzagarakis Manolis, University of Patras, Greece
- Viedma Jose, Polytechnic University of Catalonia, Spain
- White Bebo, Stanford University, USA

Additional Reviewers for I-KNOW '07

- Andreas Abecker, Research Center for Information Technologies, Germany
- Thomas Augustin, University of Graz, Austria
- Gisela Dösinger, Know-Center Graz, Austria
- Peter Haase, University of Karlsruhe, Germany
- Irene Häntschel, University of St. Gallen, Switzerland
- Stefanie Hain, University of St. Gallen, Switzerland
- Mark Hefke, Research Center for Information Technologies, Germany
- Florian Heidecke, University of St. Gallen, Switzerland
- Frederik Huitfeldt, Aalborg Universitet Esbjerg, Denmark
- Michael Kickmeier-Rust, University of Graz, Austria
- Christophe Kunze, Research Center for Information Technologies, Germany
- Alexander Nussbaumer, University Graz, Austria
- Matthias Palmér, L3S, Germany
- Alexander Stocker, Know-Center Graz, Austria
- Markus Strohmaier, Graz University of Technology, Austria
- Valentin Zacharias, Research Center for Information Technologies, Germany

Social Media Maps of Online Communities

Keynote Speaker

Marc Smith

(Microsoft Research, USA
marc.smith@microsoft.com)

Social media, the products of the growing read/write web, leaves traces behind that can be gathered, processed and visualized. These pictures of individual and collective authorship patterns illuminate the ecologies of online communities, highlighting the roles people play as leaders, answer people, discussion participants, questioners, spammers, and flame warriors. An ecological model of these roles focuses attention on the need for integration of multiple roles for successful communities. Mobile devices are changing the community scene further, creating new forms of input generated by passive sensors, leaving behind detailed trails that can themselves be shared, searched, and aggregated.

About Marc Smith

Marc Smith is a Senior Research Sociologist leading the Community Technologies Group at Microsoft Research in Redmond, WA. His group focuses on computer-mediated collective action. Marc Smith studies and designs enhancements for social cyberspaces, in particular he is interested in the emergence of social organizations like communities in online conversation and annotation environments. With Peter Kollok he co-edited the book *Communities in Cyberspace* (Routledge). Co-edited with Peter Kollok, the book explores identity, social order and control, community structures, dynamics, and collective action in cyberspace.

Community Equity – How to Implement and Measure a Social Capital System in Enterprise Communities

Keynote Speaker

Peter Reiser

(Sun Microsystems, Switzerland
peter.reiser@sun.com)

Web 2.0 is a user centric phenomenon. People can express themselves through blogs, videos, podcasts etc. The community decides what is hot through means like rating, voting and commenting, creating a dynamic social value system where good content bubbles up and less interesting content bubbles down. But what are the opportunities of Web 2.0 for an enterprise? How can a company take advantage of the social dynamics of Web 2.0 and turn it into a business value and business advantage? Peter Reiser's talk describes the building blocks of a Community Equity Measurement system that SUN Microsystems is implementing as part of its Customer Engineering CE2.0 project.

The architecture is based on reusable web services, standard protocols (Restful, ATOM, WebDAV) and extensible set of web widgets. The community life cycle methodology describes the tools, roles and psychological dynamics of a community and provides a cook book how to build, sustain and archive measurable communities. The objective is to build a dynamic Social Capital system by measuring the contribution, participation, skills and roles equity a person can gain by actively engage in communities.

About Peter Reiser

Peter H. Reiser is a Principal Engineer at Sun Microsystems. He is currently leading the Web2.0 and social network implementation for the global technical community at Sun. Prior to this he was responsible for the Knowledge & Intellectual Capital Management for Sun's Global Sales Organization. In Sun Europe he established the first community based solution practices and was head of the e-Finance Competency Center, where he was responsible for the architecture and implementation of some of largest Internet Banking and e-Commerce solutions in Europe.

Collaborative Knowledge Visualization: Researching and Augmenting Visual Practices in Knowledge Work

Keynote Speaker

Martin Eppler

(University of Lugano, Switzerland
martin.eppler@lu.unisi.ch)

How can knowledge workers integrate their diverse knowledge more productively? How can experts from different fields share what they know efficiently and create breakthrough innovations? To answer this question, this keynote address focuses on one feasible solution: facilitating knowledge integration through collaborative real-time visualization. Recent research results as well as field tests with new tools show that visual practices can improve collaborative knowledge work significantly. But not all interactive visualization methods are equally conducive to knowledge creation or sharing. The framework and concepts presented in this keynote help to identify the key levers (and barriers) for effective visual practices in knowledge work. An outlook on future developments concludes the talk.

About Martin Eppler

Martin Eppler is Professor at the University of Lugano, prior vice director of the institute of media and communications management at the University of St. Gallen where he also led the Competence Center Enterprise Knowledge Medium, and 'Privatdozent' at the University of St. Gallen. His research focus is on information quality and knowledge management, his work experiences regard consulting and the media industry. His latest book publication is "Managing Information Quality" (Springer). A lot of his publications concern knowledge visualisation.

A Guideline for Modelling and Supporting Information Access Processes

Jessica Huster

(Fraunhofer-Institute for Applied Information Technology FIT, Germany
jessica.huster@fit.fraunhofer.de)

Christoph Hesselmann

(RWTH Aachen University)

Andreas Becks

(Fraunhofer-Institute for Applied Information Technology FIT, Germany
andreas.becks@fit.fraunhofer.de)

Abstract: In many companies search solutions span departments and functional operations, independent of individual working processes in particular departments. The efficient access to the required information in relevant sub-processes is a key factor for the performance of the enterprise. Optimizing the searching and finding of necessary information enables a better workflow, and thus reduces costs. Based on well known process modelling methods we present a guideline and methods to analyse and document relevant characteristics of information intensive processes. In a next step these observations will serve as a basis for a mapping between identified requirements and suitable methods and tools for an optimal support of the process.

Keywords: Search and Retrieval, Information; Search Process, Problem Solving, Control Methods, and Search

Categories: H.3.3, I.2.8

1 Introduction

The problems, arising from not finding the relevant information, are various and crucial [Dengel et al. 02], [Rese et al. 04]. They include increased work load and cost factors because the already existing information is not found and thus created again. Problems and errors appear in daily work and even wrong decisions are made because relevant information is not known or relations between different information can't be accessed and used.

Wasting time for finding the information needed in the current task is an important cost factor for companies. This problem even becomes more crucial as the relevant information amount nearly doubles each year [Dengel et al. 02]. According to a study of Fraunhofer IAO [IAO 99] the competitiveness of companies could be increased about 30 % in average, if the existing information would be used properly. In this paper we present a guideline and methods to analyse and document relevant characteristics of information intensive processes, in order to provide a common understanding of the companies' internal information access processes and its problems which is the basis for appropriate tool support.

2 Scientific background and related work

One reason for the problem of not finding the relevant information in the appropriate working context is the low integration of current information systems into the working processes [Dengel et al. 02]. This leads to additional effort when information, created or needed within the working context, is stored in or extracted from the information system. Another important factor is the individual itself [Albertoni et al. 05], as the individual with his background knowledge and experience influences the search. This experience should be considered when establishing search systems to optimize search results. Combining these two aspects means a working process as well as user oriented and integrated support for search. There are many search tools available and information retrieval is a well known research topic. But persons still have problems to find the information which is relevant for them in that particular situation because of a gap between the working situation and the used search tool on the other hand. Research in information retrieval does focus more on system-oriented aspects than on user context.

Kampffmeyer [Kampffmeyer 99] supports the access to information by structuring the document flows. To implement a document management system he suggests a project design starting with an analysis of the current situation by questionnaire and organigram. However this approach is focused on document management systems. The necessity to take the current working situation into account when establishing a search tool is identified [Järvelin and Ingwersen 04] [Freund and Toms 05] but not solved up to now. [Järvelin and Ingwersen 04] propose a cognitive framework to connect the two research fields information behaviour and information retrieval. The framework considers important aspects from both of the fields. [Freund and Toms 05] address the gap by developing a search system able to incorporate context to increase the relevance of search results. In this paper we would like to address this gap by proposing methods to systematically survey information access processes and perform a best match between the information access situation and current search systems.

Section 3 presents our approach to characterize actual situation of information access, taking into account the working context as well as the individual. To enable access to the relevant information in the different working processes the individual requirements in the situation of access are identified and recorded. These observations help to specify the requirements for the selection of suitable methods and tools to support the information access in that situation. In this paper we focus on the analysis phase. Initial evaluation was already performed in different companies.

3 Guideline - Overview

In this section we give an overview of each phase of our guideline. We focus on the first phase which aims at a detailed and systematic analysis and documentation of information access processes in the company. The outputs provide the basis for the following phases, the recommendation and selection of potential information access methods and search tools for support. Figure 1 depicts each process step and its results, as well as the methods involved in each step. The guideline basically includes

three phases. The first phase includes a detailed analysis and characterisation of information access processes as well as the context of the information access situation itself. We integrated task models from information retrieval, process analysis and user modelling approaches to get a detailed and structured method for the analysis of information access processes. In section 4 we describe each step of this phase in more detail. Phase two and three support the selection for an appropriate search tool based on the analysis results of the first phase. In phase two search tools are classified according to the information access methods and functionality they implement and offer, e.g. linguistic or statistical methods, mining methods such as clustering and classification. On the other hand the information access profile (result of phase one) characterises the requirements for an appropriate tool support. The classified search tools are evaluated and mapped according to these requirements. The last phase includes an evaluation of the proposed mappings according to relevant economic aspects and the final choice for a tool. The economic values depend on each company and may include the costs for a product, the type of the offered licensing model, the kind of offered support, as well as integration aspects into tools used within the company existing software.

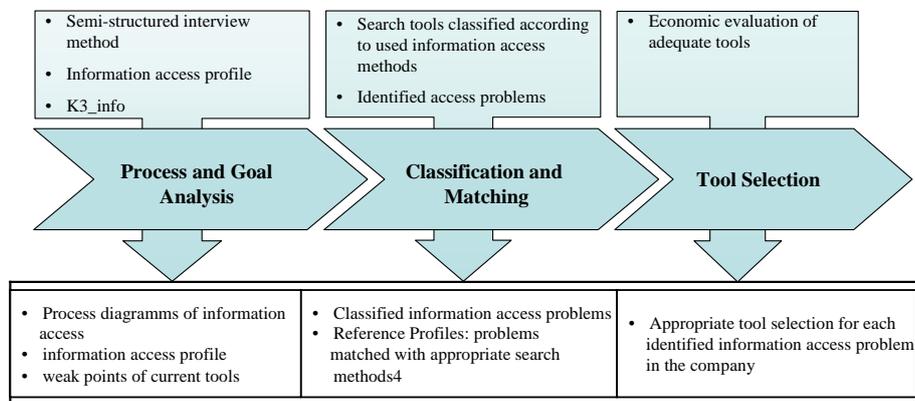


Figure 1: Guideline for Analysis and Selection of Appropriate Search Support

4 Process and Goal Analysis for Information Access

First of all, a detailed analysis of the information access processes is performed to provide an understanding of these processes and to systematically identify the current problems. The users involved in the daily working processes play an important role in this phase. Based on a *semi-structured interview* method they are questioned to survey and evaluate the processes and situation of information access, embedded in their task and working process. The identified parameters of the situation of information access are documented in the *information access profile* (section 4.1). The processes, the tools currently used for information access, as well as the current problems are documented throughout the interview based on *C3 process modelling* method [Foltz et al, 02] (section 4.2). This phase results in a documented and

evaluated process diagram, depicting also the situation of information access according the characteristics of the data accessed, the tools currently used to access the data, as well as the weak points and problems.

A *semi-structured interview* is related to qualitative analysis methods and aims at understanding a specific situation and the persons involved [Creswell 97]. The interviews are performed with representatives of the inquired field and are based on a interview structure, which on the one hand guides the interviewer and on the other hand gives enough freedom to the interviewee to depict his situation [Bortz and Döring 03], [Weßel et al. 06]. Figure 2 depicts our template for a semi-structured interview. The precise questions have to be prepared before the interview and adapted to that situation. But in principle the interview template lists the topics which have to be considered in each part. In part one, the interview starts with some general questions related to the interviewee and its working context in the company. The global parameters of the information access profile (section 4.1) are also surveyed. Part two concentrates on the current information access process and its documentation through C3_info. Part three focuses on the future situation. The interviewee is asked for suggestions and ideas for an improved information access in the future. Finally, in part four the interviewer summarises the main aspects to conclude the interview.

Two key factors in our guideline are important for the next steps.

- The recording of the parameters of the *information access profile* throughout the structured interview, as the *profile* combines important parameters to characterise the situation of access.
- The documentation of the working processes during the interview based on our extension C3_Info of C3 [Foltz et al, 02].

In the following section we describe each of the factors in more detail.

4.1 Information access profile

The information access profile was developed to describe the situation of information access and all relevant factors important in such a situation. Based on that profile the requirements for an appropriate tool support are deduced.

Figure 3 presents the profile and its four dimensions: *user*, *task*, *data*, and *environment*. These dimensions are distinguished according to global or local availability. Global dimensions, such as *user*, *task* and *environment* are valid for the whole considered working process in which the information access is embedded. Whereas the local dimension *data* describes the characteristics of each information source during the information access.

The user is the central actor [Oppermann 05] in the information access process. His role and qualification influence the activities performed throughout the process, in particular the situation of information access. An appropriate tool has to adapt to the experience and role of the user for efficient support. Furthermore, the user is involved in a specific task he has to fulfil and from which his information need occurs. The aim and type of task influences the kind and aim of information access [Oppermann 05]. Similar to [Picot and Rohrbach 95] who distinguishes three types of processes (routine process, regular process, one time process) the context of information access is characterized according to the number of iterations of the task. This parameter indicates the complexity and value of a task. A routine task is often not only easier, but also more valuable to support:

Part 1: Warm-Up
Introduce terminology: Processes in general, information access profile, information process
Interview partner: professional background, task, daily work
Organisation: vision, objectives, structure, business activities
Information access profile: document the global parameter of the information access profile
Part 2: Current Situation (annotated in C3_Info)
Overview of the working process: activities, dependencies, roles, communication, cooperation, access to information (overview), first soft spots
Reflection: Reporting of the recorded process by the interviewer, controlled by the interviewee, correction of the recorded model, if necessary
Focus on selected information access: exact activities, information sources, currently used tools, soft spots
Part 3: Future Situation concerning information access
Expectations and needs
Requests for new tools
Other ideas
Part 4: Conclusion
Interviewer summarizes important points of interview
Check for completeness through interviewee
Thanks and Leave taking

Figure 2: Guideline for Semi Structured Interview

dimension	parameter	specification	values
User (global)	qualification	experience in system handling	great, middle, low, not any
		background knowledge	high, middle, low
		frequency of use	frequently, rarely, never
	role	description	informal
		languages	#
Task (global)	objective	type	retrieval, additional information, specification of task
	type of process	type	project oriented, repeating, innovative
Data (local)	format	number of documents	#
		structure	structured, semi structured, unstructured
		spreading	one source vs. many sources
	time vs. quality	trade-off	scale
	dynamic	durability	unlimited, temporary
		change frequency	high, low, not any
	languages	number	#
which		enumeration of languages	
Environment (global)	location	working place	static, nomadic (mobile working environment)
		connection to data	online, offline
	time	working/ leisure time	working time, leisure time
		requirements on response time	high, middle, low
		time to formulate search query	#
	domain	description	informal
		type	open, closed
		spectrum	wide spreaded, limited/domain specific

Figure 3: Information Access Profile: Global and Local Dimensions, Parameters and its Values for Characterising the Situation of Information Access

- *project oriented*, the task and hence the embedded information access is related to a specific project and thus performed in a specific time horizon
- *repeating*, the task is part of daily business in the company, and hence performed repeatedly
- *innovative*, the task changes each time and the search as to be adapted to different needs

The dimension *data* specifies the kind of accessed data. Additional to the descriptive parameters format, document dynamic and document languages, which have a direct influence on appropriate search methods, a parameter representing the trade-off between time and quality is queried. Searching a big amount of data needs time, a restriction concerning the time factor may lead to suboptimal search results. In security sensitive situations it is essential to find all relevant documents hence the quality is much more important than the response time.

The last dimension environment combines further parameters which are relevant for the information access process. Beside the location and time of access, the search domain is important. A closed, precise describable domain can be searched easier than a domain comprising several topic areas and continuously changing.

4.2 Process documentation with C3_Info

Throughout the structured interview the information access process is documented. This is important due to two facts. First, the documentation during the interview provides a communication basis between interviewer and interviewee to improve the understanding of the considered process. The interviewee can give feedback on points not well understood by the interviewer. Second, the documentation serves as a basis to deduce requirements and evaluate each information access situation.

As an appropriate process modelling method we choose C3 [Foltz et al. 00]. This method, based on modelling language UML, is especially suitable for weak structured processes and considers process steps, the involved stakeholders as well as the coordination, cooperation, and communication among the stakeholders. The aim is to survey and depict the weak structured processes in cooperation with the stakeholders.

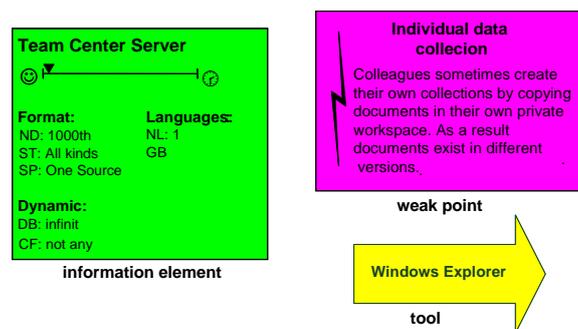


Figure 4: Additional Elements for Characterising the Situation of Information Access (C3_info)

[Schwarz et al. 01] mention the high degree of communication in knowledge intensive tasks. This factor can be considered using C3 because stakeholders are represented through their roles and the activities they fulfil within the process.

We extended the elements of C3 to document especially the characteristics of the information access. We call this extended version C3_info. The additional elements are shown in Figure 4. Each information access activity within the considered process is described in detail using these three elements. The green information element represents the local dimension data of the information access profile (see section 4.1). Additionally the currently applied tools are listed using the yellow arrow. The red boxes characterise the current problems of the information access situation. Here are the starting points for optimizing the access situation.

5 Evaluation

We have applied our guideline in two different industrial case studies (production industry, medical services) to survey the information access processes and its characteristics. We also performed a qualitative evaluation. The method was agreed on being intuitively understandable, easy to follow and providing good starting points for optimisation. The companies gain a common understanding of their (sometimes complex) working and information access processes. Problems of the current information access are systematically identified. Part 3 and 4 of the semi-structured interview (Figure 2) support the interviewer in identifying these weak points and problems, as the interviewee is often not able to directly formulate them. When the interviewer questions for wishes and optimisation of the situation the interviewee gives indirect hints to current shortcomings.

It is important to survey and document (roughly) the whole working process first before focusing on the information access itself and the local parameters. Coherences and dependencies to other working processes in the company are identified and may provide hints to first critical points.

6 Conclusions and Outlook

In this paper we presented an overview of our structured guideline in which we focused on the analysis and characterisation of the situation of information access. We have shown how best practice methods can be combined and used to systematically survey the characteristics of information access processes.

Further research work includes improvement and extension of the matching between the identified characteristics and weak points on the one hand and the requirements and appropriate search tools on the other hand.

Embedding the interview in the well known goal question metric approach better reveals the overall objective of the company. Based on the metrics the improvement made through the tool deposit of information access processes can be evaluated. Additional evaluation is performed in ongoing projects at the customer support.

References

- [Albertoni et al. 05] Albertoni, R., Bertone, A., De Martino, M.: Information Search: "The Challenge of Integrating Information Visualization and Semantic Web"; DEXA Workshops, Copenhagen, Denmark (2005), 529-533
- [Bortz and Döring 03] Bortz, J., Döring, N.: "Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler"; Springer, Berlin (2003)
- [Creswell 97] Creswell, J.W.: "Qualitative Inquiry and Research Design: Choosing among Five Traditions"; Sage (1997)
- [Dengel et al. 02] Dengel, A., Maus, H., Abecker, A.: "Dynamik betrieblicher Informationsprozesse - Adaptive Unternehmensgedächtnisse"; Networld-VDE Kongress (2002)
- [Foltz et al. 00] Foltz, C., Killich, S., Wolf, M.: "K3- User Guide"; Lehrstuhl und Institut für Arbeitswissenschaft der RWTH Aachen, (2000), 1-13
- [Freund and Toms 05] Freund, L., Toms, E. G.: "Contextual Search: From Information Behaviour to Information Retrieval"; Annual Conference of the Canadian Association for Information Science, Canada (2005)
- [IAO 99] "Wettbewerbsfaktor Wissen- Leitfaden zum Wissensmanagement"; Fraunhofer IAO, Deutsche Bank AG, Frankfurt/Main, (1999)
- [Järvelin and Ingwersen 04] Järvelin, K., Ingwersen, P.: "Information Seeking research needs extension towards tasks and technology"; Information Research, 10(1), (2004)
- [Kampffmeyer 99] Kampffmeyer, U.: "Strategien zur Einführung von Dokumenten-Management-Systemen"; Project Consult, Hamburg (1999)
- [Oppermann 05] Oppermann, R.: "From User-adaptive to Context-adaptive Information Systems"; i-com, 4, 3 (2005), 4-14
- [Picot and Rohrbach 95] Picot, A., Rohrbach, P.: "Organisatorische Aspekte von Workflow-Management-Systemen"; Information Management, 1, (1995)
- [Rese et al. 04] Rese, M., Gräfe, G., Herter, V.: "Relevance as an Information Quality Problem in Corporate Decision-Making Process"; Int. Workshop Data and Information Quality (DIQ04) 16th CAISE, Riga (Latvia), (2004), 25-36
- [Schwarz et al. 01] Schwarz S., Abecker, A., Maus, H., Sintek, M.: "Anforderungen an die Workflow-Unterstützung für wissensintensive Geschäftsprozesse"; WM2001-Workshop "Geschäftsprozessorientiertes Wissensmanagement – Von der Strategie zum Content", Baden-Baden, (2001)
- [Weßel et al. 06] Weßel, C., Weymann, F., Spreckelsen, C.: "Streamlining Qualitative Research Methods for Medical Informatics - A Methodological Approach"; Int. Congress of the European Federation for Medical Informatics (MIE), Maastricht (2006), 27-30

Semantically Integrating Heterogeneous Content: Applying Social Tagging as a Knowledge Management Tool for Process Model Development and Usage

Michael Prilla

(Information and Technology Management, Institute for Applied Work Sciences
Ruhr University of Bochum, Germany
michael.prilla@rub.de)

Thomas Herrmann

(Information and Technology Management, Institute for Applied Work Sciences
Ruhr University of Bochum, Germany
thomas.herrmann@rub.de)

Abstract: Process management is an important task in many companies. However, most of the literature on process management focuses on aspects like execution and monitoring and does neither deal with knowledge management support for the maintenance and contextualization of process models nor with the integration of such tasks into work procedures and corresponding tools. An effective knowledge management for business processes needs meta-data describing not only the processes but also their parts and details. This paper proposes a knowledge infrastructure for process modelling, usage and maintenance, which is based on a social tagging approach derived from popular social bookmarking tools. The concept of a tag-based prototype is described, which enables knowledge management support for complex sets of processes.

Keywords: Process Models, Knowledge Management, Semantics, Social Tagging

Categories: H.3.2, H.3.3, H.3.5, I.2.4, J.4

1 Introduction

Business process modelling is an essential activity of nowadays' companies. In such organizations, process models are used to define, analyze and optimize procedures. Furthermore, models are used to help employees doing their work by guiding them through a company's procedures.

Creating process models, working with them and maintaining them are difficult and knowledge intensive tasks. While modelling, several information sources have to be integrated, including e.g. documents providing knowledge on the working domain and surveys done preliminary to gather information. In a process model, all this knowledge is packed into a handy yet complex format. Therefore, process models used to document and facilitate organizational procedures include a substantial portion of an organization's knowledge. The complexity of this process-related knowledge is mainly caused by the multiple relationships of certain processes with practices and documents of the company and with its 'neighbour'-processes. Furthermore, every process includes several aspects such as the control flow, organizational structure and assigned resources and entities. It is not only complex to

make these multiple relationships explicit but it is also difficult to keep an overview over them, to make them searchable and to support the employees to internalize an understanding for the complex interrelations [Nonaka and Takeuchi 1991]. Therefore certain drawbacks exist when process models are used¹. Due to the density of knowledge packed into a set of process models, it is often hard to completely understand them for those ones using models in their daily work. Using plain names for process models and their elements imposes ambiguity and therefore makes understanding even harder. Moreover, in the multitude of process models describing various procedures in an organization, finding an appropriate process model is hard to accomplish. To understand this difficulty one has to be aware of the fact that sometimes there are only slight differences between processes which are not adequately mirrored by their names, and – even more difficult – that nearly every process includes several ramifications of the workflow and therefore does not represent only one but a multitude of procedures which somehow should be identifiable. The continuous need for adapting process models caused by changes in the underlying procedures amplifies this problem. Efforts like standardization of processes or reusing process components for describing other processes also suffer from these problems.

Taken together, throughout the entire process of creating and using processes, the need for supporting these activities by appropriate concepts of *knowledge management* becomes evident. An appropriate knowledge management concept should be able to take into account a variety of the aspects described above. Sadly, nowadays' knowledge management systems are mainly focussed on textual content. In the literature, several approaches proposing improvements for this situation can be found. However, these approaches fall short in providing knowledge management solutions linking existing organizational knowledge to process models and in integrating this support into work procedures of people. Therefore, some research towards supporting the creation and usage of process models still remains.

In this paper, an approach using a knowledge management framework applying semantic markup for process models and their components is presented. The approach is based on so called *tagging* well-known from nowadays' fast emerging social bookmarking sites like del.icio.us or flickr.com, in which hundreds of thousands of users share content by simply applying natural language keywords to it. Learning from the remarkable success of these sites, in this paper tagging is proposed as a mechanism to enable knowledge management for process model creation and usage. The corresponding approach presented in this paper is still work in progress. Therefore, there has not been a thorough evaluation of the prototype in a real setting. However, we consider the approach to be a valuable alternative to existing approaches.

In the remainder of this paper, we will describe how tagging can be used to build a knowledge infrastructure fostering work with process models. In [Section 2], we give a brief overview of related work and discuss differences and synergy effects between our approach and others. [Section 3] gives a description of tagging as a means of semantic description for knowledge management support of process modelling. After that, [Section 4] describes a framework and its prototypical

¹ Model usage here means both reusing models and their components for the creation, refinement or maintenance of models and using models for the guidance of work procedures.

implementation for supporting process model creation and usage. Furthermore, it demonstrates how tagging can be applied to models. The paper concludes with a discussion of the approach and a description of further research and development work.

2 Related Work

The approach presented here is inspired by and related to several fields of research. In the following, we briefly sketch some areas of influence, point out differences of our approach and show how it can be combined with existing approaches.

Process repositories for reuse of process parts: Reusing process models provides a means to use existing process parts in newly combined processes. Among these approaches, [Lang et al. 1996] propose “online libraries” to support this task and [Fettke et al. 2005] describe criteria for building up process model catalogues.

Management of process model aspects: Several approaches focus on the management of certain aspects of process models. Such aspects can be found in e.g. supporting process families [Rolland and Prakash 2007], ensuring quality compliance [Schmidt et al. 2007] and making design rationales explicit [Balabko et al. 2005].

Process-oriented knowledge management: Process-oriented knowledge management uses processes to structure the content of knowledge management systems. This way, knowledge can be provided for specific tasks [Maier and Remus 2001] and process models can be used to support browsing the content [Hoffmann et al. 2002].

(Semantic) Business process management: Business Process Management [Smith and Fingar 2003] mainly focuses on the management of process execution and monitoring. Semantic approaches applying ontologies as described in e.g. [Hepp and Roman 2007] and [Wetzstein et al. 2007] aim at enhancing existing systems.

Process model management and maintenance: Approaches supporting the management and maintenance of process models aim at collaborative creation of process models and their exchange. Examples of corresponding approaches can be found in [Thomas and Scheer 2006] and [vom Brocke and Thomas 2006].

Integration of Heterogeneous Content in Knowledge Management: There are several approaches addressing the problem of integrating heterogeneous content into knowledge management. Most of these approaches apply ontologies as semantic descriptions for such content [Fensel et al. 2000]. Existing approaches address e.g. the integrating of audiovisual content [Lux and Becker 2002], [Hammiche et al. 2004].

Though the approaches presented above all provide appropriate solutions for the problems they address, none of them provides a knowledge management solution combining organizational knowledge and support for a variety of aspects concerning process model creation and usage. Such support needs tight integration into work processes and must consider existing knowledge management systems and procedures.

Moreover, most of the approaches presented above use ontologies and similar fixed-scheme techniques for the description and management of process models and heterogeneous content. While these constructs provide a suitable mechanism for e.g. automation, they are rather complex and have to be properly understood in order to effectively use them, which induces a steep learning curve [Grudin 2006]. Furthermore, such techniques neglect two major factors in knowledge management

aiming at supporting human actors: *human perception* (cf. [von Ahn and Dabbish 2004]) and *differing usage of terminology* (cf. [Furnas et al. 1987]). Therefore, we suggest lightweight semantic techniques like social tagging to overcome such problems.

The approach described in this paper aims at contextualizing process models by integrating them into organizational knowledge management. Therefore, it may not only be used in a standalone manner, but may also be coupled to activities like business process management or process model management and corresponding systems.

3 Social Tagging to Support Process Modelling

As described above, the application of semantic descriptions in knowledge management often deals with ontologies or at least pre-defined metadata. According to the drawbacks associated with such techniques [Grudin 2006], it is doubtful that such pre-defined schemes can suit the needs of a multitude of users. We therefore propose social tagging as a lightweight alternative to these approaches. We are aware of existing drawbacks of tagging, including the lack of control concerning the resulting content structure and the uncertainty concerning the quality of descriptions produced by tags. However, we state that in supporting *users* in the integration of models and other content types, the advantages of tagging outperform existing disadvantages.

Users of these sites share content by simply applying free keywords (tags) to content units. These keywords allow for a completely different way of exchanging content compared to the concepts before the emergence of social tagging. Though tagging is a rather new phenomenon, there are several contributions providing insights into the ability of tagging in managing content.

One of the main advantages of tagging mentioned in the literature is its *low burden of usage*. With tagging, users do not have to learn complex classification schemes but use natural language keywords. While at first sight this seems to end up in a chaotic content collection, recent studies show that sound vocabularies emerge from tags [Marlow et al. 2006], which can be additionally fostered by suggesting tags to users [Sen et al. 2006]. Tagging provides a means of bottom-up collaborative content structuring and also bear the potential to be easily integrated into daily work practices, as it does not take much time to apply a tag to content.

Taking into account the variety of content organized in social bookmarking sites, tagging can be said to be capable of exchanging arbitrary content types, including URLs, images or even music [Sen et al. 2006]. The resulting vocabularies can be described as being both broad and precise enough to describe a domain's artefacts [Marlow et al. 2006], [Sen et al. 2006]. They are also both stable and concurrently adapting to current affordances [Golder and Huberman 2006]. Therefore we state that using tags in different applications provides the opportunity to interlink content of these applications and thus use content of one application in the other [cf. Section 4].

With respect to using tags as a knowledge management component for process models, two characteristics of tags play an important role. First, its *scalability* is relevant, meaning that tags provide a means to apply markup to a complete model, to its components and to single elements, thus producing layers of semantic descriptions

matching the structure of process models. Second, tagging enables *semantic and social navigation* of content [Golder and Huberman 2006], [Sen et al. 2006].

4 A Knowledge Management Infrastructure for Process Models

The approach to support usage and development of process models presented in this paper aims at semantically integrating process models with other content types in an organization's knowledge management. Applying the surplus expected from this integration means supporting corresponding activities of modelling right at the place they occur. This means that both content description and retrieval should be bound to tools dealing with the development and usage of process models. We state that, given the potentials of tagging as described in [Section 3], tagging provides a suitable and easy to use means to support the exchange of arbitrary content types and can therefore be seen as a *suitable knowledge management means* for this task.

From a technical viewpoint, the affordances of integrating tagging into existing systems can be considered rather low as there are a lot of frameworks available for this task. For the framework implementing our approach, we use the knowledge management system Kolumbus 2², which is already capable of attaching metadata to its content [Prilla and Ritterskamp 2005]. Currently, we are complementing this functionality by providing additional functions. In contrast to the implementation of tagging mechanisms, the integration of knowledge management for process models into activities of modelling is a more challenging task. Therefore, the remainder of this section focuses on a description of this task.

When process models are developed, activities such as receiving and perceiving relevant knowledge for this task, making it explicit and documenting it have to be seamlessly integrated. Therefore, modelling tools and knowledge management applications have to be integrated into a *knowledge infrastructure* for process models, which provides possibilities for dynamic exchange of information between these systems. With a mechanism allowing for such exchange a modeller can dynamically be provided with relevant information during the course of modelling.

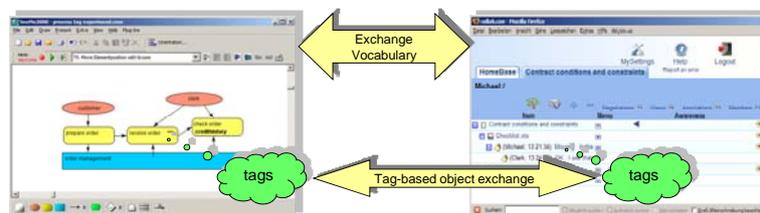


Figure 1: *Simplified Overview of Tag-based Content Exchange*

For the implementation of such an infrastructure, two main functionalities have to be available. First, the applications of the infrastructure have to *share* their tagging vocabulary. Second, content units such as documents, model elements and discussion threads have to be linked to each other via tags. [Figure 1] denotes such an exchange

² For further information on Kolumbus 2, please refer to <http://www.kolumbus2.de>

of content units between two applications used in a prototypical implementation of the framework. On the left, a model of a special modelling notation – SeeMe [Herrmann et al. 2004] – is shown as it is developed and displayed with an associated editor³. The editor is coupled to the knowledge management application Kolumbus 2, which is shown on the right. Since both applications have been developed in the same research group during the last years, there is a feasible opportunity to prototype and evaluate the functionality needed for the coupling with a tagging mechanism.

To enable tag-based content exchange, we provided the SeeMe modelling editor with a tagging component. As we aim at a seamless integration of tagging as an enabler of knowledge management and content exchange, we provide different tagging modes and visibility options for models [Figure 2]. First, tags applied to model elements can be displayed inline (A). Second, if necessary, tags can be hidden, which is depicted by a marker in the corresponding element (B). Additionally, tags can be shown on demand (C). Third, the mechanism allows for complementing tags with icons (D). The need for these mechanisms became obvious in several case studies where the modelling method was applied [e.g. Herrmann et al., 2004].

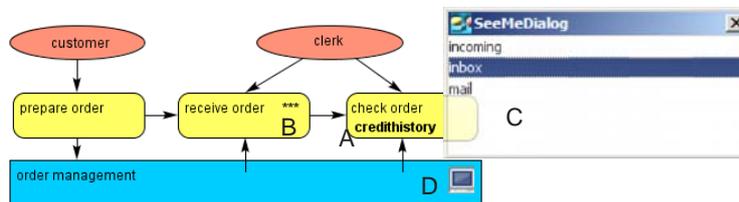


Figure 2: Tagging in the SeeMe Modelling Editor

To give an impression of the work with our framework, we provide a short scenario of its application in daily work practice, in which a modeller documents a work process. For the completion of the model, she uses information from several documents. While finishing her model, she links these documents to elements of the model in order to help others in understanding and using it. Additionally, she provides tags to the model and its elements to contextualize her work. After completing the model, she uploads it to a knowledge management system via her modelling editor. In the background, linked documents and tags describing the model are uploaded, too. In a corresponding system, the uploaded content is stored according to the tags given and related to existing content. Users can now find the model and content linked to it by browsing the content, using pre-defined filters or automated information agents. They may then provide additional tags to the model or add links to related content. After a certain amount of time, the knowledge management application contains hybrid content consisting of semantically related models and other content units. If later on another modeller deals with modelling a process and provides certain tags describing her model, her modelling tool will provide her with content matching the tags applied and thus provide information that may help with her work.

³ Further information is available at <http://www.seeme-imtm.de>

The approach described here uses web service technology to connect applications. In the corresponding framework, a knowledge management system provides a web service to access and query its content. Thus, client applications like modelling tools can receive content units and upload models and related information. For the information exchange between a web service and its clients, standard protocols and XML languages available for web services are used. Additionally, we use as special XML language for representing process models, which was extended for tagging purposes.

5 Summary and Further Work

Integrating knowledge management into the process of creating and using process models still needs research and implementation work to be done. The approach presented here aims at establishing such support by using a tagging mechanism. The applicability for knowledge exchange and structuring of tags has been shown in several studies. Tags can be easily used and integrated into daily work practice as they induce hardly any additional workload. The approach described here uses tagging to generate different levels of model descriptions and relations. It can be used as a standalone tool and to integrate and complement existing support approaches by applying suitable descriptions to the corresponding process models and their parts. Currently, the approach is still work in progress. Further work for its advancement will focus on design cycles between implementation, empirical evaluation and technical improvement of the prototype. Specific attention has to be directed refining the semantic exchange of content units.

Taking into account the success of social bookmarking services, using tagging as a knowledge management tool for process models and other content seems to be an obvious approach. However, though lots of users organize content with these services, tagging has not yet proven to be a suitable technique for the task described in this paper. Despite this uncertainty, we are confident to get positive results from the upcoming evaluation of our approach. Besides the effects on knowledge management and exchange, this evaluation will examine the group size needed for satisfying results of the approach, as social tagging is usually done in large user groups. Currently, a case study is intended to foster the documentation and usage of e-government processes in German government agencies. The evaluation of the prototype will continue with industrial field studies concerning processes in the service sector.

References

[Balabko et al. 2005] Balabko, P., Wegmann, A., Ruppen, A. and Clement, N.: "Capturing Design Rationale with Functional Decomposition of Roles in Business Processes Modeling"; *Software Process: Improvement and Practice*, 10, 4 (2005), 379-392.

[Fensel et al. 2000] Fensel, D., van Harmelen, F., Klein, M., Akkermans, H., Broekstra, J.; Fluit, C., van der Meer, J., Schnurr, H., Studer, R. and Hughes, J.: "On-To-Knowledge: Ontology-based Tools for Knowledge Management"; *Proc. eBusiness and eWork*, (2004), 18-20.

- [Fettke et al. 2005] Fettke, P., Loos, P. and Zwicker, J.: "Business Process Reference Models: Survey and Classification"; Proc. Workshop on Business Process Reference Models (BPRM), (2005), 1-15.
- [Furnas et al. 1987] Furnas, G. W., Landauer, T. K., Gomez, L. M. and Dumais, S. T.: "The Vocabulary Problem in Human-System Communication"; CACM (Communications of the ACM), 30, 11 (1987), 964-971.
- [Golder and Huberman 2006] Golder, S. and Huberman, B.: "Usage patterns of collaborative tagging systems"; JIS (Journal of Information Science), 32, 4 (2006), 198-208.
- [Grudin 2006] Grudin, J.: "Enterprise Knowledge Management and Emerging Technologies"; Proc. HICSS '06 (2006).
- [Hammiche et al. 2004] Hammiche, S., Benbernou, S., Hacid, M. And Vakali, A.: "Semantic retrieval of multimedia data"; Proc. 2nd ACM International Workshop on Multimedia databases (2004), 36-44.
- [Hepp and Roman 2007] Hepp, M. and Roman, D.: "An Ontology Framework for Semantic Business Process Management"; Proceedings of Wirtschaftsinformatik 2007, (2007), 423-440.
- [Herrmann et al. 2004] Herrmann, T., Hoffmann, M., Kunau, G. and Loser, Kai-Uwe: "A Modeling Method for the Development of Groupware Applications as Socio-Technical Systems"; Behaviour & Information Technology, 23, 2 (2004), 119-135.
- [Hoffmann et al. 2002] Hoffmann, M., Diefenbruch, M., Goesmann, T. and Herrmann, T.: "PRomisE2 - Gathering and Providing Situated Process Information in Knowledge Management Applications; Proceedings of I-KNOW 02 (2002), 170-176.
- [Lang et al. 1996] Lang, K., Taumann, W. and Bodendorf, F.: "Business Process Reengineering with Reusable Reference Process Building Blocks", Scholz-Reiter, B. and Stickel, E. (eds): "Business Process Modelling"; Springer (1996), 264-290.
- [Lux and Becker 2002] Lux, M. and Becker, J.: "XML and MPEG-7 for Interactive Annotation and Retrieval using Semantic Metadata", Proc. I-Know (2002), 965-984.
- [Maier and Remus 2001] Maier, R. and Remus, U.: "Towards a framework for knowledge management strategies: process orientation as strategic starting point"; Proceedings of the 34th Annual Hawaii International Conference on System Sciences (2001), 4023.
- [Marlow et al. 2006] Marlow, C., Naaman, M., Boyd, D. and Davis, M.: "HT06, tagging paper, taxonomy, Flickr, academic article, to read"; Proc. of the 17th conference on Hypertext and Hypermedia (2006), 31-40.
- [Nonaka and Takeuchi 1991] Nonaka, I. and Takeuchi, H.: "The Knowledge Creating Company"; Harvard Business Review, 69, 6 (1991), 96-104.
- [Prilla and Ritterskamp 2005] Prilla, M. and Ritterskamp, C.: "Playfully Assigning Keywords to Content". Proc. Workshop Computer Games and CSCW at ECSCW, Paris, (2005).
- [Rolland and Prakash 2007] Rolland, C. and Prakash, N.: "On the Adequate Modeling of Business Process Families"; Proceedings of 8th workshop on Business Process Modeling, Development and Support (2007).
- [Schmidt et al. 2007] Schmidt, R., Bartsch, C. and Oberhauser, R.: "Ontology-based representation of compliance requirements for service processes"; Proceedings of the Workshop on Semantic Business Process and Product Lifecycle Management (2007), 28-39.

[Sen et al. 2006] Sen, S., Lam, S.K., Al Manunur, R., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F.M. and Riedl, J.: “tagging, communities, vocabulary, evolution”; Proc. CSCW 2006, ACM Press (forthcoming).

[Smith and Fingar 2003] Smith, H. and Fingar, P.: “Business Process Management: The Third Wave”; Meghan-Kiffer Press (2003).

[Thomas and Scheer 2006] Thomas, O. and Scheer, A.: “Tool Support for the Collaborative Design of Reference Models. A Business Engineering Perspective”; Proc. HICSS '06, (2006).

[vom Brocke and Thomas 2006] Vom Brocke, J. and Thomas, O.: “Designing Infrastructures for Reusing Conceptual Models - A General Framework and its Application for Collaborative Reference Modelling”; Proc. Business Information Systems, Köllen (2006), 501-514.

[von Ahn and Dabbish 2004] von Ahn, L. and Dabbish, L.: “Labeling Images with a Computer Game”; Proc. CHI 2004, (2004), 319-326.

[Wetzstein et al. 2007] Wetzstein, B., Ma, Z., Filipowska, A., Kaczmarek, M., Bhiri, S., Losada, S., Lopez-Cobo, J. and Cicurel, L.: “Semantic Business Process Management: A Lifecycle Based Requirements Analysis; Proceedings of the Workshop on Semantic Business Process and Product Lifecycle Management (2007), 1-11.

Semantic Based Knowledge Flow System in European Home Textile: A Process Oriented Approach with PROMOTE

Robert Woitsch
(BOC Vienna, Austria
robert.woitsch@boc-eu.com)

Dimitris Karagiannis
(University of Vienna, Department Business and Knowledge Engineering, Austria
dk@dke.univie.ac.at)

Hans-Georg Fill
(University of Vienna, Department Business and Knowledge Engineering, Austria
hans-georg.fill@dke.univie.ac.at)

Vedran Blazevic
(BOC Vienna, Austria
vedran.blazevic@boc-eu.com)

Abstract: Semantic Technology is well established and mature enabling challenging and novel applications, but is rarely brought into real business applications. The EU-Project AsIsKnown (FP6-2005-028044) implements a show case of semantic flow systems in European home textile industry to solve two major challenges: (a) enable a homogeneous knowledge flow and (b) provide trend analysis. State of the art Semantic Web Technology is used for context, semantic and syntax transformation. One major challenge is the transformation from human knowledge into Ontologies. The PROMOTE[®] approach has been selected for the transformation from the expert knowledge into Ontologies. This paper introduces the project AsIsKnown and discusses the PROMOTE[®] approach to transform expert knowledge into Ontologies.

Keywords: Process Oriented Knowledge Management, Knowledge Management Services
Categories: H1

1 Introduction

The overall aim of the AsIsKnown project (FP6-2005-28044) [AIK, 06] is to develop a showcase for an integrated knowledge and content flow system between customer, trade facilities and producers through semantically-enabled Knowledge Technologies. This showcase will be exemplified by the home textiles industry, as it is an ever-moving target for the home textiles industry to continually change courses to keep pace with shifting consumers shopping trends [DoW, 05]. Current technology and approaches have a number of “knowledge barriers” such as isolated application, different data formats, insufficient standards, different views on the domain and last

not least what has been identified as a main risk: The ICT penetration in the home textile industry is quite low.

Key characteristics of European home textile industry - in contrast to the often confused clothing industry - are that European home textile market has nearly no import from outside Europe and has a high scrap production. The scrap production can reach up to 60%, as home textile products are mainly pre-produced. Although there is high expertise in trend analysis, the home textile industry would massively benefit by a better trend analysis and in better communication between the market participants. Another interesting characteristic of the high-price European home textile is that competitors have a strong differentiation, so the typical knowledge barrier "protection of knowledge" is only of minor relevance. The above characteristics makes the European home textile industry a very attractive showcase for Semantic Technology, as the semantic flow approach is promising and the current problems guarantee a high acceptance for new technology and establish a "win-win" situation for the participants of knowledge management (compare [Delic, 03]).

2 AsIsKnown Project Presentation

The project started in April 2006, with a project total cost of about 3.9 Mio Euros and a funding of the European Commission, Unit E2 [ECE2, 07] "Knowledge and Content Technology" of about 2.4 Mio Euros. The duration of the project is 36 months, the first review has been held end of April 2007 in Aachen. In the following we present the main idea of AsIsKnown and indicate the first result after the first third of the project.

2.1 The AsIsKnown Architecture

The AsIsKnown overall architecture has evolved during various technical and business related workshops. It aims to show both, the business aspects and the technical aspects. The focus of this paper is the technical aspect which will be elaborated in the following.

Figure 1 distinguishes the producers on the right side in members of AsIsKnown and non-members. It is the goal to establish a global knowledge management system that can be used by different types of members. Currently Vorwerk & CO Teppichwerke; Union Francaise Tapis Moquettes and the Textiles and Flooring Institute as consortium partners act as producers. They provide data about their products using traditional data bases, which are transformed syntactically and semantically into an agreed XML format set up by RWTH Aachen using state of the art technology. The ontology from Bulgarian Academy of Science is used to contextually translate the product data into the common "AsIsKnown language" to be used by all other AsIsKnown components. The Virtual Interior Designer is a commercial product of Crystal Design that provides 3D visualizations of the customers' room, and simulates the various home textile products. This system will be hosted by retailers, whereas INKU Nemzetközi Kereskedelmi is participating as a demonstrator. This axis is seen as state of the art technology that has been integrated in the first prototype in the first year of the project.

The challenge for the second prototype in the second year will concern novel feedback mechanisms of the Trend Analyzer and an ontology based orchestration of the system using an ontology repository and the Smart Profiler.

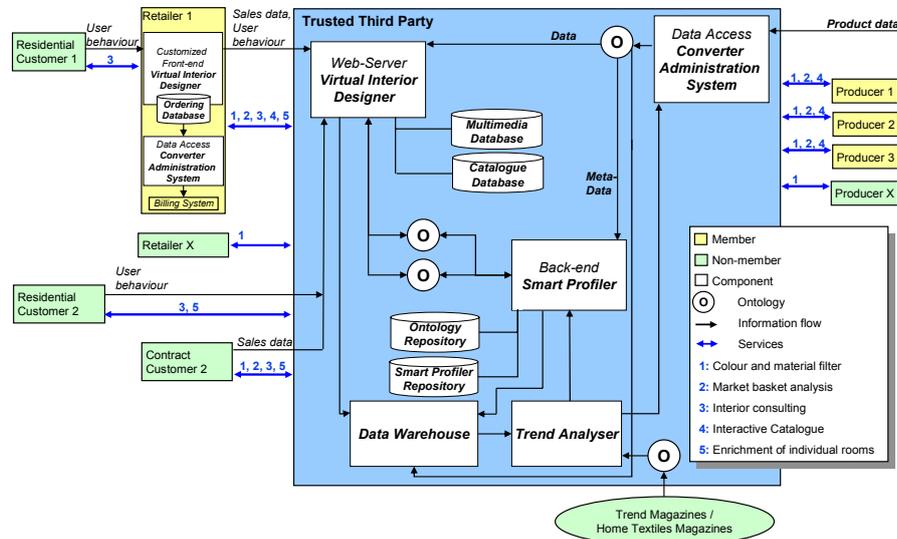


Figure 1: Concept Picture of AsIsKnown, describing the different types of members of AsIsKnown, depicting the key Knowledge Management Services and analyzing the need of the ontology and the various messaging formats

The Smart Profiler acts as the central repository of meta data that can be accessed by all AsIsKnown components. In this role the Smart Profiler also encapsulates the communication with the ontology management systems. This component is implemented by BOC using the PROMOTE[®] approach to provide meta information such as user profiles, product categories like themes or country dependencies and product characterizations to map between standards like colour codes or production types. The Trend Analyser collects available information from outside the system such as trend magazines and from inside the systems such as the data warehouse and various log databases. A visual trend analysis tool is provided by Fraunhofer Gesellschaft using data mining techniques to analyze the data and provide trends to the system. Results of the trend analysis - like the correlation between user profiles and product themes - will be integrated into the Smart Profiler in the form of models to configure AsIsKnown.

Conceptually the above AsIsKnown components are seen as Knowledge Management Services, although the technical implementation is not fully service oriented. The term Knowledge Management Services is used similar to [Woitsch, 04] and comparing to [Valente, 01] providing a logical set of functionality to a socio technical Knowledge Management System.

2.2 The AsIsKnown Key Objectives

The objective of AsIsKnown is to establish a knowledge flow within European home textile industry using Semantic Technology to establish a show case. Pioneer research is necessary for (a) the Trend Analyzer to develop a trend rule editor in combination with Ontologies, (b) the Smart Profiler for dynamic profiling using trend rules and Ontologies, (c) the Virtual Interior Designer for the semantic integration of web services and transparent knowledge resources and (d) for building the common sense ontology for home textile [DoW, 05]. The following section will discuss the Smart Profiler.

2.3 An Overview on the Smart Profiler

The Smart Profiler acts as a mediator between different knowledge bases by providing a common data format for home textiles knowledge, to provide a graphical modelling language making knowledge of the home textile industry explicit. Ontologies are currently more advanced from a technical oriented point of view, usable for ontology experts but not for end users that are seen as domain expert, who usually do not think in terms of concepts, instances or properties.

The overall goal is to develop a domain specific graphical ontology modelling framework involving domain experts by offering graphical representations of ontologies in their domain specific language.

In the following the work done for the Smart Profiler is described before discussing the approach in more detail.

First a collection and analysis of information made available by the domain experts has been performed including sample data, standards, tool-specific meta data, information material, workshops and interviews [D12, 06]. The end-user relevant model types have been built using PROMOTE[®] methodology applying a bottom-up approach.

The following sub-domains and appropriate modelling representations have been identified: (a) the sales conversation between customer and sales person can be modelled via process flows and business rules, (b) the customer profile can be modelled via profile models referring to characteristics and preferences, (c) the product categories can be modelled using concept maps and (d) the product can be described using concept maps or ontologies. PROMOTE[®] is used to implement domain specific graphical representation of the ontology to tighter involve domain experts in the ontology building process.

The current prototype of the Smart Profiler establishes the communication between the producer databases, the Converter Administration System and the Virtual Interior Designer. The second implementation cycle, considers the integration of ontologies and trend rules [D13, 06].

2.4 Related Work

The idea to involve domain experts into the building process of an ontology is based on the observation that currently there is a high dependency on few ontology experts and a time and cost-intensive period of ontology generation.

According to Uschold and King [Uschold, 96], Grüninger and Fox [Grüninger, 95] as well as Noy and McGuinness [Noy, 01] there are top-down,

bottom-up or inside-out approaches. The graphical meta models as an end-user interface extends the top-down approach in providing additional methods for end-user knowledge acquisition. According to the Methontology Methodology described in Fernández, Fómez-Pérez and Juristo [Fernández, 97] the knowledge acquisition phase – using expert interviews – will benefit in providing a graphical representation in the language of the expert. According to Swartout, Ramesh, Knight and Russ [Swartout, 97] describing the SENSUS methodology, the natural language processing will be supported by an end-user driven grouping of concepts via the graphical modelling and additionally providing relations to concepts by applying natural language processing in the textual description of the graphical objects.

In this sense the above introduced concept is not seen as a complimentary approach to ontology building, but as an extension of existing approaches to simplify the knowledge acquisition of end-users.

Current graphical representations of ontologies like OWLViz, Jambala, OntoViz, OntoVista, TGViz, Kaon, OL Graphs, RDF Gravity, Vizigator, VisoOWL visualise concepts, instances and relations to support the ontology engineer, but do not support a domain-specific graphical representation.

In this sense the above introduced concept of the Smart Profiler is seen as a new form of a domain specific graphical representation of an ontology using PROMOTE[®].

3 The PROMOTE Approach

PROMOTE has been developed in the EU project (1999-11658) [Karagiannis, 00] and since then continuously improved in commercial projects [Mak, 05] and in EU-Projects [Woitsch, 06]. The PROMOTE[®] modelling language [Woitsch, 04], had been extended by a new profile concept and OWL [W3C, 07].

Figure 2 gives an impression, how the PROMOTE[®] modelling language is used to describe the home textile theme “Toscana” indicated in the photo below. PROMOTE[®] was used for the expert workshop, by using semi-formal models to classify the different home textile themes. The PROMOTE[®] models have been specified using OWL syntax by an ontology expert who was analysing the PROMOTE[®] models. The sample in Figure 2 indicates that ontology concepts such as colours or patterns are related to the informal style like “Toscana”. This enables a “semantic-supported” search for home-textile products related to the theme “Toscana” without the need that the producer models (register) a direct relation between products and the styles. This “loose binding” principle is extended with discovery mechanisms in case there are no products found that are described by the referred concepts. State of the art mechanisms are used to reason the ontology.

PROMOTE[®] is implemented as a meta model according to the well known meta model layers by [Strahringer, 96] and comparing [Karagiannis, 07]. In this sense PROMOTE[®] is seen as an instance of a meta meta model such as MOF [OMG, 07] implementing the formal model specification of the PROMOTE meta model [Woitsch, 04].

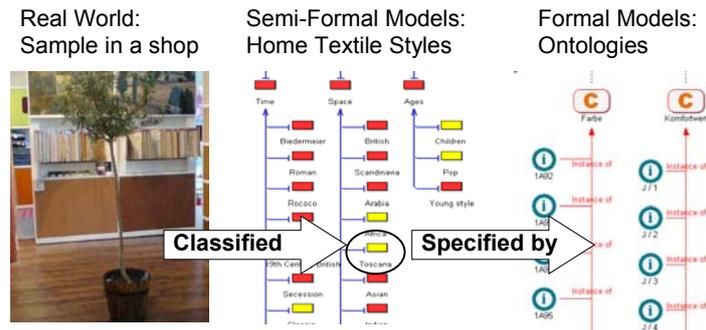


Figure 2: Sample of a PROMOTE[®] model, indicating the style “Toscana” on the left and the PROMOTE[®] formalism on the right by using a topic map representation to classify styles and specify each style with a set of OWL concepts

The integration between PROMOTE[®] and ontologies is seen as a transformation between models. When approaching the aspect of ontologies in modelling two basic directions can be identified in order to integrate ontologies with PROMOTE[®]. Ontologies can be seen as a meta model (vertical approach) or as a meta meta model (horizontal approach). For the Smart Profiler the vertical approach has been selected, interpreting an ontology as an alternative representation of knowledge models. This approach allows the use of a set of method integration patterns developed by [Kühn, 04] that provides a classification of approaches for integrating different methods from a meta modelling perspective modelling techniques, mechanism patterns, procedure pattern and patterns spanning multiple aspects for considering semantic, functional and graphical notation integration.

A trivial sample on *semantic integration* would be a 1:1 mapping from one meta model concept (CM) to one ontology concept (CO): $CM_1 \Leftrightarrow CO_1$. Interesting cases are a multiple dependency in one direction: $CM_1 \Rightarrow CO_1 \dots CO_n$ and multiple dependencies on various meta model levels: CMM_1 and M_2 and $MI_3 \Leftrightarrow CO_1$.

The *functional integration* is concerned with the situation that some functionality is better provided by meta models (e.g. simulation, analysis); some functionality is better provided by ontologies (e.g. reasoning, discovery). AsIsKnown will indicate the functionality and the algorithms applicable by providing e.g. discovery mechanisms to meta models ($f(MM)$) and simulation algorithms to ontologies ($f(O)$) like $f_1(\text{Application}) = \{f_1(O), f_1(MM)\}$.

The *graphical notation integration* is concerned with a PROMOTE[®] specific representation of ontology concepts.

The Smart Profiler enables therefore the cooperative modelling of knowledge models using PROMOTE[®] as a graphical user interface for domain experts and links the PROMOTE[®] models with ontologies using the integration concept. These mechanisms are defined as Design Time components of the Smart Profiler. The models are offered to other AsIsKnown components via the service Process Stepper, Search Engine and Knowledge Accessor, so that the Smart Profiler is seen as a mediator between other AsIsKnown components using knowledge models.

4 Conclusions and Future Work

The Smart Profiler is one component of the AsIsKnow system that builds a showcase for semantic based knowledge flow systems exemplified in European home textile industry. A key aspect of the Smart Profiler is the use of PROMOTE[®] models and ontologies during the Design Time and the provision of services that interpret the models for the Run Time of AsIsKnown. In the first implementation cycle the Smart Profiler implemented a model stack that represents the application field with PROMOTE[®] and indicates possible integration with ontologies.

In the second cycle the integration using a vertical approach will be implemented to support dynamic profiling using ontologies. The goal is to model sales process, user profiles, product categories and knowledge sources using the graphical modelling language PROMOTE[®] and integrate the AsIsKnown ontology to define the domain. The sales processes are seen as search processes within the domain that guides the user through the knowledge space. Here adaptive mechanisms introducing business rules will be applied to have flexible search processes following the approach of the FIT project (FP6-2004-27090) [FIT, 07] described in [Leutgeb, 07]. The implementation of the second cycle will be finished in May 2008.

Acknowledgements

We thank our project partners for their support in building the first prototype; especially Dr. Kiril Simov from Bulgarian Academy of Science who supported us in establishing the first integration with ontologies, Mr. Peter and Paul Aichner from Crystal Design who adapted their commercial tool to use ontologies and our coordinator Dr. Tobias Valtinat from Rheinisch-Westfälische Technische Hochschule Aachen who supports our approach in AsIsKnown.

Here we would like to mention the helpful and detailed support of our Project Officer Johan Hagman of the E2 Unit from the European Commission.

References

- [AIK 06] AsIsKnown Project (FP6-2005-28044) Homepage, <http://www.asisknown.org>, access: 20.05.2007
- [D12 06] Deliverable 12, Report on analysis of the data made available by partners, self published, 2006
- [D13 06] Deliverable 13, Report on design specification of Smart Profiler, self published, 2006
- [Delic 03] Delic K.A., Assessing the return on investment for How-to Databases, <http://www.knowledgeboard.com/cgi-bin/item.cgi?id=108093&d=101&dateformat=%o-%B>, access: 28.05.03
- [DoW 05] Description of Work of AsIsKnown (FP6-2005-28044), self published, 2005
- [ECE2, 07] EC - Unit E2, <http://cordis.europa.eu/ist/kct/index.html>, access: 20.05.07
- [Fernández, 97] Fernández M., Gómez-Pérez A., Juristo. N. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. Proceedings of AAAI97 Spring Symposium Series, Workshop on Ontological Engineering, pp. 33-40, 1997

- [FIT, 07] FIT Project (FP6-2004-27090) Homepage, <http://www.fit-project.org>, access: 05.05.07
- [Grüniger, 95] Grüniger M., Fox M., Methodology for the Design and Evaluation of Ontologies. Proceedings of IJCAI95's Workshop on Basic Ontological Issues in Knowledge Sharing, 1995
- [Karagiannis, 00] Karagiannis D., Telesko R., The EU-Project PROMOTE: A Process-oriented Approach for Knowledge Management, Proceedings of the Third International Conference on Practical Aspects of Knowledge Management (PAKM) October 30-31, 2000, Basel
- [Karagiannis, 06] Karagiannis, Dimitris, Höfferer, Peter 2006: Metamodels in Action: An overview, In: J. Filipe & B. Shishkov & M. Helfert (Eds.), ICSOFT 2006 - First International Conference on Software and Data Technologies: IS27-36. Setúbal: Insticc Press.
- [Kühn, 04] Kühn H., Methodenintegration im Business Engineering, PhD Thesis at University of Vienna, self published, 2004
- [Leutgeb, 07] Leutgeb A., Eichner H., Utz W., Woitsch R., Fill HG., Semantic and Adaptive E-Government Processes – A Field Report About Ontology Based Approaches in EU-Projects, in Proceedings of 5th EeGovDays - Prague, Czech Republic – April 13, 2007
- [Mak, 05] Mak K., Woitsch R., Der Einsatz des prozessorientierten Wissensmanagementwerkzeuges PROMOTE® in der Zentralkodokumentation der Landesverteidigungsakademie, Schriftenreihe der Landesverteidigungsakademie, 19/2005,
- [Noy, 01] Noy N. F., McGuinness D. L.. Ontology Development 101: A Guide to Creating Your First Ontology. Knowledge Systems Laboratory, 2001
- [OMG, 07] Object Management Group, Meta Object Facilities, <http://www.omg.org/mof/>, access: 07.05.07
- [Strahinger, 96] Strahinger S., Metamodellierung als Instrument des Methodenvergleichs: eine Evaluierung am Beispiel objectorientierter Analysemethoden. Berichte aus der Betriebswirtschaft. Shaker, Aachen, 1996
- [Swartout, 97] Swartout W., Ramesh P., Knight K., Russ T., Toward Distributed Use of Large-Scale Ontologies. Symposium on Ontological Engineering of AAAI. Stanford (California), 1997
- [Uschold, 95] Uschold M., King M., Towards a Methodology for Building Ontologies. Proceedings of IJCAI95's Workshop on Basic Ontological Issues in Knowledge Sharing, 1995
- [Valente, 01] Valente A. Housel T., „A Framework to Analyze and Compare Knowledge Management Tools“, In Proceedings of the Knowledge-Based Intelligent Information Engineering Systems and Allied Technologies (KES2001), IOS Press, Ohmsha 2001
- [W3C, 07] W3C, OWL - Web Ontology Language (Overview), <http://www.w3.org/TR/owl-features/>, access: 07.05.07
- [Woitsch, 04] Woitsch R., Process Oriented Knowledge Management: A Service Based Approach, PhD Thesis at University of Vienna, self published, 2004
- [Woitsch, 06] Woitsch R., Utz W., Roadmap to Akogrimo Convergence, A Sample of Process Oriented Knowledge Management with PROMOTE, In Proceeding of the BPOKI'06 (Business Process Oriented Knowledge Infrastructures), at I KNOW '06, Graz, Austria

Experiences with Introducing and Using Social Software in a Distributed Project Team

Robert Slagter

(Telematica Instituut, The Netherlands
robert.slagter@telin.nl)

Lilia Efimova

(Telematica Instituut, The Netherlands
lilia.efimova@telin.nl)

Abstract: Although co-located project teams from within one organisation can have difficulties to work together effectively and efficiently, the challenges are significantly increased in a team that consists of people from different organisations that work from different physical locations. In this paper we share our experiences of introducing new software tools (a project wiki and Skype) to support distributed project work and describe good practices, tips and tricks for the use of the new tools, also in relation to the existing toolset.

Keywords: social software, wiki, case-study, technology introduction

Categories: H.3.5, H.4.1, H.4.3, H.5.3

1 Introduction

Although co-located project teams from within one organisation can have difficulties to work together effectively and efficiently, the challenges are significantly increased in a team that consists of people from different organisations that work from different physical locations. Especially the process of introducing new technology support benefits from having the opportunity to sit together, show and discuss with each other how to use the new tools, and to share tips and tricks. In a distributed team, there is much less opportunity to do so.

In this paper we share our experiences of introducing new tools (a project wiki and Skype) to support distributed project work. Introducing these tools was not part of the planned project work, but rather a necessity to address specific collaboration needs discovered in the process of work. This paper is based on reflection and analysis of the use of these tools, performed together with the project participants at the end of the project.

We start this paper by presenting our case: the project team, its characteristics and how these characteristics are associated with technology (introduction) challenges. The paper then continues with an overview of how team collaboration was supported and how the new tools complemented the existing s: where they added value. The main part of the paper describes how we supported the introduction and use of the new tools in the distributed project team, and provides good practices, tips and tricks for the use of the new tools, also in relation to the existing toolset.

2 The context: introducing technology support in a distributed interorganisational project team

In this paper, we use examples from a project team called RUSMECO (www.rusmeco.net) to illustrate our experiences. This team worked together for two years in a research project of the European Commission. Some of the characteristics of the team made it a challenging setting to introduce new technology support, in particular the fact that the team members worked from different physical locations, worked for different organisations and had diverse professional and cultural backgrounds. In the following sections we will elaborate on these characteristics and what challenges for the introduction of new technologies they result in.

2.1 Distributed team members

The members of the RUSMECO team were located in 7 cities in 5 different countries, covering a total of 5 time-zones. This distributedness, together with the fact that the travel budget of the team was limited, made that the team had to depend heavily on technology-mediated collaboration. Additionally, it resulted in the following challenges:

- There were limited possibilities for synchronous interactions, given the time-zone differences.
- New technologies also had to be introduced in a technology-mediated way, instead of in a face-to-face setting.

2.2 Interorganisational project

In the RUSMECO project 12 organisations participated. The consortium of organisations did not work as a team before and only worked together for the duration of the project (2 years). Moreover, the team members were involved in other projects at the same time. These factors resulted in the following challenges for the introduction of new tools:

- There was a lack of shared technological infrastructure: different organisations used different tools for team collaboration.
- There were different policies with respect to installing and using technologies (e.g., some project members were not allowed to install software on their computers or had organisation-wide firewall settings that complicated the introduction of new tools).
- Members lacked shared practices about the use of technology in a project.
- There was limited motivation and resources to invest in installing new tools and learning how to use them, given the limited project duration.

2.3 Diverse backgrounds

The team members had very diverse backgrounds: both professional, cultural, as well as with respect to their experience of using collaboration technologies. The RUSMECO team was multidisciplinary and consisted of Russian and Western European members. Within the team, English was used for team-wide communication, although not all team members were fluent in English and the team

included only two native speakers. These diverse backgrounds of the team members resulted in the following challenges for the introduction of new tools:

- There was a lack of shared vocabulary and shared understanding.
- Given the different professional backgrounds, there were differences in how much coordination and communication was felt to be necessary & allowed.
- It was difficult to decide of the appropriate level of detail for training and tutorials, given the diverse technological backgrounds.
- In some cases tutorials and training material had to be made available both in English and Russian.

2.4 Overview of tools used by the RUSMECO team

One of the major challenges every team is faced with is to decide on what tools to use and for which purposes. This challenge exists since there is an abundance of tools and each type of tool is especially appropriate for only a part of the collaboration process. As described by media richness theory [Daft et al 1984] and task-technology fit [Zigurs et al 1998], there is no silver bullet in tool support for team collaboration: each tool has its strengths and weaknesses, and a collaborating team needs to find a mix of tools that fits their activities. In this section we discuss what tools were used by the members of the RUSMECO team and for what activities.

As indicated by the remark about face-to-face meetings, Table 1 shows how the team used different tools, which does not necessarily reflect their fitness for use. The table illustrates that e-mail was used for many types of team activities. Moreover, the shared workspace, originally intended to support a large part of the online activities, turned out to be used only as an archive of static information. The project wiki, which allowed for easy editing, was used for collecting input from multiple participants and for asynchronous coordinating activities in which multiple participants were involved, such as tracking to do items and deadlines, setting meeting agenda, etcetera. For more synchronous types of activities phone and Skype audio were used. In general, Skype audio was used less than the phone, also since a part of the team could not be reached by Skype, due to the settings of their organisational firewall. Note that the phone was also used for conflict resolution; even though communication on emotional topics is preferably done in face-to-face settings, the phone proved the next best option.

3 Experiences with introducing and using social software

In this section we describe our experiences with introducing new collaboration tools (a wiki and Skype) in the RUSMECO team, and how we stimulated the team members to actually use these tools. Secondly, the section describes tips, tricks and good practices for using these tools in distributed project teams.

3.1 Approach to introduce and stimulate the use of new collaboration tools

In our case, the introduction of new technologies was informed by the work of [Rogers 2003] on characteristics of innovation that support adoption: relative advantage, compatibility, ease of use, trialability and observable results. The introduction process was initiated by a few “*evangelists*”: team members with

extensive experience in using the new tools, who believed that they would address the collaboration needs better than the existing tool-mix. Introducing new technologies is not an easy task, especially in a case like this, when distributed settings and distributed authority do not make any “push” strategies possible. Thus, the new tools were introduced alongside the existing ones in a “let’s give it a try” manner. The evangelists used the tools for the project work they had to do, while making an effort to demonstrate these uses and the added value the new tools provided.

Team activity	Face-to-face ¹	E-mail	Phone	Wiki	Skype IM	Skype audio	Shared workspace	Other
Sharing information								
- one to one	*	***	**		***	**		
- with a sub-group	*	***	***	*	*	*		
- with the whole team	*	***		*			*	Newsletter
- with external people	**	***	**				*	
Collaboration								
- generating ideas, input collecting, commenting	*	**		***				
- integrating input for deliverables, refactoring				***				
- final editing, cross-indexing, referencing								Word
- working on software, presentations		*	*		**			Native applications
- archiving final versions							***	
Team coordination								
- team building, creating trust, shared understanding	*	**		**				
- decision making	*	**	***		***	**		
- conflict resolution	*	*	***			*		
- planning	*	**	**	**	**	*		
- to do items, deadlines and tracking status of activities	*	***	**	***				
- notifications & reminders		***	*					
In-person meetings								
- setting agenda		**	**	***	*	*		
- defining logistics		***	*	***	*			
- distributing minutes		***					**	
* seldom ** regularly *** often								

Table 1: Frequency of tool uses in the team

¹ As members of the team had limited opportunities to meet in person, face-to-face contact was seldom used for activities, even when this would have been most effective.

As there were limited possibilities for training at team meetings, every opportunity for face-to-face training and practicing with the new tools was seized. This involved, for example, an evening training session in the lobby of a hotel where most project members were staying. During this session the project members could practice with new tools and discuss with the evangelists where the tools would add value for their specific tasks and circumstances. In addition, tutorials with basic steps were created and made available for the project members.

For tasks that involved users and non-users of the new tools, an approach of combining tools was used. For instance, when input for a report was needed, all team members would receive an email with the request and a link to the wiki page to add their contributions. Those team members who could not access the wiki (because of settings on their corporate firewall) received a separate e-mail, stating that they could submit their contributions by e-mail to the editor as well. A similar approach was taken with combining phone and Skype calls in the same audio-conference. Although this created some extra work for the organisers, it stimulated project members to work with the new tools, thus providing an opportunity to experience the value of the tools.

3.2 Wikis in project teams: fitness of use, basic rules and tips & tricks

To facilitate the process of selecting technology for (a part of) a collaboration process, it is important to know its fitness for use. Our experience is that wikis are especially good for gathering, sharing and discussing textual and/or graphical information with a team. This makes them suitable for organising or planning events, reporting on project events and working on project deliverables. Wikis are good for activities where multiple people participate, but mostly not at the same time, such as collecting material, gathering input for a decision making process, creating and maintaining an overview of action points and, to a lesser degree, brainstorming. Wikis are suitable when there is a clear division of labour and clear responsibilities on who should work on what part of the content. They work especially well when a deadline drives participation (in line with [Grudin 2006]).

Wikis are not suitable for synchronous collaboration processes where multiple people contribute at the same time. They do not particularly fit the converging stages in a collaboration process: for integrating input and editing contributions it may be better to export the contents to a normal word processor. As in a wiki there is no single place to edit the structure, this makes them less appropriate for collecting structured, linked information (like pages with links to the previous section, next section and one level up) that requires frequent restructuring.

When introducing wikis in a project team, it is important to explain team members the following basic rules for collaboration in wikis [Burns et al 2006]:

- Keep wiki pages conversational in tone and written from a neutral point of view.
- Be polite. Expect your pages to be modified and be respectful of others' opinions. Work toward agreement.
- Don't delete other people's contributions; Instead, consider for instance moving contributions to a separate discussion page.
- Break down complex topics into multiple pages where possible. This encourages reusability at the page level.

- Provide a context for your pages and hyperlinks. Instead of “More info about etiquette can be found here”, use “More info about etiquette can be found at WikiEtiquette”.

Based on our experiences, we have formulated the following tips and tricks for introducing and using wikis in project teams:

- Create accounts for less experienced people and send them the password (with a link of where to change it), even if the system allows them to create their own accounts.
- Invest time in putting some initial information and pages on a new wiki, before inviting less experienced project members. They need this as a reference what the wiki could mean for them.
- Provide online support to help less experienced team members with editing the wiki.
- Use the wiki home page to link to frequently needed pages. Allow all team members to edit this home page. Within the RUSMECO project, the home page was edited 91 times, as new topics were being discussed and new pages became “hot”.
- For fast-paced processes, choose a wiki engine with a notification mechanism, such as RSS feeds or e-mail. This helps to notify relevant team members of updates to the wiki content.
- Establish conventions for authoring and commenting. A convention could for instance be to place comments in a different colour and always start with your name. Similarly, establish conventions for indicating the status of a page. On some pages contributions or comments are required, but other pages may contain stable information that should not be edited anymore. Make this explicit.
- Use attachments to wiki pages to store related material, such as sources of images, the final Word document or PDF file.

3.3 Skype in project teams: fitness of use, basic rules and tips & tricks

Regarding the fitness of use of Skype, our experience is that:

- Skype audio conferencing is a less expensive alternative to phone calls.
- The presence awareness indicators in Skype are useful to pick suitable moments for ad-hoc contact when there are limited possibilities for synchronous interactions.
- The instant messaging (IM) functionality of Skype is especially useful for ad-hoc coordination, while the voice communication is useful for longer conversations.
- Quick coordination via IM only works well between people who already know each other and need little introduction to the question.
- Communicating sensitive / emotional messages should not be done using IM.

Based on our experiences with introducing the IM and audio conferencing functionality of Skype, we have formulated the following good practices:

- Help inexperienced users by installing the software on their computer and

make sure the settings are correct. For instance, make sure that double click starts the chat function, instead of launching an audio call.

- Make sure users know how to connect a headphone and test whether it actually works.
- Keep track of the login names of all project members and share this information within the team (use the Skype *share contacts* option for this); provide a short description on how to add contacts to your contact list.
- Teach people about the meaning of the presence indicators and how they can indicate their availability for communication.
- Create a tutorial on the frequently needed functions of Skype and distribute this to the project members.
- It is less intrusive to use the instant messaging (IM) functionality to check whether someone is available for communication, before calling. Hence make sure that double clicking on a name in Skype starts the chat function.
- It is practical to use the IM functionality to exchange hyperlinks with team members when having a voice conversation.
- It is possible to include a normal phone in a Skype voice conference: use this for instance in case certain team members cannot use Skype.
- When adding another person to an ongoing chat conversation, copying the relevant part of the chat history helps to bring the newcomer up to speed.
- Show team members how much money they could save on their phone bills!

Notably, the options in Skype for file exchange were hardly used and video conferencing was even never used by the team. File exchange was difficult as it was blocked by some of the corporate firewalls. The reasons why videoconferencing was not used may include that setting it up requires additional effort and that some team members did not have the bandwidth required.

4 Discussion and conclusions

In this paper we share our experiences of introducing two particular tools in a distributed, inter-organisational project team: a project wiki and Skype, to support distributed project work. While the introduction of these tools was not part of the planned project work, it happened in response to specific collaboration needs discovered during the project. The introduction was triggered by a small group of project members, the *evangelists*, who observed these collaboration needs and were convinced that the new tools would fulfil these needs better than the existing tool mix. As the introduction was not a research objective, no formal fitness-of-use evaluation or team effectiveness measurement was performed. Instead, this paper is based on analysis and reflection on the use of these tools, performed together with the project participants at the end of the project.

The fact that the RUSMECO team had to decide what tools to use for which team activities is a challenge that other distributed inter-organisational project teams are faced with as well. With new tools, in our case the wiki and Skype, a team has no experience yet for which activities these tools add value. We aim to support such teams by formulating fitness-of-use criteria for wikis and Skype, provide suggestions

for the introduction process, as well as tips and tricks for using these tools. At this point, we would like to stress the essential role of evangelists in the process to introduce tools, to show the possibilities, to help other team members and to stimulate tool use: it takes energy to overcome the natural inertia of teams to adopt change.

Further research includes defining more elaborate fitness-of-use criteria for these and other software tools to support team activities, sharing success stories and good practices of introducing and using new tools.

Acknowledgements

This work has been partly funded by the European Commission through IST Project RUSMECO: Enhancing Russian SME collaboration and business development through Communities of Practice (No. IST-1999-29107). We would like to thank our colleagues in RUSMECO, in particular Liza Wohlfart, Patricia Wolf and Tanya Emshanova, for their contributions to the work described in this paper.

References

- [Burns et al 2006] Burns & Janicki (2006). A Case for Personal Knowledge Management in the Information Systems Curriculum. *Information Systems Education Journal*, 4 (111). <http://isedj.org/4/111/>. ISSN: 1545-679X. (Also appears in *The Proceedings of ISECON 2005*: §2543. ISSN: 1542-7382.)
- [Daft et al 1984] Daft, R.L. & Lengel, R.H. (1984). Information richness: a new approach to managerial behavior and organizational design. In: Cummings, L.L. & Staw, B.M. (Eds.), *Research in organizational behavior* 6, (191-233). Homewood, IL: JAI Press.
- [Grudin 2006] Grudin, J. (2006). Enterprise Knowledge Management and Emerging Technologies. In *Proceedings of the 39th Annual Hawaii international Conference on System Sciences - Volume 03. HICSS*. IEEE Computer Society, Washington, DC, pp. 57.1.
- [Rogers 2003] Rogers, E. M. (2003). *Diffusion of Innovation*, Fifth Edition. Free Press, New York.
- [Wikipedia 2007] Wikipedia (2007). Social software. Retrieved 16 May 2007 from: http://en.wikipedia.org/wiki/Social_software
- [Zigurs et al 1998] Zigurs, I., & Buckland, B. K. (1998), A theory of task/technology fit and group support systems effectiveness, *MIS Quarterly*, 22, 3, 313-334.

Sharing Digital Resources and Metadata for Open and Flexible Knowledge Management Systems

Martin Memmel and Rafael Schirru

(Knowledge Management Department,
German Research Center for Artificial Intelligence DFKI GmbH
e-mail: {mommel,schirru}@dfki.de)

Abstract: This paper discusses the requirements of a framework for sharing digital resources and metadata to meet the needs of open, flexible Knowledge Management solutions. The changing nature of the Web and its users as observed in recent years clearly establishes the need for new approaches and technologies to fully exploit the potential for working with existing digital resources. Formal metadata about the resources can be combined with information created in lightweight and user-centric approaches in order to significantly enhance resource descriptions and enable more efficient access to existing knowledge. The ALOE system, currently in development at DFKI, is one such solution and it is used here as the basis for a sample realization of an appropriate framework.

Key Words: digital resources, knowledge management, knowledge sharing, lifelong learning, metadata, resource profiles, social software, web 2.0

Category: H.3.0, H3.2, H3.3, H3.4, H3.5, H.5.0, H5.1, H5.2

1 Introduction

Today, the number of digital resources is growing at a speed never seen before [Lyman, Varian (03)], and people are confronted with the need to learn throughout their lifespan. In this situation, the role of Knowledge Management and lifelong learning becomes more and more important. It should support people in collaborating with others, and in getting just the right content in just the right time without having to leave the current task and context. Therefore, being able to share digital resources and information about them is a key factor for the success of any system trying to support users in their daily tasks.

Not only the amount of available resources has to be considered – it also has to be taken into account that the way content is created and accessed by the users has fundamentally changed in the last years. The Web is continuously evolving from a place where information was usually only consumed by most of the users to a more social and participatory system. More and more users contribute information in the Web (so called ‘user generated content’), using tools like blogs, wikis, social bookmarking services or file sharing platforms such as Flickr and Youtube. Usually, this group of technologies, platforms and tools facilitating ‘a more socially connected Web where everyone is able to add and edit the information space’ [Anderson (07)], and where sharing of resources through

an individual's social network is eased, is referred to as 'Web 2.0'¹ [O'Reilly (05)]. Instead of connecting resources Web 2.0 is focussing on connecting users with what is commonly denoted as social software.

This paper discusses the requirements for a framework to share digital resources and metadata for various contexts and applications in section 2, the ALOE system, currently developed at DFKI, is presented in section 3 and will serve as an example for the realization of such an approach.

2 Sharing Digital Resources and Metadata

In the following we will discuss the requirements for a system to share digital resources and information about these resources. We will consider the following key questions: (1) *Which types of resources will be handled by the system, and how can the resources be integrated?* (2) *How can resources be described properly to enable access and further functionalities?* and (3) *How can users and other systems and applications interact with the system?*

2.1 Resources

As already stated, the amount of resources that can be used to support users in their daily tasks is continuously growing. There are numerous different types of multimedia resources (e.g., HTML, PDF, MPEG), and they are stored in distributed locations and repositories. For all these resources, and also for resources created by users themselves, integration possibilities should be provided. Integration should be possible by inserting the resources directly into the system or by referencing them via a URI. In the latter case, the problem of URI persistence has to be considered.

2.2 Metadata - Describing Resources

There are different ways and standards to describe digital resources, e.g., Dublin Core² or LOM³. However, these approaches usually suffer from several problems (see [Doctorow (01)]) that can only be partly solved with technology. The main problem is that there is the implicit assumption in the structure of most metadata formats which suggests that there is a one-to-one relationship between a resource and the metadata that describes it [Downes (04)]. But there is no 'single and correct' way to describe a resource. A lot of the information depends on the context in which a resource was created, by whom it will be

¹ For a concrete definition see http://radar.oreilly.com/archives/2005/10/web_20_compact_definition.html

² <http://dublincore.org/documents/dces/>

³ <http://ltsc.ieee.org/wg12/files/LOM.1484.12.1.v1.Final.Draft.pdf>

used, and as a means to what end. Wiley et al. therefore distinguish between objective (e.g., the size of a file) and subjective (e.g., the degree of interactivity of a resource) metadata [Wiley, Recker, Gibbons (05)]. A one-to-one relationship also neglects that metadata may change during the lifecycle of a resource [Cardinaels, Duval, Olivie (06)].

It is obvious that centralised approaches, where only a single authority is responsible for the definition of metadata, are a bad idea. Instead, any attempt to describe resources should embrace diversity. The use of application profiles that ‘consist of data elements drawn from one or more namespace schemas combined together by implementors and optimised for a particular local application’ [Heery, Patel (00)] is one possible solution. Using less restrictive approaches (e.g., application profiles do not allow to introduce new data elements) such as ‘resource profiles’ instead of single metadata sets [Downes (04)] can also be considered. A resource profile is defined as a ‘a multi-faceted, wide ranging description of a resource’. It does not conform to a particular XML schema, instead, it is a patchwork of metadata formats (potentially created by different authors) which are assembled as needed in order to form a description that is most appropriate for the given resource.

In any case, the possibility to annotate various descriptions for each resource should be offered. Nevertheless, there should be some mandatory metadata. Firstly, to *enable basic functionalities* such as search and access (containing, e.g., the name and location of a resource). Secondly, about the *technical format of a resource and the technical requirements to use it*, and thirdly for *intellectual property rights* with information about the way in which a resource may be used.

2.3 Access by users

As already discussed in section 2.2, centralistic approaches have a lot of weaknesses. Thus, the aim must be to attract enough stakeholders that can provide valuable information about resources, at best working as a self-sustained community. But it is not enough just to attract a lot of users, there are certain conditions enabling the ‘Harnessing of collective intelligence’ as one of the Web 2.0 key principles [O’Reilly (05)]. A ‘wise crowd’ can be characterized by four conditions [Surowiecki (04)]: *Diversity of opinion* (each person should have some private information), *Independence* (people’s opinions are not determined by the opinions of those around them), *Decentralization* (people are able to specialize and draw on local knowledge), and *Aggregation* (some mechanism exists for turning private judgements into a collective decision).

If we want to ensure that these conditions are fulfilled, first of all we have to attract enough users that will contribute to the system. This of course requires dissemination efforts, but it is also very important to provide a user interface following the principles of simplicity [Nielsen (00)] and joy-of-use [Reeps (04)],

and to encourage users to participate, e.g., by using reward mechanisms. To ease the flow of information, possibilities to import (e.g., tags from a social bookmarking system) and export (i.e., users ‘own’ their information like their profile or their tags) of existing information should exist. Microformats can be used for the export of, e.g., contact details of users or reviews of resources. The aggregation of information can be realized in several ways: E.g., allowing access to all comments concerning a resource in an easy way, showing existing tags in form of lists or clouds, showing average ratings, etc. Last but not least, users should be offered the possibility to use functionalities in their usual contexts and applications, so that they can contribute with different views on resources. Using widgets is one way to realize such an integration.

2.4 Access by systems

Not only users, but also other systems and components can provide and use resources and information about them. So we need more than ‘just’ good user interfaces or widgets. An infrastructure allowing an easy creation of mash-ups and complex functionalities using the data provided from the system is required.

Access to the data as well as the functionalities of the system may be provided by making use of Web services. As an alternative to robust Web services which make use of ‘heavyweight’ techniques like SOAP⁴ and WS-*, often more lightweight or simplified programming models are demanded [Anderson (07)]. REST⁵ is a technology which is likely to meet these requirements.

Furthermore, the system should offer means of notification, e.g., about new resources concerning a certain topic, as well as the transmission of the new content itself. This can be realized by providing feeds using formats such as RSS or Atom.

3 The ALOE system

The ALOE system⁶ is currently being developed in the project CoMet⁷ at the Knowledge Management Department of DFKI. The aim of the project is the development of a system which offers possibilities to share digital resources and metadata about them. ALOE provides a rich user interface (see figure 1) to motivate users to participate, and allows data to be exchanged via a Web service API. Gathering information in this way is intended to achieve a richer description of resources, thus enabling the development of advanced retrieval and personalization techniques in the future.

⁴ <http://www.w3.org/TR/soap/>

⁵ <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>

⁶ <http://aloe-project.de/>

⁷ <http://www.dfki.uni-kl.de/comet/>

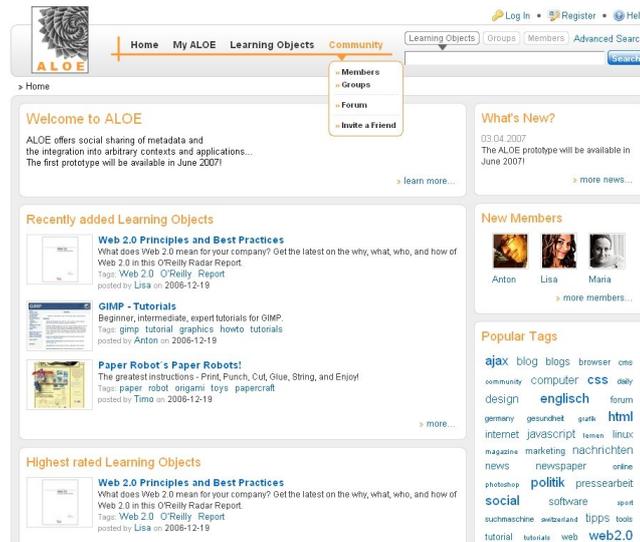


Figure 1: Screenshot of the ALOE user interface

In the following, the system architecture, the metadata which is used to describe resources, and ALOE's core functionalities are presented.

3.1 Architecture

ALOE consists of two components: The FLOR Web Service and the FLOR Connector. The **FLOR Web Service** is the Web interface to ALOE's **F**lexible **O**bject **R**epository (FLOR). It has been implemented using the Apache Axis⁸ technology thus providing access to the functionalities and the information stored in the system via services. The advantage of using this technology is that arbitrary clients can access the system by using SOAP. In ALOE, resources as well as user-defined metadata sets are stored in a Jakarta Slide content repository. Jakarta Slide offers full WebDAV support which enables direct access to the resources and user-defined metadata sets via their URI. Transactions and locking are provided so that data integrity can be ensured. The metadata which describes a resource and the metadata about a user-defined metadata set (see section 3.2) is stored in a MySQL database. That way metadata may be aggregated thus enabling different views on the resources registered in the system. The **FLOR Connector** realizes the user interface to ALOE's Flexible Object Repository. It has been implemented by using the JavaServer Faces⁹ (JSF)

⁸ <http://ws.apache.org/axis/>

⁹ <http://java.sun.com/javaee/javaserverfaces/>

technology. Compared to technologies like JavaServer Pages¹⁰ (JSP), JSF offers several advantages. For instance HTTP requests can automatically update the data of user interface components. Also fine grained event handling mechanisms are offered (e.g., for value change events in user interface components like list boxes) that could otherwise only be provided via client-side technologies. The use of JSF together with JSP causes issues concerning the rendering of user interface components ([Haiges, May (04)] pages 215-216). That's why the Facelets¹¹ technology was used in order to render the view. Facelets is an alternative view technology to JSP which has its strengths amongst others in templating and the composition of user interface components. As Facelets is based on XHTML and provides support for expression languages it can easily be integrated into our JSF application.

3.2 Metadata

To fulfill the requirements presented in section 2.2, ALOE stores a subset of the Dublin Core Metadata Element Set for every resource which is registered in the system (see table 1). It contains a minimal set of information which is necessary to work with resources of arbitrary formats (i.e., to make them retrievable, accessible and usable). In order to achieve a richer description, resources may additionally be described with metadata sets in arbitrary formats (e.g., LOM for learning objects or information about the context in which a resource was used). This allows for the creation of resource profiles which are likely to overcome the earlier stated problems of inappropriate metadata annotation.

3.3 Functionalities

ALOE provides the following basic functionalities:

- **Insert:** A resource or a user-defined metadata set can be inserted by uploading it as a file into ALOE's WebDAV repository, or by using a reference to the resource, i.e., its URI.
- **Annotate:** ALOE offers mechanisms to rate, tag and comment on resources. They can also manage own tags, favorite resources, etc.
- **Search & Navigate:** ALOE provides different search filters, e.g., users can search for resources which contain certain keywords in their title, description or tags. An advanced search is provided that allows to search for keywords in selected metadata terms. The information annotated by the users allows to browse content via tags (social browsing). Further, retrieved resources can

¹⁰ <http://java.sun.com/products/jsp/>

¹¹ <https://facelets.dev.java.net/>

Mandatory Resource Metadata	
dc:contributor	Person who inserted the resource into ALOE.
dc:creator	Author of the resource.
dc:date	Date of insertion.
dc:description	Description of the resource.
dc:format	Either MIME type or a proprietary format.
dc:identifier	URI which identifies the resource uniquely.
dc:rights	CC license which is associated with the resource.
dc:title	Title of the resource.
Metadata of a User-Defined Metadata Set	
dc:contributor	Person who inserted the metadata set into ALOE.
dc:creator	Author of the metadata set.
dc:date	Date of insertion.
dc:description	Description of the metadata set.
dc:format	Metadata format (e.g., DC)
dc:identifier	Identifier of the metadata set.
dc:relation	URI of the described resource.

Table 1: An excerpt of the metadata used in the ALOE system

be ranked according to different criteria, e.g., alphabetically, most viewed, best rated.

- **Preview:** Together with a preview image, various information about a selected resource (e.g., user-defined metadata files, comments, tags, ratings) may be displayed.
- **Group Management:** Users of ALOE can initiate and join groups of interest. Whenever they contribute a resource or a tag into the system, the users may decide whether the resource or tag is visible for all users of the system, for members of selected groups or only for themselves.

4 Summary and Future Work

Any approach that supports knowledge sharing and that meets the needs of today's users, should allow to incorporate any kind of digital resource available on the Web or a local repository, as well as content generated by the users themselves. To ease the use of the digital resources, information enabling basic functionalities, about technical requirements for the usage, and about the way in which a resource may be used (intellectual property rights) should always

be provided. Additionally, approaches such as application profiles or resource profiles are recommended to allow an adequate description of resources. A user interface that stimulates users to participate, options to import and export information in an easy way, methods to aggregate the collected information, and the possibility for users to use functionalities in their usual contexts and applications are very important to ensure that the harnessing of collective intelligence will be successful. Providing access to the system via services and making use of standardized formats for the representation of information ease the information exchange with other systems. The ALOE system realizes such an approach, and we plan to deploy and evaluate the first prototype at DFKI, where we focus on motivating researchers to share information about publications and conferences. ALOE will be enhanced with various functionalities in the future, e.g., feed support, semantic search and automatic generation of selected metadata.

References

- [Anderson (07)] Anderson, P.: “What is Web 2.0? Ideas, technologies and implications for education”; JISC Technology and Standards Watch, 2007.
- [Cardinaels, Duval, Olivié (06)] Cardinaels, K.; Duval, E.; Olivié, H.: “A formal model of learning object metadata”; in Wolfgang Nejdl and Klaus Tochtermann, editors, EC-TEL 2006 - Innovative Approaches for Learning and Knowledge Sharing, volume 4227 of Lecture Notes in Computer Science, pages 74-87. Springer, 2006.
- [Doctorow (01)] Doctorow, C.: Metacrap: “Putting the torch to seven straw-men of the metatopia”; 2001, electronic document, date of publication: August 26, 2001, retrieved March 14, 2007, from <http://www.well.com/~doctorow/metacrap.htm>.
- [Downes (04)] Downes, S.: “Resource profiles”; Journal of Interactive Media in Education, 5, 2004, ISSN:1365-893X.
- [Haiges, May (04)] Haiges, S; May, M.: “JavaServer Faces. Web Development mit dem Standardframework”; entwickler.press, 2006.
- [Heery, Patel (00)] Heery, R.; Patel, M.: “Application profiles: mixing and matching metadata schemas”; Ariadne, Issue 25, 2000, <http://www.ariadne.ac.uk/issue25/app-profiles/intro.html>.
- [Lyman, Varian (03)] Lyman, P.; Varian, H. R.: “How Much Information”; 2003, retrieved July 12, 2007, from <http://www.sims.berkeley.edu/how-much-info-2003>.
- [Nielsen (00)] Nielsen, J.: “Designing Web Usability: The Practice of Simplicity”; New Riders Publishing, 2000.
- [O’Reilly (05)] O’Reilly, T.: “What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software”; 2005, electronic document, date of publication: September 30, 2005, retrieved February 12, 2007, from <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>.
- [Reeps (04)] Reeps, I. E.: “Joy-of-Use: eine neue Qualität für interaktive Produkte”; Master’s thesis, University of Konstanz, 2004.
- [Surowiecki (04)] Surowiecki, J.: “The Wisdom of Crowds”. Doubleday, 2004.
- [Wiley, Recker, Gibbons (05)] Wiley, D. A.; Recker, M.; Gibbons, A. S.: “Getting axiomatic about learning objects”; in David A. Wiley, editor, The Instructional Use of Learning Objects: Online Version. 2000, retrieved from the Internet on February 1, 2006 from <http://reusability.org/axiomatic.pdf>.

Web 2.0 as Platform for Inclusive Universal Access in Cooperative Learning and Knowledge Sharing

Tomáš Pitner

(University of Vienna, Research Lab for Educational Technologies, Austria; Masaryk
University Brno, Czech Republic, tomp@fi.muni.cz)

Michael Derntl, Thorsten Hampel, Renate Motschnig-Pitrik

(University of Vienna, Austria
{michael.derntl, thorsten.hampel, renate.motschnig}@univie.ac.at)

Abstract: The goal of this paper is to identify and analyze Web 2.0 services and usage models appropriate for assisting the realization of Inclusive Universal Access (introduced by Derntl and Motschnig-Pitrik as an extension to Universal Access) particularly in cooperative learning and knowledge sharing environments. The analysis is backed by previous research and experience either at the authors' or other educational institutions. It can serve as a brief reference of a broad spectrum of Inclusive Universal Access scenarios which have proven to be successful together with their Web 2.0 support. It will hopefully also help Web 2.0 service developers to get deeper insight into the different aspects and effects of the new web services in terms of their applications in cooperative learning and knowledge sharing. Furthermore, the paper aims to reveal the gaps and future topics for research, development, and practice.

Keywords: Universal Access, Inclusive Universal Access, Person-centeredness, CSCW, scenario, Web 2.0, groupware, blog, wiki, teamwork, discussion, sharing, planning, active participation, project-based learning, blended evaluation, peer teaching, learning contracts

Categories: C.2.4., H.3.5., H.4.3., H.5.3., I.7.4, K.3.1

1 Web 2.0 and Inclusive Universal Access

When Tim O'Reilly coined the term Web 2.0 he created a new word for an old, well-known phenomenon. He was talking about new ways of interaction on the web. As multidimensional the term Web 2.0 may be today, as important is the discussion about "people and the web". Web 2.0, though very loosely defined and often criticized as hyped and overestimated, has gained enormously high popularity among developers, businesspeople, and – most importantly – among users, including knowledge workers.

In our conception of Web 2.0 we are mainly referring to the next generation of computer supported cooperative work (CSCW), which has traditionally been a highly interdisciplinary approach, involving computer scientists, media scientists, psychologists, and sociologists. What can make Web 2.0 define the next generation of CSCW is Universal Access (UA) [Stephanidis and Savidis 01], aiming to involve all users in all aspects of participation. For the first time, people can communicate and contribute online, express their views without any specialized skills.

The potential of Web 2.0 in the area of learning and knowledge sharing is broader than just being a set of tools supporting UA. From the classical UA viewpoint, services and systems are universally accessible when they are inclusive of all

potential users, adaptable to specific needs, usable even by non-specialist target users, and involving users in all phases of the system life cycle. We recognize that in many ways these aspects apply to the concept and intent of Web 2.0 as well. However, UA primarily focuses on technology and design aspects of accessibility. In recent research [Derntl and Motschnig-Pitrik 07] we extended and integrated the UA concept with non-technological, human aspects that can contribute to facilitating social and personal growth of students in learning and knowledge sharing settings. We refer to this extended UA concept as *Inclusive Universal Access (IA)*. It proposes a significant extension to the concept of UA in so far as *IA aims to actively involve learners in all aspects learning and assessment; to primarily address them on all levels of learning including intellect, skills, and personality; and to employ universally accessible tools to support the educational activities.*

The goal of this paper is to identify and analyze Web 2.0 services and usage models appropriate for assisting Inclusive Universal Access in Cooperative Learning and Knowledge Sharing. The analysis is backed by previous research and experience either at the authors' or other educational institutions. It can serve as a brief reference of a broad spectrum of IA scenarios which have proven to be successful together with their Web 2.0 support. It will hopefully also help Web 2.0 service developers to get deeper insight into the different aspects and influences of the new web services in terms of their applications in cooperative learning and knowledge sharing. And, generally, it can help to find the gaps and identify future topics for research, development, and practice.

2 Guiding Principles for IA Scenarios

The discussion of IA dimensions revealed a number of central, overlapping characteristics of IA scenarios, which can be summarized by the following guiding principles [Derntl and Motschnig-Pitrik 07].

- **Levels of learning:** Address students on all levels of learning, i.e. intellect, skills, and personality.
- **Responsibility, empowerment:** Offer students to take responsibility for the educational process by involving them in setting learning goals, in evaluation of their own and their peers' learning, and in assessment of course quality.
- **Active participation:** Inspire students to actively participate in educational activities including practical exercises, collaboration with peers, presentations of results and projects, proposing and solving problems of personal relevance, or evaluating their own and their peers' contributions.
- **Interpersonal sharing:** Engage students in interpersonal sharing and communication with their peers and the instructor.
- **Person-centeredness:** Promote and live the attitudes of acceptance, realness, and understanding to a perceptible degree and equally try to express these attitudes in the design and conduct of learning activities.
- **Universal access:** Use methods and tools for learning and teaching support (e.g. a web-based learning platform) in compliance with classic UA.

3 IA Scenarios Supported by Web 2.0 Services

In this section we summarize the characteristics of the proposed IA scenarios and present appropriate Web 2.0 support for each one. Thereby we concentrate on principles, while mentioning very concrete services as well.

3.1 Considering goals and expectations

Rationale. This scenario is employed early in the course (e.g., in the initial meeting) to involve learners in defining overall as well as individual goals for the course, such as “What do I want to achieve in the course?” It has been shown that furthering learner orientation by having the participants formulate questions and goals first is an effective strategy for improving their learning. This scenario particularly supports inclusion, user involvement, and adaptability.

Web 2.0 support. Many Web 2.0 services allow establishing *groups* (like *MySpace*, <http://myspace.com>), *blogs* (like *Blogger*, <http://www.blogger.com>), or host a free wiki (*Wikidot*, <http://www.wikidot.com>). An initial discussion about the overall course goals can be launched there. The respective blog entries or wiki pages can be appropriately *tagged*, to ensure distinctiveness among the whole content. Almost all content-creation services additionally offer *aggregation/syndication* feeds like Atom and RSS (“Really Simple Syndication”). With those feeds the individual contributions can be integrated, presented, and reused.

3.2 Teamwork

Rationale. Teamwork is used for complex problems that require application and hands-on experience with relevant subject matter. Thereby we offer students freedom of choice of their team mates. To increase transparency and effectiveness of teamwork, we additionally provide each team with virtual *communication facilities*, *online workspaces*, and a web-based *journal* to keep track of teamwork activities. As teamwork requires coordination, communication and cooperation it is capable of furthering interpersonal skills and a constructive personality development.

Web 2.0 support. Regarding *communication facilities*, some of the team discussion has private or closed character not suitable for a common discussion. In such a case, an e-mail or instant messaging (IM, chat) solution may be preferred, e.g. Google services offer a package of integrated communication tools, such as e-mail client and instant messenger. Similar functionality may be found integrated in many other Web 2.0 services. Regarding *online workspaces*, a lot of groupware-like services can recently be found on Web 2.0. *Online Office Zoho* (<http://www.zoho.com>) or *Google Docs & Spreadsheets* (<http://docs.google.com>), for instance, provide such a facility for online creation, import, editing, publishing, and sharing of the most common office document types (i.e., texts and spreadsheets). If these common formats are not sufficient, generic online storage and sharing services like *MediaMax* are available (<http://www.mediamax.com>). Note, however, that privacy may be an issue, as the basic/free version of the service often does not allow definition of data access restrictions (e.g., public, private, group). The collaboration support is also the core business of specialized services like *Backpack* (<http://www.backpackit.com>) or *Teamwork* (<http://www.twproject.com>). If the actual teamwork includes intensive

online interaction, virtual environments like *LiveOnline Rooms* (see later) represent an appropriate approach.

3.3 Project-based learning

Rationale. In Project-based learning (PBL) students tackle complex problems in an iterative and incremental approach. Projects and case studies additionally are suitable application domains for the teamwork scenario. In IA courses students propose concrete project topics within the course's topic frame. PBL offers process flexibility and supports user involvement, adaptability, and inclusion. It is usually combined with blended evaluation of projects (see later).

Web 2.0 support. In this scenario, the project team will probably use many of the tools presented. The abovementioned full-fledged project management Web 2.0 services provide comprehensive team project support online, whereby the free version has seriously limited functionality. However, to define just project milestones, plan meetings, presentation sessions or events, there is a handful of simpler Web 2.0 services oriented to manage calendars and promote events, like *CalendarHub* (<http://www.calendarhub.com>), *Google Calendar* (<http://www.google.com/calendar>) – both providing sharable calendars – or *Eventful* (<http://eventful.com>) for event promotion, respectively. For periodic project reporting, blogging or docs & spreadsheets services represent appropriate tools allowing posting the reports and sharing them within a specified audience.

3.4 Sharing and presentation of contributions

Rationale. In conventional settings it is common that students have no idea about the tasks their peers are working on. For instance, in a typical seminar initially topics for papers are distributed, then elaboration of papers proceeds without much student-to-student interaction, and finally papers are presented during lengthy presentation blocks. In IA this is overcome by *fostering sharing* and *discussion of contributions* and by providing space for *periodical presentations of work-in-progress* to give students the opportunity to learn and benefit from each other.

Web 2.0 support. If simple collaborative authoring tools like blogs and wikis are considered to be too informal or insufficient for sharing ideas, specialized Web 2.0 tools, namely *Zoho Show* for creating presentations or *SlideShare* (<http://www.slideshare.net>) for publishing and sharing can be utilized. To persistently capture live presentations it is often sufficient to record just computer desktop content (as video track) and accompanying comments (as audio track) creating a screencast. There exist many tools to do this job, including but not limited to, *Camtasia* (commercial) and *V3 Recording Tool*, a freeware tool based on *CamStudio* and running on nearly all flavors of the Windows operating system [Pitner et al. 06]. The screencast media file size is distinctively smaller than camera video while preserving the recording quality and user experience, and requiring only a cheap recording hardware like a handsfree set or a simple webcam.

3.5 Peer teaching

Rationale. The aim of this scenario is that students become more involved in the teaching process. This can for instance be realized by including tutorials on specific

topics and tools. However it can be even deeper embedded into the instruction, for example when *students or teams prepare presentations and moderated course units* under supervision of the instructor. On the personal level research on tutorials has shown that technical skills are increased and that peer teaching recipients tend to show “greater confidence, more motivation to work, and an improved attitude” toward the subject.

Web 2.0 support. At Masaryk University in Brno, Faculty of Informatics, peer teaching is conducted in several courses. However, the most successful application of peer teaching in the software development area has been achieved so far in the so-called *Tech Preview Parties* (<http://iris.fi.muni.cz/techpreviewparty>) organized regularly and covering various topics, mostly on advanced software development. From the technological point of view, a combination of the university’s learning management system and a wiki based on a local instance of *MediaWiki* (<http://www.mediawiki.org>) has been employed. A necessary step in peer teaching is collecting, ordering and integration of the peers’ contributions. Fortunately, recent Web 2.0 services mentioned above provide syndication feeds (RSS/Atom). If feeds are unavailable, clipping services like *Clipmarks* (<http://www.clipmarks.com>) might help here to collect the contributions and allows posting them to a blog.

3.6 Interactive elements

Rationale. Various forms of interactivity among students, instructor, tutors, and experts can be employed to further student participation as well as interaction and communication skills. Examples include *brainstorming, collective elaboration of theories, online market spaces for ideas and contributions, topic workshops, (online) consultation hours* at regular intervals, etc.

Web 2.0 support. A clever combination of common Web 2.0 tools, such as IM (chat) and discussion services can provide a sufficient fundament for nearly all tasks occurring in brainstorming and collective elaboration sessions. Freely available mind-mapping software, such as *FreeMind* (<http://freemind.sourceforge.net>), or better online brainstorming environments like *bubbl.us* (<http://bubbl.us>) allowing editing and sharing, serve as good examples here. For even richer interaction, computer-mediated communication environments like *LiveOnline Rooms* may be employed (see later). On the other hand, when just a simple collaboration taking the form of document reviewing, commenting or just collecting is required, Web 2.0 annotation tools like *Diigo* (see later) can be applied.

3.7 Computer-mediated communication

Rationale. To further communication with peers, tutors, and instructors students are provided with virtual communication facilities such as online *discussion forums* and *chat rooms*. These facilities can be used for synchronous and asynchronous communication in online phases for coordination, cooperation, and elaboration.

Web 2.0 support. The advantage of Web 2.0-based synchronous and asynchronous communication over closed, single-purpose environments is obvious. However, in the near future, when broadband connection will reach a critical mass among the students, we may expect the rise of a new generation of virtual presentation and collaboration tools, such as *LiveOnline Rooms* [Gorjany 06]. In comparison with traditional video-

conferencing, the setup does not require more than installing an auto-updated lightweight client (browser plug-in) and connecting a web camera. The functionality is, on the other hand, richer, allowing uploading and showing PowerPoint slides, and on-screen hand drawing.

3.8 Problem proposals

Rationale. Students are offered to *propose problems* or *topics* to be *elaborated*, *presented*, or *considered* in the course. This scenario can for example be realized in PBL or learning contract settings (see below).

Web 2.0 support. When proposing problems or topics, a mixture of (collaborative) content creation and annotation tools might be handy to annotate what the teacher or the peers proposed. For example, *Diigo* (<http://www.diigo.com>) offers a browser toolbar to easily bookmark, highlight and/or annotate an arbitrary web page and share the annotations publicly, within a group, or keeping them private.

3.9 Learning contracts

Rationale. In this scenario the *instructor sets basic task requirements* and a *topic frame* and within which students or teams are given a significant degree of freedom to *propose specific topics* they wish to work on. The instructor approves the proposals, facilitates the elaboration process, and co-evaluates the results according to goals defined in the contract.

Web 2.0 support. The tools presented earlier for the scenarios “considering goals and expectations” and “problem proposals” can also be applied here because learning contracts have similar characteristics, but on a very personal level. Additionally, elaboration of learning contracts usually follows a step-wise, incremental procedure, which should make Web 2.0 support for this scenario compliant with any tools employed for the “project-based learning” scenario.

3.10 Collecting feedback and opinions

Rationale. Students are involved in the course quality assessment process by collecting feedback and opinions in various forms, such as structured feedback in *questionnaires* or *open feedback* in class, in online *reaction sheets*, *surveys*, and *feedback forums*. The feedback may concern specific activities or the whole course. It impacts upcoming course units and is used to improve the scenarios.

Web 2.0 support. Blogs with their commenting facilities represent a great tool to quickly collect informal and unstructured feedback – you simply post an entry and let the student comment it. Nevertheless, if a more structured feedback in a form of a questionnaire is preferred, Web 2.0 offers some ready-to-use services, such as *Free Online Surveys* (<http://freeonlinesurveys.com>).

3.11 Blended evaluation

Rationale. This scenario aims to involve students in the evaluation of their own and their peers’ contributions. All student evaluations are reviewed by the instructor as a quality assurance measure: the final decision on grades lies with the instructor.

Web 2.0 support. Any peer contribution (project, report, task solution) can be peer-evaluated either with generic tools enabling qualitative/narrative evaluation, such as discussion forums, blogs and wiki, with online questionnaire services mentioned above, or simply by “digging or burying” using *Digg* (<http://digg.com>) or similar web resource evaluation system – if privacy is not an issue, because the ratings (“digs”) are publicly visible. In a more structured approach, questionnaire and feedback tools can be used as well for this scenario.

4 Conclusions

We have shown that there are multiple ways of employing Web 2.0 services and tools for meaningful cooperative learning scenarios based on the principles of Inclusive Universal Access. Pedagogical advantages of three specific scenarios have also been demonstrated in [Motschnig-Pitrik and Derntl 07]. Specifically, Web 2.0 can bring significant benefits for learning and knowledge sharing communities when educators are aware of the following potential advantages and drawbacks.

By using Web 2.0 tools and services learning and knowledge sharing communities of any size can benefit from network effects, as there are usually large existing communities around popular Web 2.0 services (e.g. Google Groups). These effects are difficult to be achieved by any closed LMS [Wilson 06]. Also, if community members are already active users of the service(s) there are no entrance barriers for these users [Alexander 06]; the border between learning and the real, personal world of users virtually disappears, which can produce a feeling of comfort and security. As the number and diversity of Web 2.0 services is steadily increasing, there are almost certainly services available for all kinds of informal, ad-hoc or self-directed teaching and learning episodes (e.g., seminars, workshops). In content creation activities (e.g., seminar report or creating presentation slides), the growing popularity and content body of open/shared content initiatives (e.g. Creative Commons) can greatly simplify dealing with copyright issues [Kelly 07]. However, as English is the primary language on the web, many services are only useful when use of a foreign language is not an issue for the non English speaking users [Pitner 07].

The drawback of using and aggregating various external Web 2.0 services in a learning community is that it is usually required to sign up for each service separately, so community members will have to memorize or keep notes of various user credentials. This is remedied only by single sign-on services, e.g. those offered by Google. Even though syndication and aggregation feeds such as RSS allow users keeping track of any “action” in the community it may still be a burden to rely on the availability and stability of services of multiple providers [Kelly 07]. Additionally, similar services offered by different providers may differ significantly in usability, navigation scheme, and supported features. This might be less of a problem for experienced web users or computer science students, but has to be taken into account when employing Web 2.0 services for inexperienced communities of web or computer users. Also, from the technological point of view the resource requirements (hardware, connection) on the client-side need to be satisfied by all potential users [Kelly et al. 07].

Another aspect to be considered by educators is that use of new Web 2.0 services per se does not necessarily add value to the learning and community processes. Added

value is achieved only with a sound underlying pedagogical approach [Delgado 07]. Finally, the advantages of new Web 2.0 services can unfold only when educators, curricula, learners, and other community members are sufficiently open and flexible to tolerate different viewpoints and integrate them into their teaching and learning.

Acknowledgements

The research has been partially supported by Czech National Program Information Society, project “E-learning in the Semantic Web Context”, No. 1ET208050401; by the EC Leonardo da Vinci Programme “Versatile Vocational and educational training Vehicle (V3)” project; and by the “Technology Enhanced Learning” project at the University of Vienna. The authors sincerely thank their colleagues Jürgen Mangler and Klaus Spiegl for their continuing endeavor to make web services available.

References

- [Alexander 06] Alexander, B.: “Web 2.0 - A New Wave of Innovation for Teaching and Learning”; in EDUCAUSE Review, March/April, 2006.
- [Delgado 07] Delgado, D.: “Live mini-paper on informal learning, connectivism, e-learning 2.0, Elgg and Moodle”; <http://eduspaces.net/davidds/weblog/159359.html>, accessed May 15, 2007.
- [Derntl and Motschnig-Pitrik 07] Derntl, M. and Motschnig-Pitrik, R.: “Inclusive Universal Access in Engineering Education”; Proc. 37th ASEE/IEEE Frontiers in Education Conference, Milwaukee, WI, 2007.
- [Gorjany 06] Gorjany, M.: “Web2.0 - Community zur gemeinsamen Nutzung neuer LiveOnline”; Nov 27, 2006, <http://lionpool.web2.0campus.net/>, accessed May 16, 2007.
- [Kelly 07] Kelly, B.: “Deployment Strategies For Web 2.0”; <http://www.ukoln.ac.uk/web-focus/events/seminars/nottingham-2007-03>, accessed May 15, 2007.
- [Kelly et al. 07] Kelly, B., Sloan, D., Brown, S., Seale, J., Petrie, H., Lauke, P., and Ball, S.: “Accessibility 2.0: People, Policies and Processes”; Proc. International Cross-Disciplinary Workshop on Web Accessibility, Banff, Canada, 2007.
- [Motschnig-Pitrik and Derntl 07] Motschnig-Pitrik, R. and Derntl, M.: “Three Scenarios on Enhancing Learning by Providing Universal Access”; Universal Access in the Information Society, Springer Verlag, (2007, in press).
- [Pitner 07] T. Pitner: “Web 2.0 in Education - A Survey”, May 15, 2007, <http://tomaspitner.blogspot.com/2007/05/web-20-in-education-survey.html>.
- [Pitner et al. 06] Pitner, T., Ráček, J., Schumacher, L. and Miche, G.: “A Framework to Enable Enriched Training Experience”; Proc. ICL Conference, Villach, Austria, 2006
- [Stephanidis and Savidis 01] Stephanidis, C. and Savidis, A.: “Universal Access in the Information Society: Methods, Tools, and Interaction Technologies”; Universal Access in the Information Society, 1, 1 (2001), 40-55.
- [Wilson 06] Wilson, S.: “Web 2.0 a learning and teaching viewpoint”; Presentation at the Institutional Web Management Workshop 2006, <http://www.slideshare.net/lisbk/web-20-a-learning-and-teaching-viewpoint>, accessed May 16, 2007.

Knowledge Management in the Web 2.0 – Semantic Positioning as an Integrated Approach

Patrick Erren

(Heinz-Nixdorf-Institute
University of Paderborn, Germany
erren@mail.uni-paderborn.de)

Abstract: Distributed cooperative contexts provide a challenge for successful Knowledge Management. Users in different places around the globe need to interact with one another in goal-oriented projects. In classical Computer Supported Cooperative Work applications this interactive component is often neglected. In contrast Web 2.0 offers a strong author-recipient interaction that may be translated successfully into cooperative work contexts. An integrated approach based on virtual knowledge spaces using semantic positioning to communicate meaning efficiently between project members enhanced with Web 2.0 functionality is introduced in this paper.

Keywords: Cooperation, Knowledge Management, Web 2.0, Virtual Knowledge Space, Semantic Positioning, Medi@rena,

Category: K.3

1 Interaction in Cooperative Knowledge Management Software

Knowledge has always been an important factor in leading. It has played a deciding role for success, beginning with monarchs and tyrants of old up to politicians and industry leaders of today. Yet only recently have we come to embrace the term of knowledge as a good that can be managed; we are living in a knowledge society. Strangely however we still speak of the information age. The qualitative difference of the two terms information and knowledge has been discussed over long terms. The general consensus seems to be that knowledge is superior to information by providing an interlinked context. This paper therefore understands knowledge as information applied by humans in differing ways depending on the given context. This tie to humans explains why knowledge transfer in cooperative work is still a challenge.

Since a lot of today's knowledge work is accomplished with computers and more and more project teams are distributed globally, the research area of Computer Supported Cooperative Work (CSCW) came into being. The term CSCW itself was coined by Irene Greif and Paul M. Cashman in 1984 during a workshop [see Greif 1988]. It describes how cooperative work can be supported by computer technology, with a focus on situations of distributed teams. The specific tasks of CSCW in addition to the cooperation itself can be described as coordination and communication according to [Teufel et al. 1995]. It can be argued that without guaranteeing at least a constant communication between team members cooperation and coordination efforts aren't possible, since both require consultation. Thus supporting communication especially in distributed settings is a basic requirement in the area of CSCW.

In the area of communication growing bandwidth has enabled the development of technologies that allow participants to communicate more and more naturally with each other. The development ranges from telephone and simple asynchronous e-mail over synchronous text chat up to video conferencing.

The area of coordination needs to manage knowledge flows staying true to the set project goals and assigning (knowledge) resources according to needs of current tasks. Therefore tools for coordination include shared group calendars and other scheduling programs that help assigning human resources to specific tasks and to set deadlines for milestones. While the tools on this level have to be considered mature, the area of managing document-based knowledge sources on the access and designation level is still much less elaborate. Team members need a common knowledge pool where knowledge or at least information can be placed in the form of documents. Furthermore these documents should not only be presented conservatively in file manager lists, but in respect to their contextual meaning, so to support knowledge transfers. [Chapter 3] will detail how Semantic Positioning may help coordinators and team members to store their knowledge in readily accessible ways, so that the respective context isn't lost.

The cooperation of individual teams is accomplished mostly on the base of documents (project code and documentation, media files...) as is common for almost any knowledge work. It is important to have a rights and role management in place that helps to explicitly state who is allowed to access documents in which way.

Knowledge documents are constantly in use by a multitude of team members and thus their contents are often related as well. Such a relation may span more than two elements or just refer to a specific meaning of the document itself. The visualization of these relations or the context should be directly evident to users without the need of determining it from reading the contents [Bannon & Bødker 1997]. Documents within a common space should have a persistent (but also if possible dynamic) meaning that can be perceived directly and instantly by looking at the workspace. This helps team members assessing the relevance of elements to their task and enables them to construct relations to their individual knowledge. There aren't however many applications that cater for an immediately visible way of leaving comments or annotations on documents or even folders or allowing for spatial arrangements that reflect paper based working environments (e.g. a person's desk).

It can be argued that user interactions are limited in CSCW software. It is often impossible to comment documents or to deduce their meaning and relational context instantly without the need of reading them. Web 2.0 has opened many new perspectives in the area of user integration into the generation of content. Many static author reader cycles of the first generation internet are broken down and replaced by cooperative functions. The following chapter will give an introduction into Web 2.0 technologies and identify its major advantages as well as its shortcomings.

2 Better Interactivity for Collaboration with Web 2.0

Web 2.0 has become a common buzzword that is often uttered in reference to interactivity and user created content. This stems from the central advances over the classical internet. First of all Web 2.0 can be accessed just like its predecessor by any regular web browser, thus basically eliminating the barrier of entry for members of

our knowledge society. Most of the presented contents have been created and uploaded by users and not the owners of the same site. This insight discloses the main new value of Web 2.0: Every user of a web page becomes a potential author. The classic web with its long interaction cycles between authors and recipients, where feedback about errors could only be given via e-mail or guest books, seems to be outdated [see Gillmor 2004].

A special aspect is the sheer scale of thousands if not millions of users visiting and contributing to pages. While not every user exerts their right to comment or rate posts, there are individual videos on *YouTube.com* with over a thousand comments. It is however highly questionable in how far these comments are still relevant to the author or other readers. The described general level of interactivity in Web 2.0 is even surpassed by sites like *Wikipedia.org*. Wikipedia calls itself a user created online encyclopedia. The basic premise of Wikis is *cooperative writing* and can be regarded as a new level of interaction, because each article has a possible set of infinite authors.

Web 2.0 also bears a few general problems especially in the light of using the new technologies for cooperative purposes. Companies as well as learning institutions foster goal-oriented working. Within Web 2.0 there is no ultimate purpose. Most applications just apply to the sector of private entertainment. Users only upload self-created content to possibly receive a large feedback. This isn't however very useful in a business sense, especially with CSCW technologies in mind.

Many Web 2.0 pages do not even feature the simplest form of user awareness. E.g. often it is impossible to tell if a specific user (as an author or commentator of a specific item of interest) is currently online or what he/she is currently doing. Also when a new comment is added, it will only appear to a reader of the same page after reloading it. The area of synchronous communication in a team is thus still restricted to external tools, such as instant messengers.

Another limitation is the one-dimensional display of a Web 2.0 page's objects in a simple list (sorted by rating, time of creation etc.). The context and the meaning of objects are not immediately available. Web 2.0 does however use a simple way of providing at least some context for posts. This is achieved by so called *tagging*. Tagging basically means to assign keywords to any kind of object, which are then used for filtering or search purposes. In most applications the tags are just created by the author, but for example Flickr.com (a page where user's can post their photo albums online) allows users to add their own tags to any photo they are allowed to view. This is a form of cooperative behavior, though again not a directed one; the data of hundreds of users is merely accumulated and then evaluated, they do not consciously work together. Also tags in their current form can only contain information about a specific object but not its relation to other objects.

The last fundamental problem of Web 2.0 with regards to cooperation is that specific sites specialize on just one multimedia sector (e.g. YouTube on videos, Flickr on photo albums etc.). These applications stand side by side and do not allow for constructs consisting of different media types or formats. For knowledge work however it would be important to mix documents of varying formats in one collection and use a general semantic concept to order these files. An integration of all the major Web 2.0 application services within one general application has not yet surfaced. Even current mash-ups only combine two or more services instead of focusing on formats and functionality.

3 Virtual Knowledge Spaces and Semantic Positioning

Documents in operating systems are usually represented as *objects with attributes*. One of the advantages of such an object model is that objects can be upgraded with new attributes by specific applications while not losing their original functionality. *Virtual knowledge spaces* are defined as object-oriented spatial interlinked structures that can be explored via the internet in cooperation [Geißler et al. 2004]. They feature persistent objects through a database and enhance these with role- and rule-management as well as extended communication functions (awareness of users in the same space, text-chat...). Knowledge spaces act as a kind of external memory for working groups due to the combination of cooperation and editing functions on persistent objects. The goal is to allow for working with visual arrangements of objects. This is based on the mentioned desk surface as the central place of knowledge work. The size, position and look of any object can be saved in its attributes, available because of the important object-orientation.

Semantic positioning describes the process of arranging objects graphically with the requirement that objects already gain meaning solely by their position in the construct [Erren & Keil, 2006]. Since the position is only another attribute of an object, overlays become possible in contrast to regular desktops (which usually employ a background grid to arrange icons making it impossible to overlap icons). This marks an important step in indicating relations between objects graphically. On the one hand it allows for stacks like on a real desk on the other hand it allows for arranging objects on top of a (graphical) background. This could for example be a color gradient between two extreme positions. Objects placed on top of this gradient receive an immediately discernable meaning regarding their contents, in the light of their position on the scale between the two extreme concepts. We call these kind of constructs arranged via semantic positioning *knowledge structures*.

We distinguish four variants of arranging documents with semantic positioning. These arrangements each employ a different visual method of showing relationships between objects:

Topologies use dimensions (usually in the form of axes) to assign concrete positions to values. Semantic closeness of documents regarding to a specific dimension is thus displayed through spatial proximity on the respective axis. *Logic sets* use (visual) containedness to display membership of documents. Those contained in the same sets (sub-sets, intersections...) are semantically closer, according to specified membership criteria, defined and displayed in the sets. *Relational graphs* use combinations of nodes and (named) edges to explicitly state different kinds of relations between documents. Note that a document can either be placed on edges or directly represent a node. Semantic closeness in relational graphs is usually displayed either explicitly through edges or lies in the nature of the graph (for example parent-child trees). Lastly there are *combinatorial arrangements*. These usually take the form of matrixes where each cell is defined as a combination of two vectors with respective meaning. Documents placed in these cells are semantically close when they share the same vectors equaling the same combinations and specific cell.

While each arrangement type serves its own function of visually assigning quickly comprehensible meaning to documents in relation to one another, their combination within knowledge structures offers even deeper ways of expressing

complex statements. We call these types of knowledge structures that use more than one type of arrangement *overlays*. This stems from the fact that the elements that make up the arrangement (lines, arrows, shapes etc. which we termed PicMents) have to be visually overlaid to achieve the desired result. Currently the depicted relations are not stored or evaluated by the processor, since the beneficial results were primarily intended for human perception.

In the following chapter a concept for an application integrating the multiple benefits of Web 2.0 combined with the advantages of CSCW software based on virtual knowledge spaces is introduced. This application uses semantic positioning as its central instrument to provide groups of objects in a virtual knowledge space with a visual context conveying inter-object relations and meaning.

4 An integrated concept: Medi@rena

A Medi@rena is defined as a co-active (cooperative, coordinating, communicative) virtual knowledge space enabling semantic positioning with openness for evaluation capabilities (this will most likely become a requirement). Semantic positioning requires a two dimensional cooperative whiteboard that allows for persistent arrangement and editing (creating, commenting, changing, deleting...) of documents, folders or spaces. A virtual knowledge space can contain files or objects, gates to other knowledge spaces or even sub-spaces and folders, thus allowing to cope with larger amounts of files. In addition graphical elements (Drawings, Images...) can be used to define relations in a visible way which enables recipients to immediately perceive a document's contextual meaning. Overlays are possible due to layers, in which objects and graphical elements are placed. Every object can be annotated and commented upon. Most importantly all of the named features are accessible synchronously for multiple users allowing for cooperation due to a strong rights and role management in the background. The goal is however not to create a new form of Web 2.0 application, where AJAX (e.g. <http://ajax.asp.net>) would be the preferred development framework, but an integration of the Web 2.0's advances into a CSCW environment, based on virtual knowledge spaces. Through this marriage of virtual knowledge spaces covering the advantages of CSCW applications with Web 2.0 evaluation and cooperative writing capabilities, the concept of a Medi@rena is a new step of integrating user interactivity into knowledge management software. Thus a Medi@rena becomes a dynamic representation of the working process and context.

In a first development step at the University of Paderborn a rich-client application based on the Eclipse (www.eclipse.org) platform has been implemented that is extendable through plugins. In communication with a server architecture based on virtual knowledge spaces, this *Medi@rena Composer* already allows for the creation of simple document arrangements that can be enhanced with simple drawings and images. Communication so far has only been implemented in the form of a text-chat (and server-wise email functionality), but the commenting or annotation of documents is in place. Changes to a specific arrangement by someone are immediately reflected to all other users present in the respective virtual knowledge space. Cooperative evaluation, rating, writing and refinement functions from Web 2.0 are also present in early forms [see Fig. 1], demonstrating how regular cooperative contexts can be

enhanced with respect to interactivity, which the author will describe in the following by two examples:

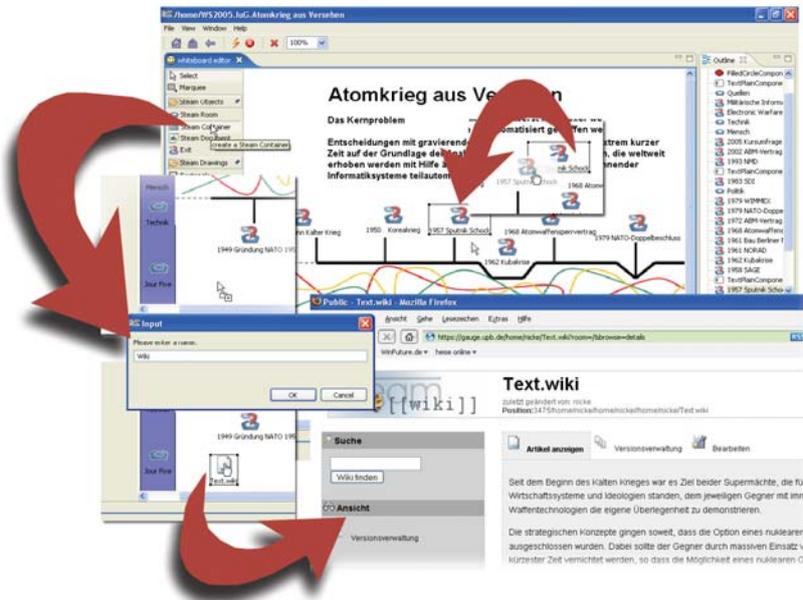


Figure 1: Interaction in Medi@rena to create a knowledge space and a wiki in the visual knowledge structure.

The first example was realized for an electronic review process of milestones during projects. The goal was to cooperatively review products in the form of seminar papers over the course of a semester. Each student's contributions were evaluated by all others at specific milestones with respect to their state of readiness for publication and/or entering the next project phase. A simple traffic lights system was utilized for the rating with green representing a ready, well received product, yellow showing that there are still some problems with the contribution (it may still pass into the next phase) and finally red indicating some major problems that led to a denial of passing the document. The individual votes were kept confidential and at the end an aggregated result for each contribution was used to decide about its readiness to pass into the next editing phase.

Discourse structuring is a second example for the step-by-step refinement of positions or theses. The actual implementation focuses on so called pyramid discussions [Blanck 2005], where the graphical structure of a pyramid is used for the refinement process. A set of opinions are placed at the pyramid's base and then in a pair-wise debate refined into new theses for the next level. For example eight original positions would result in four new positions on the next level, then two and then a final thesis comprised of parts of all the original positions. Each thesis can be read and commented in cooperation and the resulting new theses are written cooperatively. There is a certain set of rules governing the process for instance one can only read

other positions on a specific level, after an own (or jointly formulated) thesis has first posted on the same level. Thus the pyramid is a graphical representation of the whole process structure with the step-wise combination or refinement of contents.

These examples show that shared rating and commenting structures like they are present in Web 2.0 can be used to enhance cooperative working contexts in the integrated form of a Medi@rena.

5 Related Work

Other projects exist which are based on different uses of knowledge structures; two of these shall be mentioned and compared to this paper:

iMapping [Haller 2006] is a technique for creating semi-structured visual representations for the modeling of knowledge. iMap in this context are comparable to a large white- or bill-board where information items can be positioned similar to post-its but also nested into each other. The resulting structures are intended to be persistent, however in difference to the approach described in this paper, the focus of iMapping seems to lie more in opening up the navigation through knowledge structures via semantic zooming than to define types of spatial arrangement and the associated expressible meaning.

Habanero [Driggers et al. 1997] is a framework for the construction of distributed collaborative applications based on Java. It offers tools for text messaging, a shared whiteboard and a Viewer for different file formats. Custom extensions can be added in the form of so called "Hablets". The whiteboard in this implementation also serves the purpose of graphical representations, however more in the form of short and simple sketches than complex and persistent arrangements or knowledge structures.

6 Résumé and Outlook

The paper introduced the Medi@rena framework as a base for Knowledge Management catering to the three c's and adding a graphical communication layer to provide contextual meaning combined with interactive commenting and rating in goal-oriented scenarios. This provides a remedy for the missing goal orientation and integration of functions in Web 2.0 and thus enables to take advantage of the new developments in author-recipient-interactivity in CSCW applications. The added dimension of enabling the cooperative arrangement of graphical knowledge structures of any kind of documents with graphical elements through semantic positioning helps to devise a more efficient communication within the work area of distributed teams.

The most recent development of our virtual knowledge spaces has added further Web 2.0 functionality like Wikis and RSS-feeds which still need graphical representations in the Medi@rena Composer. Further research on how semantic positioning can also be used to store relations, opening up the system for responsive scenarios is currently being conducted. One currently researched possibility in consistence with the strong object-oriented concept of virtual knowledge spaces is that of translating the contextual meaning expressed through graphical arrangements into attributes of an object. The results would be special forms of tags, which could be assigned by graphically dragging an object to a specific position. The assigned

meaning could then be used for sorting, filtering and searching. This may ultimately lead to responsive structures where for example a document can tell the user that it may have been placed in violation of attributes defined in other arrangements

References

[Bannon & Bødker 1997] Bannon, L., Bødker, S. (1997): *Constructing Common Information Spaces*, Proc. ECSCW 97, Lancaster, Sept. 1997, 81-96.

[Blanck 2005] Blanck, B.: Diskutieren mit der Methode der »erwägungsorientierten Pyramidendiskussion« – ein Beispiel für computerunterstütztes erwägendes Lernen. In: Berntzen, D., Gehl, M., Hempel, M. (Hrsg.): *Zukunftswerkstatt Lehrerbildung: Neues Lehren und Lernen durch E-Learning. Der didaktische Mehrwert von E-Learning-Konzepten in der Lehrerbildung*, Tagungsdokumentation. Münster, 7. Juli 2005. Münster, p. 70-98.

[Driggers et al. 1997] Driggers, B., Alameda, J., and Bishop, K.A.: Distributed Collaboration for Engineering and Scientific Applications Implemented in Habanero, A Java-based Environment. In: *Concurrency: Practice and Experience, The Association for Computing Machinery*, New York, NY, USA, (1997), Volume 9, Number 11, p. 1269-1277.

[Erren & Keil 2006] Erren, P., Keil, R.: Semantic Positioning as a Means for Visual Knowledge Structuring. In: Nejdll, W., Tochtermann, K. (eds): *EC-TEL 2006*, LNCS, Volume 4227, p. 591-596, Berlin, Heidelberg, Springer-Verlag.

[Geißler et al. 2004] Geißler, S., Hampel, T., Keil-Slawik, R.: Vom virtuellen Wissensraum zur Lernumgebung – Kooperatives Lernen als integrativer Ansatz für eine mediengestützte Bildung. In: *i-com, 3.Jahrg., Heft 2*, (2004), p.5-12.

[Gillmor 2004] Gillmor, D.: *We the Media: Grassroots Journalism by the People, for the People*. O'Reilly Media, (2004).

[Greif 1988] Greif, I.: Remarks in panel discussion on "CSCW: What does it mean?", *CSCW '88. Proceedings of the Conference on Computer-Supported Cooperative Work, September 26-28, 1988, Portland, Oregon*, ACM, New York, NY, (1988).

[Haller 2006] Haller, H.: iMapping – a Graphical Approach to Semi-Structured Knowledge Modelling. In: Lloyd Rutledge (eds), *Proc. SWUI2006*, Athens, GA, USA. Nov, (2006).

[Teufel et al. 1995] Teufel, S., Sauter, C., Mühlherr, T., Bauknecht, K.: *Computerunterstützung für die Gruppenarbeit*. Addison-Wesley, Bonn, (1995).

Design of Personalized Knowledge Management in Web 2.0 Network

Malinka Ivanova

(Technical University – Sofia, Bulgaria)

m_ivanova@tu-sofia.bg

Ekaterina Vasilyeva

(University of Jyvaskyla, Finland)

ekvasily@cc.jyu.fi

Inga Beleviciute

(Vilnius Gediminas Technical University, Lithuania)

inga@gama.vtu.lt

Abstract: The knowledge is defined as combination and organization of data and information in given context and Knowledge Management (KM) provides capturing, storing and reusing of knowledge objects. In Web 2.0 world the knowledge is represented in form of microcontent object and KM 2.0 proposes creation, sharing and leveraging the microknowledge in a collaborative way. The microknowledge in Web 2.0 network can be controlled through designing the instructional strategies that will provide user learning paths and activities and that will give possibilities for sharing of those same learning activities and microknowledge with others. The present paper is aimed to reflect of the research needs and the new challenges in the mentioned above three areas: KM, Web 2.0 technologies and Learning Design (LD). It is focused on designing of personalized learning using IMS LD elements. The paper analyzes the possibilities of applying Web 2.0 technologies for defining a broaden set of activities and creating the rich environments with microknowledge objects and web services in one successful scenario.

Keywords: Knowledge management, Web 2.0 technologies, IMS Learning Design, personalization, adaptation

Categories: H.5.1, H.5.2, H.5.3, H.5.4, H.3.5, H.1.2

1 Introduction

Knowledge includes both the experience and understanding of the people in the organization and the information artifacts, such as documents and reports, available within the organization and in the world outside [Marwick, 2001]. Polanyi divided knowledge into two types [Polanyi, 1996]: explicit knowledge, the knowledge included in documents or books and tacit knowledge, the knowledge that can be acquired by experience, communication.

Knowledge Management provides mechanisms for creating capturing, storing, transferring and reusing of knowledge objects. One of the most difficult tasks for the information technologies nowadays is to capture a tacit knowledge and to convert it

into explicit and searchable form. Web 2.0 technologies could be used to solve this problem.

In Web 2.0 world the knowledge and knowledge management have received the new meaning and dimension. The knowledge is named microknowledge and it is represented in form of microcontent object that has at least one unique identity with at least one unique address on the network, and that encapsulates no more than a small number of central ideas [www.mindingtheplanet.net, 2003]. KM 2.0 has appeared to support knowledge processing mechanisms in Web 2.0. KM 2.0 supports creation, sharing and leveraging the microknowledge in a collaborative way, through participation, dialogue, discussion, and observation. The Web 2.0 tools that are more frequently used in university and enterprise environments nowadays are: weblogs, wikis, social bookmarking, tagging, RSS/Atom syndication, podcasts, search engines, social networks, where the microcontent can be metadata, learning content, information content, instructions, comments, feedbacks [Leene, 2006], [Klamma, 2006]. Among the advantages of using the Web 2.0 technologies in the learning process there are: (1) facilitated process of knowledge objects creation by learners/educators, (2) knowledge objects enhanced by tagging and group use, (3) increased frequency of small communication/awareness knowledge objects, (4) multiple interventions over a period of time, (5) involve multiple learners/educators in the overall intervention, (6) reporting tools and (7) assessments tied directly to knowledge objects, integration of metrics and community use information to track utility.

The best way to use the most of these advantages and to improve learning process is to create instructional strategies that will control user learning paths and activities at using all of these different tools. The design of instructional strategies could also allow sharing the same learning activities and microknowledge with other professionals (Figure 1).

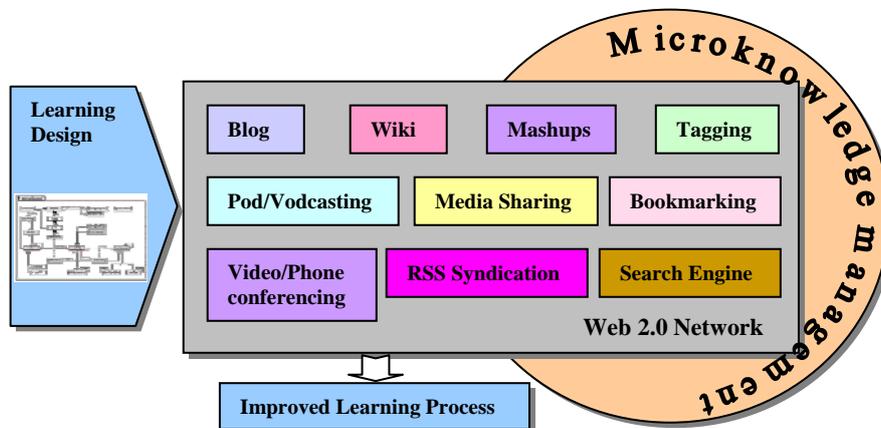


Figure 1: Microknowledge of Web 2.0 for Learning Design

Different instructional strategies can be designed for different learners to personalize their learning activities and learning content and to provide adapted services to them according their knowledge, personal interests and goals. IMS Learning Design specification provides a framework within which it is possible to describe the structure of tasks and activities, their assignment to roles, and the workflow of a Unit of Learning, and also provides a platform-independent notational convention to allow sharing and reuse of these designs [IMS Learning Design Specification, 2003].

This paper discusses the research needs and the new challenges in the mentioned above three areas: KM, Web 2.0 technologies and LD. It is mainly focused on personalized learning design task where IMS LD elements are used as well as on applying Web 2.0 technologies to define a broaden set of activities, and creating the rich environments with microknowledge objects and web services in one successful scenario. The rest of the paper is organized as follows. We discuss IMS learning design elements in section 2. Section 3 is dealt with Personalized Knowledge Management. We conclude with brief discussion in Section 4.

2 IMS Learning Design Elements

The objective of the IMS LD specification is to provide a containment framework of elements that can describe any design of a teaching-learning process in a formal way [Koper, 2006]. The first general idea behind learning design is an idea of active learning. It means that people learn better when actively involved in doing something (i.e. are engaged in a learning activity). The second idea is that learning activities may be sequenced or otherwise structured carefully and deliberately in a learning workflow to promote more effective learning. The third idea is that it would be useful to be able to record “learning designs” for sharing and reuse them in the future.

The LD elements are structured via conceptual models in three different levels – A, B, C. LD Level A includes the core vocabulary needed to support pedagogical diversity: Learning Objectives and Prerequisites, Roles, Activities, Activity Structures, Environment, Play, Act, and Method. LD Level B adds Properties and Conditions to level A, which enable personalization and more elaborate sequencing and interactions based on learner portfolios. LD Level C provides a notification capability. Every level is built on the previous one.

The core components of Learning Design identified by the IMS LD specification are based around the conceptual entity of a Unit of Learning. This is the smallest unit that satisfies one or more learning objectives (Figure 2) [Koper, 2004]. In practice this may be a course, a module, a lesson or single activity.

The next generation of the web has transformed the opportunities that are open to the Learning Designer. The Learning Designer can develop learning plans for using the existing and emerging Web 2.0 tools; as well as innovative programs that invite participation and provide transparency of the learning process.

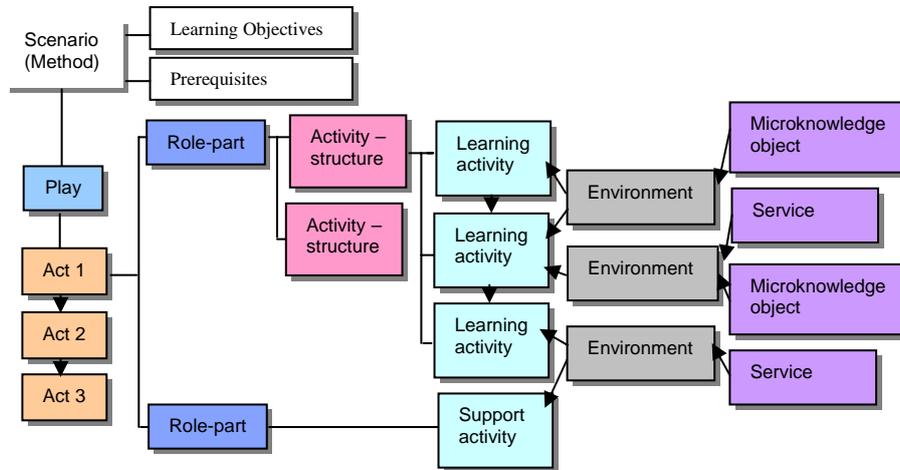


Figure 2: Conceptual View of Unit of Learning

3 Personalized Knowledge Management

Personalized support for learners becomes even more important, when eLearning takes place in open and dynamic learning and information networks like this based on Web 2.0 technologies. The pedagogical aspect of personalized learning using distributed information in dynamic and heterogeneous learning networks is still an unsolved problem in eLearning research.

The adaptive eLearning methods in context of IMS LD specification are discussed in [Burgos, 2006], [Specht, 2007]. IMS LD is able to support six main types of adaptation: learning flow based, content based, interactive problem solving support, adaptive user grouping, adaptive evaluation and changes in run-time.

In this paper is examined learning flow based adaptive method where learning process is dynamically adapted to explain the contents of the course in different ways.

Below is provided one example of Unit of Learning designed to demonstrated possibilities for personalized paths in Web 2.0 environment. The pedagogy of this Unit of Learning is presented as method (Table 3) which is successful only when the five learning objectives are achieved. Two main roles are described: the role of learner and educator. The learner actively participates in learning process and educator monitors and advices learner. Learning activities and activity structures perform in two environments (Table 2). The activities could be arranged in the way present in Table 1:

This example shows the learning flow of the learner based on her/his performance. Personalized learning paths are defined according to the learner's achievements. Personalization as the key element of the learning process offers possibilities to serve flexible, effective and efficient education, to adjust to various learning conditions and

needs. Web 2.0 networks have the potential to support collaborative microknowledge creation, to foster community building and to extend the traditional learning models. Applying Learning Design concepts to education in Web 2.0 networks led to organizing, controlling and exchanging learner’s personal microknowledge.

<i>Learning Objectives</i>	<i>Activities</i>	<i>Activity Structures</i>
Learning Objective 1: Understand and summarize the problem	Activity 1: Study MK in Wiki, Blog Activity 2: Use Podcast for understanding and communication Activity 3: Search additional MK using tags Activity 4: Pass PreTest	Activity Structure 1: Activity 1,2,3,4
Learning Objective 2: Divide and define its main items	Activity 5: Divide problem of main items Activity 6: Describe each item using blog as personal MK management tool	Activity Structure 2: Activity 5,6,7,8,9,10,11
Learning Objective 3: Prepare analysis of studied MK and create document	Activity 7: Analyze studied MK and find bottlenecks Activity 8: Prepare and publish document	Activity Structure 3: Activity 7,8,9,10,11
Learning Objective 4: Discover solution/s of the problem and evaluate each solution	Activity 9: Discover and define possible solution/s Activity 10: Evaluate solution using social software	Activity Structure 4: Activity 9,10,11
Learning Objective 5: Make conclusions	Activity 11: Make and share conclusions Activity 12: Pass Test	Activity Structure 5: Activity 5,6,9,10,11

Table 1: Activity and Activity Structures

<i>Environment</i>	<i>Description</i>
Environment 1	Blog, Wiki, Podcast, Search Engine, Social Network Sites
Environment 2	eLearning System, analyzing and documenting software

Table 2: Environments

<i>Method (Scenario)</i>				
<i>Play 1</i>				
<i>Act №</i>	<i>Role</i>	<i>Activity</i>	<i>Activity Structure</i>	<i>Environment</i>
1	Learner	---	Activity Structure 1	Environment 1
2	Learner, Educator	---	Activity Structure 2	Environment 1 and 2
3	Learner	Activity 12 – achieve Objective 1 and 3	---	Environment 2
4	Learner, Educator	---	Activity Structure 5	Environment 1 and 2
5	Learner	Activity 12 – achieve Objective 1,2,3,4,5	---	Environment 2

Table 3: Method

4 Conclusions

Nowadays it is the challenge for educators and learning designers to systematically develop new patterns and strategies suitable for personalized learning in Web 2.0 environment. This makes it necessary to understand the evolving new kinds of learning experiences, and to further enhance and enable them with the right tools, applications and contents.

In this paper is presented one model of Unit of Learning for personalized learning where is specified the method including activities and environment related to Web 2.0 technologies.

We argue that there is necessity in the further research and development of new instructional strategies of eLearning personalization in Web 2.0.

References

- [Marwick, 2001] Marwick, A.: "Knowledge management technology"; IBM systems journal, 40, 4, (2001), 814-830.
- [Polanyi , 1996] Polanyi, M.: „The Tacit Dimensijon“; Routledge & Kegan Paul/ London (1996).
- [www.mindingtheplanet.net, 2003] <http://www.mindingtheplanet.net>; "Defining Microcontent", December (2003).
- [Leene, 2006] Leene, A.: „The MicroWeb: Using MicroContent in theory and practice“; MicroLearning conference, Innsbruck, Austria, June (2006).
- [Klamma, 2006] Klamma, R., Amine Chatti, M., Duval, E., Fiedler, S., Hummel, H., Thora Hvannberg, E., Kaibel, A., Kieslinger, B., Kravcik, M., Law, E., Naeve, A., Scott, P., Specht, M., Tattersall, C., Vuorikari, R.: „Social Software for Professional Learning: Examples and Research Issues“, (2006).
- [IMS Learning Design Specification, 2003] IMS Global Learning Consortium: Learning Design Specification, (2003), <http://www.imsglobal.org/learningdesign/index.html>.
- [Koper, 2006] Koper, R., Bennett, S.: "Learning Design: Concepts", (2006).
- [Koper, 2004] Koper, R., Olivier, B.: "Representing the Learning Design of Units of Learning", Educational Technology & Society, 7 (3), (2004), 97-111.
- [Burgos, 2006] Burgos, D., Tattersall, T., Koper, R.: "Representing adaptive eLearning strategies in IMS Learning Design"; TENCompetence Conference, (2006), Sofia, Bulgaria.
- [Specht, 2007] Specht, M., Burgos, D.: "Implementing Adaptive Educational Methods with IMS Learning Design", (2007).

Towards an “Enterprise n+1” through the Use of Web 2.0 Design Patterns Enriched by Semantic Web Infrastructure

Peter Reiser, Henry Story, Franz Novak

(Sun Microsystems Inc. {peter.reiser, henry.story, franz.novak}@sun.com)

Andreas Blumauer

(Semantic Web School, Austria
a.blumauer@semantic-web.at)

Abstract: In many respects Web 2.0 and Knowledge Management (KM) are strongly related to each other. From a KM perspective the Web 2.0 evolution can serve as a pool of ideas for new ways of knowledge sharing, knowledge organisation and for the development of new architectures of measurable knowledge management systems. KM projects are usually developed in a process-oriented, goal-driven environment, embedded in complex organisational structures, whereas typical Web 2.0 applications like del.icio.us, flickr or friendster are building “their own context”. This paper will examine Tim O’Reilly’s eight generic Web 2.0 design patterns in terms of their applicability for a measurable KM System in an Enterprise 2.0. Two use cases will be presented and it will be discussed which of the design patterns could be enriched by technologies from the semantic web which will be summarized as a concept named “Enterprise n+1”.

Keywords: Web 2.0, Semantic Web, Social Semantic Web, Knowledge Management System, Social Capital, Knowledge Organisation, Social Tagging, Web 3.0

Categories: H.3.1, H.3.3, H.5.3, J.4

1 Introduction

The increasing use of Social Software in the context of the World Wide Web shows, that members of knowledge- or tagging communities like del.icio.us (<http://del.icio.us/>) are ready for new paradigms of virtual collaboration, even if the motivation of each user to share knowledge through these platforms can vary quite strongly and the majority of Web 2.0 users are still rather passive consumers [Trump, 07]. Furthermore, the underlying business models of Web 2.0 platforms are quite often not clear, and only a small percentage of new start-ups have succeeded so far. Nevertheless, many ideas throughout the Social Software Community have been transferred into Knowledge Management Systems Architectures in the last few months. In this new context expectations on social software vary from technical aspects like “easy to use”, “no need for integration into existing systems”, “web-based software and no more installations necessary” [Fraunhofer, 05] to rather economic considerations like “increasing revenues through acquiring new customers, better customer service and support” or “reducing costs through a more efficient way for employees to interact with each other” [The Economist Intelligence Unit, 07].

Stimulated by those new developments in the World Wide Web, many scientists and practitioners are working on what may be called Knowledge Management 2.0 with the aim of transferring principles acquired from social software into an organisational context supporting business processes [Tochtermann, 07]. This raises its own set of questions. Some that come to mind are for example: under which circumstances is “social tagging” accepted and integrated by employees into their daily business? Or how can the impact of such additional knowledge work be measured?

This paper will examine which of the principles of Web 2.0 can be applied in a company’s knowledge management processes, and how their limitations can be extended by the use of semantic web technologies. This thinking will lead us to two use cases revealing how the creation, exchange and reuse of semantic data (as described in [Ankolekar, 07]) can be realised in an Enterprise 2.0 setting with the aim of moving towards Enterprise n+1 [Reiser, 07].

2 Web 2.0, its design patterns and relevancy to KM systems

In 2004 O’Reilly Media Vice President Dale Dougherty coined the term “Web 2.0” during a conference planning brainstorming session, which quickly attained global buzzword status, being applied by any and all to every product that needed a little extra shine. Even serious attempts at nailing down the term, such as Tim O’Reilly’s well known paper [O’Reilly, 05] did little to focus public enthusiasm. “Web 2.0” veered dangerously close to becoming an attitude rather than a technology, making it somewhat difficult to build a serious IT strategy on. So to better understand what impact this phenomenon may have on measurable KM systems, we need to revisit O’Reilly’s original Design Patterns:

1. *The Long Tail*

In “The Long Tail” [Anderson, 04] Andersen pointed out that information technology is turning mass markets into a million niches. As the incremental costs of making goods available are lowered, companies can offer massive variety in their catalogue instead of the one size fits all blockbusters. This effect is even more true for knowledge, which by its essence is diverse and which benefits most by the information revolution. As more and more people participate in the knowledge process, making more information than ever available, as we move from an economy of scarcity to an economy of abundance, the problem also shifts from one of finding information at all to being completely overwhelmed by it. The key to increased information productivity is therefore to improve the match making process.

Now the Long Tail in an organization is not so much the documents that encode the knowledge as the people who know. By making it easy to share information, you not only expose the hidden knowledge, you also expose the knowers themselves. The benefit comes from seeing who knows what, in being able to engage them in more effective roles.

Just as companies try to reach out to the entire web, to the edges and not just the center, to the long tail and not just the head, so must more decentralization in KM System be strived for. These must work with the decentralised nature of knowledge, linking individuals to individuals, picking up data wherever it exists, and linking it

together. The universal linking nature of the semantic web infrastructure, with its ability to relate globally dispersed resources, clearly points to a very central role for this technology.

2. Data is the Next Intel Inside

Tim O'Reilly states that "Every significant internet application to date has been backed by a specialized database". Indeed, enterprises are not only facing an unforeseen growth of complexity of data, content and knowledge, they are also challenged by the always increasing need for integrating data. Classification and semantic annotation (by a combination of user-driven, expert-driven and automatic measures) of all the information in an enterprise is the key for a successful implementation. While the use of RDF is not usually part of the Web 2.0 story, it is clear that this plays especially well to its strengths. Being designed for data integration world wide, with the use of the well known URI to create a global information space, RDF is perfectly suited to be the new Intel Inside for the global distributed corporate database.

3. Users add Value

„Involve your users both implicitly and explicitly in adding value to your application." From a KM-perspective this means, that any metadata users add to a knowledge object (by tagging, rating, commenting, and even clicking on things) the more precisely important aspects of any asset can be calculated and evaluated. Amazon.com for example, ranks products by computing the "most popular" ones not only on sales but also factors some call the "flow" around products. In order to describe these actions or preferences, which take place on the WWW of information resources identified by URLs; one needs to describe them as actions on, preferences for, tags on resources identified by URLs. RDF, the Resource Description Framework, which uses the Universal Resource Identifiers as its corner stone, clearly serves as the enabler for the above mentioned Web 2.0 design pattern.

4. Network Effects by Default

The demand to "set inclusive defaults for aggregating user data as a side-effect of their use of the application" is one of the most obvious options how to transfer this design pattern straight into a Knowledge Management System: For example, each user tag improves the tag recommender system of a KM system which eventually helps to get better search results. Network effects can take many shapes or forms. Tags for example can also be disambiguated by linking them to a wiki, and allowing the wiki page owners and users to vote on their precise meaning. "The service automatically gets better the more people use it", but the service needs not be a single service. Interacting services (such as a wiki and a tagging engine) can use the distributed knowledge of the enterprise in completely unsuspected ways.

5. Some Rights Reserved

For the most efficient data-sharing this design pattern demands to "follow existing standards, and use licenses with as few restrictions as possible". Looking at today's intranet solutions, content management systems or other systems where users usually generate or distribute content, it becomes obvious, that a strong culture of ownership

is making it a lot more difficult to merge information than it should be. Although more flexible ways to define IPRs exist (like creative commons, <http://creativecommons.org/>), this lesson has not been integrated in current KM systems. Data cannot flow if it siphoned off behind legal barriers. One structural way to help align corporate interests with individual ones would be to make the cost of secrecy in an enterprise apparent.

6. The Perpetual Beta

The statement “One of the defining characteristics of internet era software is that it is delivered as a service, not as a product” reflects best what is meant by “perpetual beta”. When applying this design pattern, which is strongly linked to open source development practices, in a Knowledge Management System, we can support our knowledge intensive processes in a much more flexible way. Instead of storing all knowledge in a centralised database, we should provide smart services which are constantly developed on top of insights gained from monitoring user behaviour together with other users acting as co-developers.

7. Cooperate, Don't Control

This design pattern has been discussed for years throughout the KM community. Neither Web 2.0 nor Knowledge Management is a technological revolution: “The transformations the Web is subject to are not driven by new technologies but by a fundamental mind shift that encourages individuals to take part in developing new structures and content.” [Kolbitsch, 06] The question in the context of Knowledge Management is: How can we stimulate this mind shift? In our first use case we will consider if measuring the value of user contributions could be an answer.

8. Software Above the Level of a Single Device

The idea of the Web as a platform – “What applications become possible when our phones and our cars are not consuming data but reporting it?” makes technologies which support semantic interoperability on top of metadata standards even more necessary. From a KM-perspective this means that knowledge generation and annotation must happen on top of standard formats like RDF. From a technical perspective Tim Berner’s Lee’s proposal of an RDF bus [Berners-Lee, 05] deploying RDF mapping tools like D2R [Bizer, 03] seems to be an applicable solution which is already at hand.

3 Use cases

3.1 CE 2.0 – Customer Engineering at Sun Microsystems

CE 2.0 is a KM project currently developed at Sun Microsystems which has defined the following goals:

- Capture and Re-use of Intellectual Capital (IC) during a sales and delivery cycle
- Measure the efficiency of Re-use in the context of Business metrics (Win Rate, Revenue, Margin)

- Create a Community Equity system to build a measurable Social Capital model for virtual communities.

CE 2.0 foresees the development of an adequate architecture, a community life cycle methodology and the Community Equity system.

CE2.0 architecture is based on reusable web services, standard protocols and an extensible set of web widgets. All parts of the architecture are realised with the latest Web 2.0 technologies developed inside and outside of Sun. The CE 2.0 team identified four key services which an enterprise needs to control in order to build a scalable and measurable Community Equity infrastructure:

- *Security/IdM and Entitlement* provides cross services Single Sign On (SSO) and role based entitlement across all services
- *Tagging* as a central metadata store with a strong Ontology & Taxonomy management
- *Search* as a service for dynamic mash-ups based on metadata and semantic technologies
- *Community Equity* as an aggregation of all Community Equity values of all communities, members and content, which we describe later in this section

User tagging based on an enterprise wide ontology (SwoRDFish) [Gupta, 02] was implemented as a first step in a few applications leading among other things to prove that re-use of content increases win rates, project margins and revenues [Reiser, 06]. As a next step tagging functionality is expanded to allow a combined folksonomy and taxonomy driven tagging of content. An enterprise wide tagging service is implemented which dynamically recommends user created and enterprise controlled tags. For the enterprise controlled tags a semantic data models is used, which will be implemented as RDF/OWL ontologies. Furthermore a metadata analysis service is implemented to extract metadata from various documents to complement the tag recommendation. User applied tags are dynamically written back to Open Office documents. The *Community Life Cycle Methodology* was developed to drive a consistent community model which is called "Community in a Box". It describes the tools, roles and psychological dynamics of a community and provides a "cook book" how to build/sustain and archive communities. The objective of the *Community Equity Measurement System* is to build a dynamic Social Capital system by measuring the Contribution, Participation, Skills and Roles equity a person can gain by actively engaging in communities and social networks.

- The Contribution Equity measures the contribution of a person related to his/her role(s) in a company e.g. the value of blog entries, mailing list contributions etc. but also includes Intellectual Properties (IP) like patents.
- The Participation Equity measures the active participation of a person especially the feedback a person has provided to other community contributions like ratings, comments or tags. In addition the re-use of

existing content can be measured through the tracking of documents tags [Reiser, 06].

- The Skills Equity measures the skills a person has which is calculated based on data like enterprise skills provided by the enterprise HR system (e.g. formal skills assessment) and skills tags which are self declared skills of a person. The person can rate his/her skills tags. If the person decides to publish his/her skills, the community can rate/comment on a person's skills tag (peer review) and the person can gain additional Skills Equity.
- The Role Equity (RQ) is the aggregated value of the business, formal and informal community roles of a person in an enterprise e.g. the formal job title and level, formal community roles (Project lead, PE) or informal community roles (Community of Practice participation, CoP).

CE 2.0 is an example for a project where several Web 2.0 design patterns were combined with technologies from the semantic web. As Schaffert [Schaffert, 06] pointed out, the Social Semantic Web is a coin with two sides: It can be "Semantically Enabled Social Software" or a "Socially Enabled Semantic Web". CE 2.0 addresses the first strategy, since it is first of all a KM project built on principles of the Social Software. Existing Wikis throughout Sun, tagging behaviour and the heavy use of Weblogs are the basis for many KM activities. But, as mentioned above, the potential which user-generated content could offer for KM systems can be leveraged by the use of semantic web infrastructure.

3.2 Use case: Beatnik – Semantic Address Book

Beatnik [Story, 07] is a Semantic address book prototype that consumes and produces foaf and vcard files. It is designed to look very similar to a well known address book software in order to build on well established intuitions. The advantage of a semantic address book are numerous and simple to understand.

Firstly adding new addresses is just a matter of dragging and dropping a foaf file onto Beatnik. With over fifteen million foaf files generated by LiveJournal (See: http://blogs.sun.com/bblfish/entry/15_million_foaf_files) alone more and more such information is becoming available. Beatnik will just increase the deployment of these files, making its usage even more appealing.

Secondly, keeping up to date with one's contacts is automated. Since the foaf file is published at a well known URL all the author of the foaf file needs to do is to change his foaf file for all his contacts to have up to date information to work with. Change your address and you just need to update your foaf file using Beatnik and everyone connected to it will have their information updated the next time they do an HTTP GET on the URL.

Thirdly, foaf has a relation (foaf:knows) linking a person to their acquaintances, thereby making it possible for a foaf file to become browseable. Starting from someone's foaf file one can explore a network of acquaintances, by simple clicking on someone's name. The Beatnik foaf user agent can help one add complete or partial information about the people one knows to one's foaf file, depending on some privacy policy. The strength of a relation can be calculated by the richness of the return links. If one's foaf file says that person A knows person B, and person B's foaf file states the inverse relation, then this confirms the relationship.

Beatnik is the combination of Web 2.0 design pattern “Network effects by default”, a very easy to use interface and technologies from the semantic web. The usefulness of an Address Book like Beatnik grows with the number of applications that are able to make use of the information contained therein. Apple's Mail reader for example can communicate with the Apple Address Book and retrieve the particular image associated with an email author, and display it in the header of the email. This provides very useful mnemonic information. For Beatnik therefore to be able to make use of this information it needs to integrate well with widely used applications. Tools from the Semantic Desktop research area can be used here.

4 Conclusion: Towards an Enterprise n+1

No matter how the creation, exchange and reuse of semantic data will be realised, the success of any KM system relies heavily on the *existence* of semantic data which can be used corporate wide. The creation of semantic data can be done by users as heavily discussed in the Web 2.0 domain, by experts, or by machines using technologies like information extraction or natural language processing.

The exchange of semantic data can be built on semantic web infrastructure, for example by setting up an RDF bus for the realisation of a semantic layer on top of existing databases and applications. However, concepts for an Enterprise n+1 should not only also consist of strategies for data integration but also of systems which stimulate the exchange of information and knowledge on a community-level, which was pointed out by use case 1 – CE2.0. Finally, the second use case has demonstrated, how important easy to use interfaces and tools are, which hide away the semantic web infrastructure from the users.

- “Enterprise n+1” is the next transformation of companies which have been incorporated already a majority of the Web 2.0 principles, and therefore make use of semantic web based information management:
- Networked, project-oriented collaboration will be transformed into semantically interlinked, community based collaboration.
- Mash-ups will not only recombine existing information but will also be coupled to business processes.
- Not only information will be open and linked to other resources but also people as social networks and therefore be able to rate available knowledge
- Software no longer will be available only as a service but as an open source service
- Web 2.0 design pattern “perpetual beta” will be replaced by “real time” services
- Enterprise n+1 will rely on a three-layered approach, like the A-O-I-Model [Blumauer, 06] combining smart applications built on user interfaces like the semantic desktop, the organisation level and Semantic Web infrastructure

References

- [Anderson, 04] C. Anderson, "The Long Tail".
<http://www.wired.com/wired/archive/12.10/tail.html> (2007-05-18)
- [Ankolekar, 07] A. Ankolekar; M. Krötzsch; T. Tran; D. Vrandečić: „Two Cultures. Mashing up Web 2.0 and the Semantic Web”, In: Position Paper, WWW 2007
- [Berners-Lee, 05] T. Berners-Lee, „Putting the Web back in Semantic Web”.
<http://www.w3.org/2005/Talks/1110-iswc-tbl/> . In: ISWC 2005, Galway
- [Bizer, 03] C. Bizer. D2R MAP - A database to RDF mapping language. In: WWW (Posters), 2003.
- [Blumauer, 06] A. Blumauer; T. Pellegrini: „Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen“, In: „Semantic Web – Wege zur vernetzten Wissensgesellschaft“, T. Pellegrini, A. Blumauer, Springer Verlag, 2006
- [Fraunhofer, 05] Fraunhofer Wissensmanagement Community (Hrsg.), „Wissen und Information 2005“, Fraunhofer IRB Verlag, 2005
- [Gupta, 02] R. Gupta; M. Risteff: „An Emerging Force in Knowledge Management“, Technical White Paper, Sun Microsystems.
<http://www.sun.com/products-n-solutions/edu/whitepapers/pdf/Knowledge-Mgt.pdf> (2007-07-05)
- [Kolbitsch, 06] J. Kolbitsch; H. Maurer: „The Transformation of the Web: How Emerging Communities Shape the Information we Consume“, In: Journal of Universal Computer Science, vol. 12, no. 2 (2006)
- [O'Reilly, 05] T. O'Reilly: „What is Web 2.0?”.
<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> (2007-05-21)
- [Reiser, 06] P. Reiser: “Messbares Wissensmanagement”, In: Mit Wissensmanagement besser im Wettbewerb, Proceedings der KnowTech 06, CMP-WEKA Verlag, 2006
- [Reiser, 07] Peter Reiser's Weblog, „Web 2.0 applied in an Enterprise – a huge business opportunity”. <http://blogs.sun.com/peterreiser/resource/RWWeb+1.pdf> (2007-05-20)
- [Schaffert, 06] S. Schaffert: “Semantic Social Software: Semantically Enabled Social Software or Socially Enabled Semantic Web?”, In: Semantic Systems – From Visions to Applications, Proceedings of the Semantics 2006, Verlag der OCG, 2006
- [Story, 07] H. Story: “Beatnik – On the road to the Semantic Desktop”.
<https://sommer.dev.java.net/source/browse/sommer/trunk/misc/AddressBook/www/Beatnik.pdf> (2007-05-21)
- [The Economist Intelligence Unit, 07] The Economist Intelligence Unit: “Serious business – Web 2.0 goes corporate”, 2007
- [Tochtermann, 07] K. Tochtermann: “Neueste Trends im Wissensmanagement”, In: ÖGAI Journal 26/1, Jahrgang 2007, Nr. 1, 4-8
- [Trump, 07] T. Trump; W. Klingler; Maria Gerhards: “Web 2.0 - Begriffsdefinition und eine Analyse der Auswirkungen auf das allgemeine Mediennutzungsverhalten“, result Gmbh und Medienforschung des SWR im Eigenverlag, 2007

Making Expert Knowledge of Adaptations of E-Learning Material Available with Patterns

Birgit Zimmermann

(SAP Research, SAP AG , Germany
birgit.zimmermann@sap.com)

Christoph Rensing

(TU Darmstadt, Germany
Christoph.Rensing@KOM.tu-darmstadt.de)

Ralf Steinmetz

(TU Darmstadt, Germany
Ralf.Steinmetz@KOM.tu-darmstadt.de)

Abstract: Adapting E-Learning material allows re-using existing material even in changed usage scenarios. But this adaptation is a complex task. To achieve a perfect result expert knowledge is needed. Often adaptations have to be performed by persons who are not experts in performing these tasks. To enable those persons to achieve a satisfying result they need to be supported. Patterns are one possibility to make expert knowledge on certain tasks available to other persons. In this paper an approach is presented how expert knowledge of performing adaptations of existing E-Learning material can be collected and made available with patterns. This approach can also be used to collect knowledge of other processes, e.g. in companies.

Keywords: Knowledge Transfer, Patterns, Adaptation, Re-Purposing, Re-Use

Categories: D.2.10 – Design, D.2.13 - Reusable Software

1 Introduction

As the creation of E-Learning material is a costly task the re-use of existing material is an important topic in current research [Duval, 03]. But it is often not possible to re-use existing material as the usage scenario has changed and the existing material does not completely fit the new needs. It is then necessary to adapt the material to the changed requirements.

Adapting existing E-Learning material to changed requirements is a complex task: There exist lots of different adaptations like layout changes, translations, didactical changes etc. Often a combination of several adaptations has to be executed to achieve good results. In addition E-Learning material uses quite a lot of different formats [Zimmermann, 06b]. To be able to work with all formats users need knowledge of the tools used to work with the particular formats. This means that expert knowledge is needed to achieve a perfect adaptation result. But often persons have to perform the adaptations who are subject matter experts but not experts in performing all needed adaptations. To be able to achieve a satisfying result nevertheless, those persons have to be supported in performing the adaptations.

Patterns are a widespread methodology to provide expert knowledge and best practices. A pattern describes a proven solution to a recurring problem that occurs in a specific context. Patterns document expert knowledge of the topic they deal with. In this paper an approach is presented how to use patterns to capture the existing knowledge of adaptations and to make this knowledge available to those persons who have to perform adaptations without having expert knowledge.

The structure of the paper is as follows: The second chapter provides an introduction on adaptations of existing E-Learning material; a short overview of patterns is presented in chapter three. Chapter four describes how expert knowledge of performing adaptations can be collected with patterns and how the resulting patterns can be used to transfer the collected knowledge to other persons. Finally a summary and an outlook on future work are given.

2 Adaptation of Existing E-Learning Material

As mentioned before the adaptation of existing E-Learning material in order to make it suited for changed requirements is complex. In this chapter an overview of this task will be given.

The adaptation of existing E-Learning material often is a prerequisite for a reasonable re-use of existing material: Whenever the new usage scenario differs from the original scenario an adaptation is needed. There are a lot of different aspects in respect of which the material can be adapted. In [Zimmermann, 06a] 15 different kinds of adaptations are identified as an outcome of a user survey.

Having a closer look on the changes done during the adaptations you will see that the adaptations can be grouped. There are three major categories of adaptations:

- Adaptations that primarily change the appearance, the *layout* of the material
- Adaptations targeting on changing the *content* of the material
- Adaptations that can have an effect on layout or content but that are primarily *caused by technical requirements*

In [Zimmermann, 06b] the 15 adaptations have been assigned to the three categories. The following figure shows this assignment.

Layout	Content	Technical reasons
<ul style="list-style-type: none"> • Design • Printability • Screen resolutions • Accessibility 	<ul style="list-style-type: none"> • Translation • Learning objective • Terminology • Degree of interaction • Semantic density • Learning strategy • Difficulty of the course • Duration of the course 	<ul style="list-style-type: none"> • Transformation into several formats • End devices • Bandwidths

Figure 1: Classification of adaptations

An accurate analysis of the adaptations [Zimmermann, 06a] has shown that there are some adaptations that are performed in a very structured way. Most of these adaptations can be performed at least partly automated by a tool. But there are also some unstructured adaptations that are performed mainly based on experiences. For these adaptations it is not possible to specify exactly what has to be done in which order. In addition it is mostly not possible to offer a reasonable automated tool support. But for both kinds of adaptations (structured and unstructured) it is possible to detect best practices, describing what has to be taken into account when performing the adaptations.

There are several other approaches dealing with re-use or re-purposing of existing E-Learning content like [Hörmann, 05] or [Obrenovic, 04]. But these approaches do not take into account that the usage scenario might have changed and the existing material is not completely suited for the new scenario. Other researchers concern the adaptation of material with respect to the learner like the work in the AHA project [Stash, 06]. All these approaches do not take into account that it might be necessary to adapt the existing material before re-using it.

3 Patterns

Patterns represent a well accepted method to capture expert knowledge on best practices. Originally patterns were invented by the architect Christopher Alexander to describe the solution to recurring problems in urban planning and building architecture. By now patterns are used in many different areas as software design, security, pedagogy, etc to provide best practices.

Patterns have a long history in documenting best practices and expert knowledge. Alexanders's pattern language [Alexander, 77] documents knowledge of designing towns, buildings, and construction. The POSA book [Buschmann, 96] covers the knowledge of the authors, leading practitioners in software development, of software architecture. The Gang of Four (GoF) has collected best practices in object oriented software design [Gamma, 95]. Martin Fowler's analysis patterns "*are groups of concepts that represent a common construction in business modeling*" [Fowler, 96]. Schumacher et al. published a collection of patterns on basic security problems and their solutions [Schumacher, 05].

Patterns are written in natural language. Mostly a certain structure is used to provide the information a pattern captures. A lot of different formats exist but there is a certain agreement on the mandatory elements of a pattern [Meszaros, 96]: A pattern has to be identified by its *name*. A pattern describes a *solution* to a *problem* occurring in a certain *context*. *Forces* describe the trade-offs that must be considered when applying the pattern. In addition there are several other elements that can be used to describe a pattern.

We started working on a pattern language describing the adaptation of existing E-Learning material in order to make it suited for changed requirements. A first set of patterns of this language has been published in the proceedings of VikingPloP06 [Zimmermann, 06c]. In this paper it is described how these adaptation patterns have been created and how they can be used to transfer the collected expert knowledge to other persons.

4 Pattern Based Knowledge Transfer

As explained before expert knowledge is needed to achieve a perfect outcome of E-Learning material adaptations. But there exist also persons who are experts in the domain covered by the material but who only occasionally deal with adaptations. Nevertheless they have to perform adaptations. To allow these persons to achieve a satisfying result they need support. In the following it is described, how patterns can be used to document best practices as a basis for such a support.

4.1 Collecting Knowledge with Adaptation Patterns

Patterns are a well accepted way of documenting best practices. As they are written in natural language they are easy to understand. No special training is needed to read a pattern, in contrast to many modeling formalisms used during analysis and design phase in software design. Many of the experts working with E-Learning material adaptations have no knowledge of software design. Patterns offer them a comfortable way to provide their knowledge to other persons.

Patterns are not invented but found. They document proven solutions to existing and recurring problems. Therefore the knowledge provided by a pattern first has to be detected, before it can be written down. We used an iterative process to write the patterns documenting the knowledge of experts of adaptation processes [Fig. 2]:

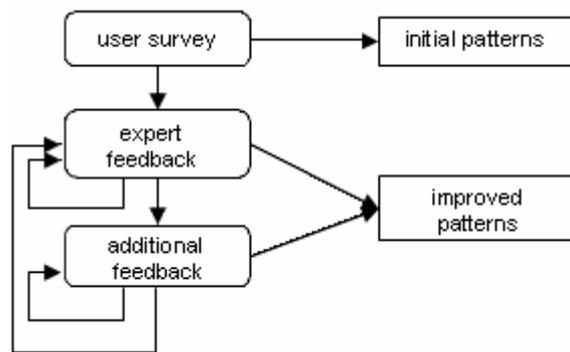


Figure 2: Pattern creation process

We started our work with a user survey [Zimmermann, 06a]. In this survey we asked persons working in the area of content creation how they perform certain adaptations. We analyzed the descriptions we got from those persons. Then we formulated a first rough version of initial patterns that represent a basic idea for each adaptation.

In a second step for each pattern we talked to experts of executing the adaptation described in the pattern. By this we refined our initial patterns and came to improved patterns. For patterns it is very important that they cover proven solutions. Therefore we determined for each pattern in which situations the solution described in the pattern has been applied successfully. Depending on the adaptation some of the

patterns went through several iterations during the second step. This means that we talked to several persons and refined the patterns again and again.

In a third step we presented some of the results at a pattern conference [Zimmermann, 06c], where we got additional feedback. This again is an iterative process: Publishing a paper at a pattern conference starts with the so called “shepherding process”. Shepherding is a one-to-one review. The reviewer (the so called *shepherd*) has to be an experienced pattern writer. The shepherd helps the authors of the patterns to improve their work and to assure that the work really represents patterns. When the patterns are in a good shape they are prepared for the *Writer’s Workshop*. The writer’s workshop offers a structured process for additional feedback on the patterns. When a pattern has gone through the shepherding process and the writer’s workshop, it has improved again and is ready to be published.

By now a reader can be sure that the patterns do not reflect the personal opinion of the authors. They are based on expert knowledge and are proven by a lot of experienced pattern writers.

The strong emphasizing of collecting knowledge from experts in the first two steps of our procedure reminds of the Delphi method [Linstone, 02]. This method can be used for example in the creation process of ontologies. The Delphi method is a structured process to collect and distill expert knowledge. It is based on written expert surveys. In contrast to the traditional proceeding of the Delphi method we decided to perform both the initial interview in the first phase as well as the feedback in the second phase not as a written survey. The answers we expected from the experts would have been too long to be written down in a bearable amount of time. Therefore we did the initial survey in the form of a structured telephone interview that has been recorded and then written down. Based on the results of this survey the initial patterns were created. They were taken as a basis for the feedback round with experts during the second phase.

Shepherding and writer’s workshops, the elements of the third phase in our approach, are the traditional process to publish patterns in the proceedings of a pattern conference. They are not taken into account in the Delphi method. But nevertheless they are very valuable as they bring the feedback of experts in pattern writing into the patterns.

The patterns that are outcome of this proceeding are a valuable source of knowledge to experts as well as to novices or laymen.

4.2 Elements of an Adaptation Pattern

There exist a lot of different pattern formats. Some pattern authors even decided to take no fixed format at all, like Martin Fowler for his analysis patterns [Fowler, 96]. As a fixed format increases readability and understandability we decided to take a fixed format for the adaptation patterns. The format is based on the GoF format [Gamma, 95] as well as on the POSA format [Buschmann, 96]. It consists of the elements shown in table 1.

4.3 Providing Knowledge Based on Adaptation Patterns

The purpose of our work is to support non-experts in performing E-Learning material adaptations. The adaptation patterns themselves are a valuable source of knowledge.

Someone who has to perform an adaptation can look up what has to be done in the patterns.

As explained before some of the adaptations can be (at least partly) automated whereas for other adaptations it is not reasonably possible to automate the adaptation. To offer a more comfortable support for both kinds of adaptations we decided to create a wizard. Based on the knowledge collected with the patterns this wizard guides users through the adaptations and offers automated support where this is reasonably possible. If no automated support is available the wizard guides users step by step through the adaptation by providing information what has to be done in each adaptation. (This information is available in the adaptation patterns.) In case of an unstructured adaptation tips and tricks and hints on how to proceed are offered. In addition it is always possible to view the whole pattern for each adaptation and to take this as a guideline.

The patterns are provided in an XML-notation. They are rendered to HTML and displayed in a browser window. Thus it is possible to render the patterns differently for an expert or a novice. This is useful as an expert often needs less information than a novice: Whereas experts use the patterns to remember what they have forgotten, e.g. after a long holiday, novices need detailed information. They have never or only seldom performed the task they want to get information about. Figure 3 shows an example for this: a part of the design pattern rendered for a novice user.

When novices are more experienced they have the possibility to change the options of the wizard and to declare themselves as experts. This does not only impact the rendering of the patterns but also the whole wizard: For novices many explanations to all steps of an adaptation are provided. In contrast experts do not need as much information. They like a fast support that allows them to do their job more shortly. This is taken into account in the wizard.

Name	Unique identification of the pattern.
Classification	An asterisk to identify the degree of confidence of the authors into the pattern (no asterisk = lowest degree of confidence).
Intent	A short overview what the pattern is about.
Context	The situation in which the pattern can occur.
Problem	The problem that has to be solved.
Example	Helps the user to understand the pattern.
Forces	The trade-offs that have to be considered.
Solution	Tells the reader how to solve the problem.
Known Uses	Situations where the pattern has proven to be successful.
Consequences	The positive and negative consequences of performing the solution.
Related Patterns	Patterns of other pattern authors that deal with the same or a similar problem.
Connected Patterns	After performing the solution of this pattern it might become necessary to execute the connected patterns.
Used Patterns	Other patterns that are used when performing the solution.

Table 1: Elements of an adaptation pattern

Design Adaptation

You want to do the following: Adapt the design of materials to match incoming requirements.

As for many kinds of content the design is very important for E-Learning content. Therefore you should always take care of a design matching all requirements. If there is a change in design requirements it is necessary to adapt the course to the new requirements. There are several reasons for a change in the requirements, e.g. if a course was originally designed for one company and should be re-used in another company or if the style guide of a company changes.

The solution described here helps you to solve the following problem: You want to adapt your course to changed requirements concerning the (corporate) design. What do you have to do in order to achieve a design that fits the new requirements?

The figure displays two screenshots of an e-learning interface for a course titled "Order to Cash".

The top screenshot, labeled "LOGO", shows a course menu on the left with options for Unit 1, Unit 2, Unit 3, Unit 4, and Assessments. The main content area is titled "About this Course" and contains the following text: "The Order to Cash course is designed to introduce you to the business activities starting from the time a customer order is placed to the time a payment is remitted." Below this, it states "The Order to Cash course will prepare you to:" followed by three bullet points: "Make informed business decisions that are founded on a conceptual understanding of SAP's Order to Cash business solutions.", "Bring you SAP Order to Cash product to a baseline level that will enable you to participate effectively in more detailed courses.", and "Identify how a company can use SAP solutions to support and optimize its Order to Cash business process." A photograph of a woman wearing a headset is positioned to the right of the text.

The bottom screenshot, labeled "My Company", shows a similar layout. The course menu is on the left. The main content area is titled "About this Course" and contains the text: "The Order to Cash course is designed to introduce you to the business activities starting from the time a customer order is placed to the time a payment is remitted." Below this, it states "The Order to Cash course will prepare you to:" followed by two bullet points: "Make informed business decisions that are founded on a conceptual understanding of My Company's Order to Cash business solutions." and "Bring you My Company Order to Cash product to a baseline level that will enable...". A photograph of a woman at a desk is positioned to the left of the text.

Figure 3: The design pattern

5 Summary and Outlook

The adaptation of existing E-Learning material to make it suited for changed requirements is an important prerequisite for the re-use of existing material. As this is a complex task it is desirable to provide the needed knowledge to non-experts who have to perform the task. In this paper an approach has been presented for collecting expert knowledge of adaptations with patterns. It has been shown how this knowledge has been collected and how it can help other persons in performing an adaptation.

With the approach presented here it is also possible to collect knowledge of other processes and to make it available to other persons, e.g. to all employees within a company. This offers an easy to use enhancement to existing knowledge management practices.

By now we are working on collecting knowledge on additional adaptations as well as knowledge of additional processes and making it available via patterns.

Furthermore we plan to improve the wizard guiding users through the processes by a component helping users to identify the adaptation suited for their requirements.

Acknowledgements

This work is supported by the German Federal Ministry of Economics and Technology in the context of the Content Sharing project.

References

- [Alexander, 77] C. Alexander, et al.: A Pattern Language: Towns, Buildings, Construction. Oxford University Press, 1977.
- [Buschmann, 96] F. Buschmann, et al.: Pattern-Oriented Software Architecture – A System of Patterns. Wiley and Sons, 1996.
- [Meszaros, 96] G. Meszaros, J. Doble: A Pattern Language for Pattern Writing, 1996. <http://www.hillside.net/patterns/writing/patterns.htm>
- [Duval, 03] E. Duval, W. Hodgins, A LOM research agenda. Published in Proceedings of the twelfth international conference on world wide web, 2003.
- [Fowler, 96] M. Fowler: Analysis Patterns: Reusable Object Models. Addison-Wesley, 1996.
- [Gamma, 95] E. Gamma, et al: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley, 1995.
- [Linstone, 02] H. Linstone, M. Turoff: The Delphi Method: Techniques and Applications, 2002. <http://www.is.njit.edu/pubs/delphibook/>
- [Hörmann, 05] S. Hörmann et al.: ResourceCenter - A Digital Learning Object Repository with an Integrated Authoring Tool Set. In: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2005, Montreal 2005; pp. 3453-3460.
- [Obrenovic, 04] Z. Obrenovic, D. Starcevic, B. Selic: A Model-Driven Approach to Content Repurposing. IEEE MultiMedia, 11 (1), 2004; pp. 62-71.
- [Schumacher, 05] M. Schumacher: Security Patterns - Integrating Security and Systems Engineering. John Wiley & Sons, 2005.
- [Stash, 06] N. Stash, A. Cristea, P. De Bra: Adaptation to Learning Styles in E-Learning: Approach Evaluation. In Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2006; Chesapeake 2006; pp. 284-291.
- [Zimmermann, 06a] B. Zimmermann, et al.: A Requirements Analysis of Adaptations of Re-Usable (E)Learning Content. In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2006; Orlando 2006. S. 2096 – 2103.
- [Zimmermann, 06b] B. Zimmermann, C. Rensing, R. Steinmetz: Formatübergreifende Anpassungen von elektronischen Lerninhalten. In Proceedings of DeLFI 2006: 4. e-Learning Fachtagung Informatik, 2006; Darmstadt 2006. pp. 15 – 26.
- [Zimmermann, 06c] B. Zimmermann, C. Rensing, R. Steinmetz: Patterns for Tailoring E-Learning Materials to Make them Suited for Changed Requirements. In Proceedings of VikingPLoP 2006; Helsinki 2006.

Using Taxonomies to Support the Macro Design Process for the Production of Web Based Trainings

Abdelhak Aqqal

(TU Darmstadt, Germany
abdelhak.aqqal@KOM.tu-darmstadt.de)

Christoph Rensing

(TU Darmstadt, Germany
Christoph.Rensing@KOM.tu-darmstadt.de)

Ralf Steinmetz

(TU Darmstadt, Germany
Ralf.Steinmetz@KOM.tu-darmstadt.de)

Abstract: Today, Web Based Training (WBT) starts to be widely used as a new way of teaching. Unfortunately, this mode of teaching imposes new requirements and constraints. It has made the creation of learning materials a complex and demanding task for the instructors because it takes much time and demands a multitude of skills, in particular technical skills that must be developed and continuously updated. Hence, we propose a collaborative authoring methodology based on division of labor as a way to produce WBTs where the processes of production are clearly separated to meet the existing and needed skills of persons involved in WBT production. This paper presents an efficient method to support instructor's guidance during the first phase of the WBT production called the Macro Design using the Rhetorical Structure Theory (RST) and taxonomies we developed.

Keywords: E-learning, Production of Web Based Training, Taxonomies, Collaborative Authoring, Knowledge Modelling, Semantic Design, Instructional design support tool.

Categories: H.4.0, H.5.4, I.6.5, K.3.1

1 Introduction

Since the integration of web technologies in teaching environments, education has undergone a shift in paradigm. An example of this shift is seen in Web Based Trainings (WBTs) that can be offered at any time and at any location as long as an Internet-enabled computer is available. However, this new mode of teaching has made the creation of learning materials a complex and demanding task for the instructors because it demands a multitude of skills, in particular technical skills that must be developed and continuously updated [Aqqal, 07]. In contrast, an instructor (at school or university as well as in a company) is a domain expert first and usually lacks technical skills needed for WBT authoring and media creation [Aqqal, 07]. Hence, one of the important requirements of an adequate approach to produce WBTs is that the technical efforts spent by the instructor in authoring and media creation should be reduced to a minimum. This helps instructors to refocus on instructional aspects rather

than technology [Helic, 02]. With this idea in mind, we have proposed a collaborative authoring methodology (figure 1) based on division of labor as a way to produce WBT where processes of the production are clearly separated to meet the existing and needed skills of persons involved in WBT production. WBT production should be done in three different levels of abstraction: the semantic, logical and physical levels handled respectively by three processes: the so called “Macro Design”, the content modelling and the content authoring & media creation. In addition we define vertical to these processes a production management process in order to harmonize the collaboration between actors.

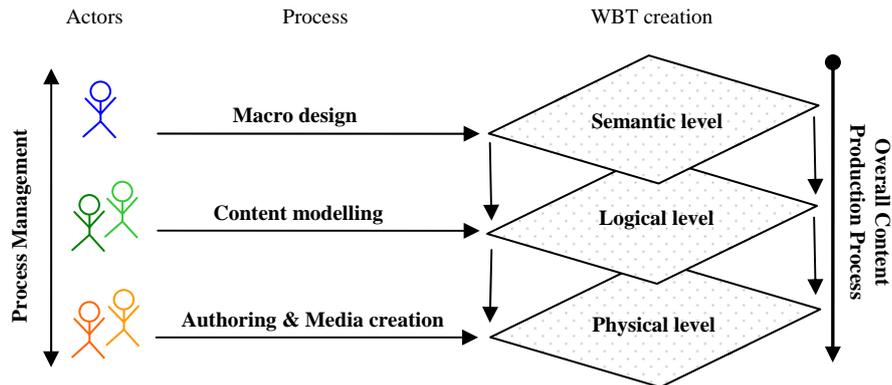


Figure 1: The proposed approach for the overall content production

The so called “Macro Design” will be explained in the next section in detail. We explore why taxonomies are necessary to support the creation of an abstract representation of WBTs. Section 3 describes the application of the Rhetorical Structure Theory (RST) as a mechanism to enhance the expressiveness of WBT design and to assist instructors when designing WBTs. We shortly introduce our developed taxonomies that enrich and extend the RST to meet our requirements. An example is given to illustrate our approach. The fourth section surveys related work and discusses the shortcomings regarding our requirements. Finally, we present some conclusions and remarks for further work in this area.

2 The Macro Design : Introduction & requirements

2.1 The need of the Macro Design for WBTs production

In contrast to existing ways of WBT production, we postulate a phase in addition to content modelling, authoring and media creation which is often neglected or not fully taken into account. This phase, temporarily called “the design thinking”, covers instructor’s ideas about what kind of WBT to produce, about a motive, reasons for a specific target group, and about a list of themes needed to be taught. The instructor defines implicitly cognitive boundaries of main concepts of his WBT and semantic relations among these concepts according to both knowledge and learner domains. The “design thinking” is done in the mind of the instructor only. Tool support starts in

the content modelling phase nowadays. Commonly WBT modelling uses the table of content paradigm. Such a table of content records the main concepts used in content authoring only. The relationships between the main concepts as well as the instructional impact can not be expressed in such a simplified model. We introduce, the “Macro Design” as an explicit modelling phase corresponding to the “design thinking” in order to record what instructors have in mind and to forward instructors ideas and intentions to all others involved in the WBT production, from the instructional level to the technical level [Aqqal, 07]. The capturing of such intentions aims to enhance the awareness and comprehension of the production context and will increase, as a metadata, consequently the chance to re-use parts of a produced WBT.

Simply stated, the Macro Design could be summarized into answering explicitly the following:

1. Why to produce a WBT and for which audience?
2. What to produce (in term of knowledge)?
3. In which form to produce this WBT and why in this form?

In this paper, we principally focus on the third question. Here, our goal is to develop a mechanism supporting instructors to transform the intentions resulting from their design thinking into an explicit how-to-product specification given via small editing steps. The next section specifies main requirements the proposed approach should fulfil and introduces the need of taxonomies to support the Macro Design.

2.2 The requirements and the need of taxonomies for the Macro Design

Our goal is to build a tool supporting Macro Design without overhead for the instructor. To realize this, *Macro Design* has to meet some requirements; most importantly it has to be simple and intuitive. The tool should not impose a certain pedagogical model for the instructor to avoid any semantic mismatch conflict between instructor intentions and the model mapping his intentions. Secondly, guidelines are needed to determine how the instructor should express his intentions, how to supervise and progress the whole production process. This can be done via step-by-step guidance. Therefore, a semantic taxonomy is required as vocabulary for the representation of the WBT including “design thinking” data. Thus, the instructor will be supported in instructional design and the structuring & expression of his domain knowledge away from more technical content authoring and media creation (figure 2).

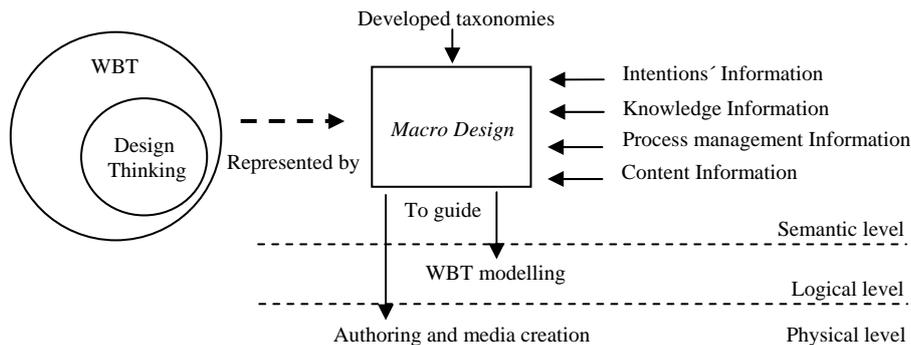


Figure 2: Using the Macro Design to support the WBT production

Semantic taxonomies in general are declarative classifications of different semantic elements in terms of a uniform vocabulary. We distinguish in our area of application at least two kinds of taxonomies: (1) taxonomy used for the representation of the WBT into aggregation of semantic parts called semantic units and (2) taxonomy of semantic relations among these elements and the way they are connected in the WBT to reflect the instructor's intentions and the knowledge mapping as well.

To suit our scenario of use, developed taxonomies have to support an instantiation by queries and should establish the correspondence between instantiated elements and the instructor's intentions. It should also reinforce a separation between the different production levels so that each level will be mastered before progressing to the next. Using taxonomies in this way for the WBT production provides many advantages over traditional authoring methodologies. Notably, an increased separation of design and authoring levels as well as an abstraction mechanism to support a step-by-step production via suggested proposals given to the instructor instead of free-to-write forms. Thus, the production is easy, fast, and deterministic. The next section introduces our developed taxonomies and their usage.

3 Towards an adaptation of Rhetorical Structure Theory (RST)

3.1 Introduction

The goal of each training apart from skill training is to transfer knowledge from a given domain to the learner. Formally, it can be expressed as follows: Web Based Trainings transfer knowledge from the WBT knowledge domain (the WBT domain) to the learner knowledge domain (the learner domain). Both WBT domain and learner domain are collections of concepts, where a concept is an independent unit of knowledge. For example, "how to insert an image into a Web page" could be a concept in WBT domain called "HTML introduction". In *Macro Design* both WBT domain and learner domain have to be described by the instructor.

3.2 Development of a taxonomy for semantic units

To get a WBT model representing the ideas of an instructor, the instructor has to be supported to determine the elementary units of the WBT first. Additionally, a general way describing semantic interrelationships among these units should be provided. Many related authoring approaches proposed hypotheses about what constitutes an elementary WBT unit. These hypotheses are based either on logical criteria (e.g. paragraph, section) or physical criteria (e.g. size, layout, image or page) [Aqal, 07].

For our scenario of use, we developed an initial taxonomy where we distinguish 8 types of WBT Units and their instances to fit the *Macro Design* adequately (table 1). Our segmentation of WBT documents is rather grounded on semantic basis, where fragmentation and modularization of WBT units is determined by the existence of a certain meaning or didactic function in each unit. This unit, called "a semantic unit", should be stand alone and didactically well-recognized. For instance, an illustration composed of an image and its description in paragraph format will be not considered as two units but only as one. This way of modelling fulfils our requirements. It leads

to a separation between the different production's levels. If so, the instructor has the ability to abstractly define desired content in form of a set of semantic units.

Semantic Unit	Semantic Rule	Examples of instances
Principal unit	Concept presentation	Definition (concept, theory, etc.)
Alternative unit	Concept restatement / unit's reformulation	Summary, abstract, preview
Illustration unit	Concept illustration	Simulation, elaboration, example
Activity unit	An activity description	According to the learning design
Assessment unit	Measure and evaluation of grasped knowledge	Test, exam, quiz, evaluation
Reference unit	To refer or designate a used concept or unit	Metadata, glossary, references, bibliography
Supplement unit	Supplement, information about a concept/unit	FAQ, Help, Read more, index
Connection unit	Join units to bridge semantic transitions	Background, planning, motivation, table of content

Table 1: The developed taxonomies of the semantic units in the WBT

This taxonomy categorizes in a matrix typical semantic units and their instances needed for WBT production. It also assures a minimal associative linking between a given semantic unit and its "typical" logical formatting since those semantic units are composed of logical units and seem to respect certain aggregation likelihood.

3.3 Using RST to support the Macro Design

So far, our intent was to generate a plausible taxonomy of a WBT's units to allow the *Macro Design*. To preserve our prior concepts we have adopted the Rhetorical Structure Theory (RST) as an additional mechanism to support the *Macro Design*. The RST is used, in our context, as a navigational model to contextualize and freeze the instructor's intentions beyond a simple hierarchical structuring of sections. RST [Mann, 87] is a framework for analyzing discourse structure and speech statements by positing hierarchical relations between spans of text. These relations are defined functionally, in terms of what their intended effect on the reader is.

RST has been chosen because it has many features meeting our requirements. First, RST is a natural and neutral mechanism for semantic modelling that specifies a rigorous set of annotation guidelines without imposing any prior model for the conception. Secondly, RST respects perfectly our developed semantic taxonomy and its requirements. It assumes that a text is divided semantically into autonomous units according to speaker intentions. These units are related by named rhetorical relations

and structured into two kinds (a nucleus and a satellite) that reflect their importance according to the speaker's intention. We also suppose that WBT segmenting (implicit or explicit) and relations between segmented elements reflect instructor intentions.

Finally, since the discourse generation and WBT production are two analogue processes, the taxonomy of semantic relations developed already by the RST community [RST, 07] is seen to be relevant for our scenario of use. This taxonomy of relations should be extended beyond the application area it was originally designed for. The extended taxonomy should be significant enough in converting the WBT structuring into a way to explore the instructor's intentions.

3.4 An RST based Macro Design Component as part of production tool

To implement a RST based tool to support the *Macro Design* we had to adapt RST formalism to our scenario of use. This RST based modelling will be implemented in a tool and allow the instructor to express his Macro Design (didactic modelling, domain knowledge modelling and WBT segmenting) for the ongoing production process. To understand, we simplify briefly in the figure 3 an example of design via RST from a given learner and WBT domains to an abstract representation of the WBT.

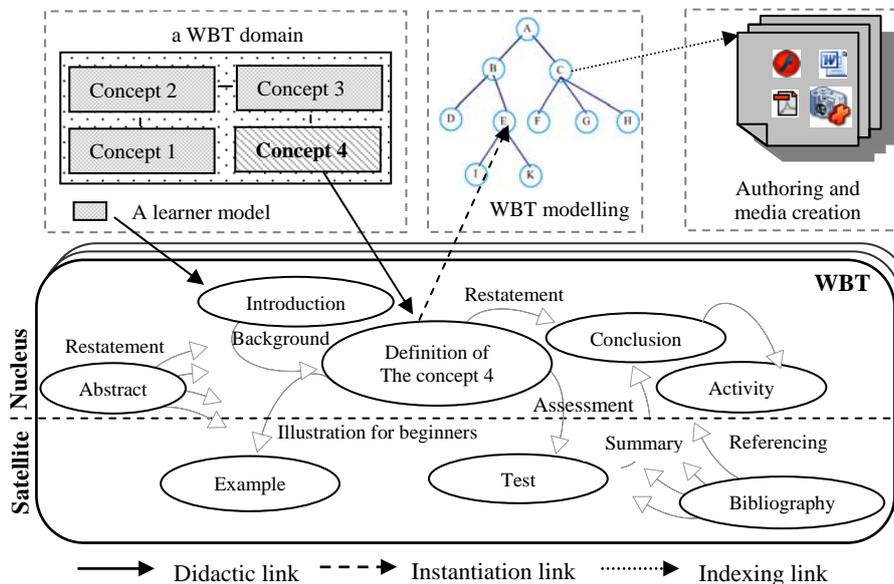


Figure 3: Using RST and taxonomies to support the Macro Design

In this example, the WBT semantic modelling shows only WBT segmenting into didactic elements and rhetorical relations among them to express some of the instructor's intentions. Here, the first thing which has to be considered is what the concepts that our WBT domain should include are, and what the concepts known by the learner (i.e. Learner domain) are. This specification will be quite easy since we have adopted a neutral and simple way of knowledge modelling. The second step is to map those concepts to certain semantic units which serve as abstract containers of

knowledge. All information which is needed about each semantic unit has to be defined explicitly such its mapping to given concepts, its semantic features or rules (e.g. nucleus/satellite), its intentional relations with other units and authoring properties if required. This specification is based on the RST framework and on our developed taxonomies. Later on, the resulting representation of the WBT when it is completed should be instantiated into a specific WBT model so that the last step to do is to enable this model by authoring and creating needed media. The modelling and authoring of WBTs must fulfill the representation and requirements given by the instructor and should be done via iterations by the process management [Aqqal, 07].

4 Related Work

Over the past years, many approaches (in academia and industry) were purposed to support the WBT production by tools [Pernin, 06] [Aqqal, 07]. However, few suppose that the WBT production is done in a collaborative way supporting different roles and skills. Hence, using existing tools for a collaborative way of working will be quite fuzzy. In particular, these tools fail usually to support Macro Design as stated in the previous section [Aqqal, 07]. Unlike our requirements, the traditional way of WBT design focuses on a flat structuring of WBT toward developing a modular view to fulfil existing technical standards requirements [Verbert, 06]. Indeed, these standards (e.g. LOM, SCORM Content Aggregation Model) have been generally limited to the modelling of object-oriented schemas because they have put significant effort into developing mechanisms to manage the reuse of materials located in a repository in form of learning objects not necessarily semantically annotated [Aqqal, 07].

In addition, there are a number of taxonomies and ontology based models that conceptualize learning resources and their mapping to the knowledge domain. Verbert and Duval [Verbert, 06] [Zouaq, 06] studied six content models and showed that they could be mapped on their abstract model called ALOCoM [ALOCOM, 05]. So far, ALOCOM refers only to slide presentations as materials to be authored [Bergsträßer, 06]. Similar approaches are found in [Bergsträßer, 06] and [Zouaq, 06]. Unfortunately, these approaches are based either on repurposing or retrieving existing resources and do not go beyond the traditional way of WBT authoring.

On the didactic level, Bloom's Taxonomy [Bloom, 56] is a relevant taxonomy to expresses educational objectives and serves as a sort of checklist to answer the first two questions listed in the section 2.1 rather than to follow all semantic connections of a WBT in terms of intentions, degree of modularity and the authoring process. IMS Learning Design (IMS-LD) [IMS, 07] considers that the focus of learning is the activity and not the content [Aqqal, 07] [Pernin, 06]. Thus, by being so abstract, generic and constructivist oriented, IMS-LD does not meet all our requirements.

5 Conclusions and further research directions

In this paper we presented a novel conceptual contribution to the Web Based Training creation. We motivated a collaborative production as way to meet instructors' skills for an efficient WBT production. The proposed methodology points out the so called "Macro Design" as an independent task to be supported. The Macro Design is

innovative in two ways. First, it extends the existing way of content design by supporting instructors in explicating their intentions and instructional data that are often not captured. Second, it demonstrates the possibility to use the Rhetorical Structure Theory (RST) as a communicative mechanism to give an explicit perception of the expected content. Hence, we have developed taxonomies that are RST adapted and fit the Macro Design requirements. We will continue to further refine these taxonomies to fully suit our scenario of use.

As proof of concept, we plan to implement an extension of The ResourceCenter tool [Aqal, 07] to support the Macro Design by addition of a Macro Design component up on this tool and to support the processes management. ResourceCenter was chosen because it constitutes an open source, browser based and instructor-friendly tool. Moreover, it supports the content modeling and authoring separately and implements already some required functionalities that we need for the collaborative production. One area of interest is the evaluation of the Macro Design concepts on the computer-communication networks domain of knowledge to be web trained but our methodology can easily be reapplied on other knowledge domains as well.

References

- [ALOCoM, 05] ALOCoM , 2005, <http://memling.cs.kuleuven.ac.be/alocom/documentation.php>
- [Aqal, 07] Aqal, A., Rensing, C., Steinmetz, R. (2007), The Macro Design as an Own Task in WBT Production: Ideas, Concepts and a Tool. EC-TEL 2007, LNCS 4753, pp.420-425, September 2007, Crete, Greece
- [Bergsträßer, 06] Bergsträßer, S., Faatz, A., Rensing, C., Steinmetz, R. (2006), A Semantic Content Representation Supporting Re-Purposing of Learning Resources. I-Know, September 2006
- [Bloom, 56] B. S. Bloom: "Taxonomy of Educational Objectives: The Classification of Educational Goals"; Susan Fauer Company, Inc, 1956, pp. 201-207.
- [Helic, 02] D. Helic, H. Maurer, N. Scerbakov: Aspects of Collaborative Authoring in WBT Systems. Proceedings of International Conference on Advances in Infrastructure for Electronic Business, Education, Science, and Medicine on the Internet, January 2002. (ISBN 88-85280-62-5)
- [Mann, 87] W. Mann, S. Thompson: Rhetorical Structure Theory, A theory of text organization. Technical Reports ISI/RS-87-190, 1987.
- [Pernin, 06] J. P. Pernin, A. Lejeune, Learning design models for computers, for engineers or for teachers?, Learning Networks for Lifelong Competence Development workshop, organized by the EU 6th Framework Integrated Project TENCompetence, Sofia, Bulgaria, April 2006.
- [RST, 07] The Rhetorical Structure Theory (RST), 2007, <http://www.sfu.ca/rst/index.html>
- [Verbert, 06] K. Verbert, J. Jovanovic, E. Duval, D. Gasevic, M. Meire, Ontology-Based Learning Content Repurposing: The ALOCoM Framework, 2006, In International Journal on E-Learning, volume 5. 67-74.
- [Zouaq, 06] A. Zouaq, R. Nkambou, C. Frasson, (2006). The Knowledge Puzzle: An Integrated Approach of Intelligent Tutoring Systems and Knowledge Management. 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06) pp. 575-582

Conceptual Data Model and Meta Framework: Modelling ICT Competence Profiles

Roman Povalej, Peter Weiß

(Institute AIFB, University of Karlsruhe (TH), Germany
{povalej|weiss}@aifb.uni-karlsruhe.de)

Abstract: Information and communication technology (ICT) skills/ competence frameworks are an important prerequisite for ICT competence development and related quality assurance for recognition and transferability of qualifications. In this paper it is argued the importance of interoperability through common or explicit semantics of ICT competence profiles. This requires modelling on basis of shared frameworks. Thus, existing frameworks have been analysed. The paper presents underlying structure and data models of some prominent systems that have achieved status of proprietary standards. Moreover, a conceptual model is derived on basis of a comprehensive analysis of respective meta data of skill/ competence grids. Presented work provides the theoretical foundation for further development of information systems for the management of knowledge, skills, competences and qualification. Applications are skill/ competence catalogues and repositories, as well as web services for data exchange between human resource information systems.

Keywords: Certification, Data Modelling, ICT Skills, Meta Framework

Categories: E.1, H.1.0, J.0

1 Introduction

Knowledge, skills and competences (KSC) and qualifications are crucial throughout the whole working life: during schooling and education, for (the first) letter of application and for a successful professional advancement. Doubtless this is unlimited valid in every nation across Europe, across the world. To make them comparable and consequently to simplify the mobility of potential employees across Europe and worldwide and also to support the Electronic Data Interchange between human resource information systems (containing e.g. descriptions of core competences, job roles/ profiles, individual portfolios) they and the related frameworks have to be analysed. Within The Copenhagen Declaration [Copenhagen, 02, p.2] this is expressed as follows: "The Investigating how transparency, comparability, transferability and recognition of competences and/or qualifications, between different countries and at different levels, could be promoted by developing reference levels, common principles for certification, and common measures, including a credit transfer system for vocational education and training." The e-Skills community and involved stakeholders are currently making an effort to establish a common European Framework for e-Skills and competences in Europe [Cedefop, 05, p.47]. Existing qualification and certification schemes are being reviewed in the context of learning provision that leads to clarify the underlying profiles, terminologies and curricula ([HARMONISE, 07], [Weiß, Stucky, Dolan, Bumann, 05]). The work presented in

this paper is part of certain pillars as basics being necessary to develop web services supporting Human Resource Processes like assessing, hiring, modelling information systems, staffing, etc.

2 Supply and Demand of e-Skills

The core competences of a company orient by marketplace scilicet orient by both the needs of their potential customers and business competition. To be on the long run successful the company has to adapt oneself to the continually altering circumstances like e.g. requirements, demands and needs of the belonging business marketplace. The employees belong to the most important resources of a company, their KSC and qualifications are part of the core competences and so they contribute to the success of the company. Hence the employers or industry determine the needed e-Skills requirements (profile A) to be most successful at the business marketplaces. These e-Skills are built by ICT (information and communication technology) Practitioner Skills, ICT User Skills, and e-Business Skills plus belonging levels (e.g. basic, core, advanced, and expert) [ESF, 04]. The employers demand from education institutions to satisfy their e-Skills requirements. However, the education institutions determine their curricula (profile B), more or less influenced by the employers. After education persons become part of the e-Skills stock (profile C) with their own portfolio (profile D). They build the e-Skills supply (profile C+D) by participating at the labour market. Employers use this labour market to satisfy their needs of e-Skills (profile A) by searching for potential employees fitting their core competences respective the open job profiles. Now there are several e-Skills profiles (A, B, C, D, C+D) which have to be comparable to each other to estimate their value and to make the right decisions for hiring employees. Because of lack of a common language to describe each of them, it is a big challenge to match them so that they are comparable. The degree of difficulty is much higher on the European level. A holistic approach is required. And in this way e-Skills certification systems play an important role but need to be more transparent and comparable, alike the accordingly skills frameworks respective job profiles.

3 ICT Frameworks in Europe

Analysis of several ICT frameworks is a basic prerequisite to develop a meta framework for describing existing reference as well as user frameworks (in [Weiß, Povalej, Stucky, 05] the working method is presented in detail). Meta and reference models offer a common language to be used for the formal description, analysis and comparison of ICT frameworks and ICT profiles. Exemplarily the ICT frameworks “The Skills Framework for the Information Age” (SFIA: <http://www.sfia.org.uk/>) and “The eCompetences and Certifications Observatory” (eCCO: [eCCO, 05], [Marinoni, 05]) are described for a better understanding of ICT frameworks: (1) **SFIA** provides a common reference model for the identification of the skills needed to develop effective Information Systems making use of ICT. It enables employers of IT professionals to carry out a range of Human Resource activities against a common framework of reference; including skill audit, planning future skill requirements,

development programmes, standardisation of job titles and functions, and resource allocation. The functionality to acquire and to manage certifications is not considered, scilicet the assessment of the skills can be compared but they are not proofed. The operative area (e.g. knowledge transfer or expert network) is also not considered. Because of missing functionality of a gap analysis many fields of application are omitted (e.g. team or recruitment planning). (2) The **eCCO** project was born just to satisfy the following needs: The challenge is to find a basis for sharing a common vision and framework with regard to competences descriptions, job profiles definitions, and ICT qualifications. Flexible job profiles, transparent learning programs, basic common language, are all shared needs. It aims at finding a basis of consent in order to build a shared “ICT grammar” fitting competences and job profiles. The developed eCCO conception should have following extensions: output of the best matched profiles accordingly to a certain matching level in percent; output of suggested trainings courses regarding to the extracted profiles; output of assessments with regard to the suggestion to get external workforces rather than training own ones; company-wide staff analysis through aggregate the gap analysis and consideration the employee view to the strategic level: team planning and staff planning would be possible and an advantage.

For the sake of completeness here are more important initiatives and frameworks which should be mentioned: CEN/ISSS Workshop ICT Skills [CEN/ISSS, 05], e-Skills UK (<http://www.e-skills.com/>), European Certification of Informatics Professionals (EUCIP: <http://www.eucip.org/>), The Advanced IT Training System (AITTS/APO-IT: <http://www.apo-it.de/>), The Computing Technology Industry Association (CompTIA: <http://www.comptia.org/>), The European e-Skills Forum [ESF, 04], The European Information & Communications Technology Industry Association (EICTA: <http://www.eicta.org/>), The e-Skills Certification Consortium (eSCC: <http://www.e-scc.org/>), etc.

4 Data Models for Electronic Data Interchange

There are various human resource information systems which have to interchange information like e.g. vacancy information (job profiles) and information about potential job applicants (individual portfolios). Especially employment agencies have to interchange information with each other and especially with employers and job fairs (e.g.: <http://www.careerbuilder.com/> or <http://www.monster.com/> or newspapers). To promote this Electronic Data Exchange between different information systems it is necessary to use a general and/or standardised data model describing job profiles and individual portfolios of potential applicants. Job profiles (represent need to manage tasks) as well as individual portfolios (represent offer to manage tasks) contain amongst others relevant KSC and qualifications which are crucial throughout the whole working life. Both IEEE Reusable Competency Definition [RCD, 07] as well as HR-XML [HRXML, 04] are standardised data models for describing competences in a re-usable way (‘competence’ and ‘competency’ are used in this paper synonymously [Abbs, Freebairn, 05]). Employers or users can define once certain competences and re-use them again and again, e.g. in different job advertisements or describing several job profiles. Further on employee can use the already defined competences to describe their own individual portfolio.

Indeed such existing data models have some weak points respective they do not fulfil all requirements extracted out of the supply and demand of e-Skills. For example [HRXML, 04] does not difference between required and acquired profiles. The main problem of [RCD, 07] is the lack of standardisation within applying it (e.g. you can use different identifier for the proficiency level, which complicates the comparability between several databases). The following clues could be extracted for our conceptual data model [Povalej, Weiß, 07]: the re-usability and the flexibility are strengthened through an object oriented approach; one competency consists of the components skills, context and proficiency level; context and proficiency level has to be clearly definable which is precondition for profile matching and domain-wide re-usability; free description area out of RCD is needed and it represents the smallest skills unit; additionally the component grade is introduced (as an example Figure 1 shows the component profile with the components: skills, proficiency level, context, competency, profile, profile unit, grade; Figure 2 shows the profile unit with the attribute components and competency and grade component).

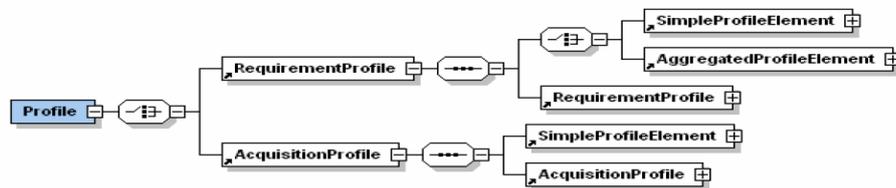


Figure 1: Exemplary Component Profile of Our Conceptual Data Model

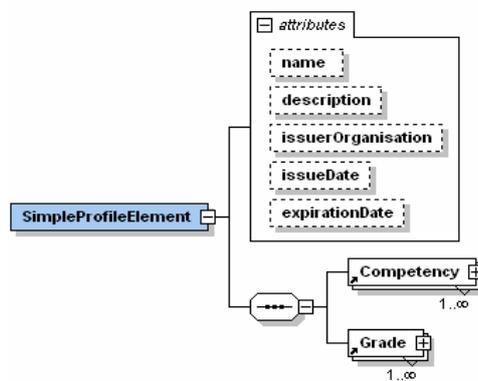


Figure 2: Exemplary Profile Unit of Our Conceptual Data Model

5 Conceptual Meta Framework

In order to achieve transparency and comparability of ICT worker profiles the smallest building block or unit of qualifications (QU) has to be defined ([CEN/ISSS,

05], [Weiß, Povalej, Stucky, 05], [Weiß, Stucky, Dolan, Bumann, 05]). QU can be described as a function with three dimensions: (1) qualification component (QC), (2) working area (WA), and (3) the belonging level (L). QC represents the general description of QU and it ought to be expressed as a set of learning outcomes (LOC), which define the relevant KSC to operate unmistakably in the underlying WA. L indicates the depth of KSC (e.g. basic, core, advanced, expert). These results are summarised in the definition of ICT worker sector profile P. An ICT worker profile is defined as the selection of the tuple (WA, QC, L) and can be described as matrix P with the job role q as follows (notation is shown in Table 1):

$$P_q = L_q = \left(l_{qij} \right)_{j=1..n}^{i=1..m} = \left(wa_i, qc_j \right)_{j=1..n}^{i=1..m} \quad (1)$$

The following example should highlight the applying of our approach: (P) Database Administrator; (QC) to install and administrate database software; (WA) banking, backbone area; (L) advanced: installation and administration of this database software have to be made independently. With this short description it is possible to develop a new or to choose out of the offered training courses by comparing them through applying of our approach.

Element/ concept	Notation
P	set of profiles; $P = \{q_1, \dots, q_m\}$
L	set of levels (vertical)
QC	qualification component as general description of the smallest unit of a qualification to be described; QC may be expressed as set of LOC on basis of KSC in relation to the intended application context (QC = LOC) (e.g. general European Qualification Framework descriptions)
WA	set of working areas (horizontal level e.g. 'sales and marketing', 'management and administration'; this is applied synonymously with the terms 'field of work', 'function', 'business process' etc.); $WA = \{wa_1, \dots, wa_m\}$; if higher granularity is required, WA may be subdivided in further sub-categories
q	job role e.g. software developer, information system analyst, etc.
i, j	indices to define KSC category based on typology (WA, QC)
m, n	indices for existing numbers of categories for WA, QC

Table 1: Overview of Applied Notation [Weiß, Povalej, Stucky, 05]

With the assistance of our developed approach it is now possible to describe, to analyse and after that to compare existing ICT frameworks, and also to model new reference and/or user frameworks, which are adapted to the working field of a company (e.g. a human resource information system).

In Figure 3 our conceptual meta framework is exemplarily applied on the ICT framework [EUQuaSIT, 04], which is "a European project that aims at contributing to more transparency of ICT work, skills and qualifications and its relationship. It intends to analyze the specific demands of companies within their ICT workforce and to what extent and how different ICT qualification and training strategies in the

supply and demand of e-Skills (or further KSC and qualifications) on the labour market is crucial for the development of a common European framework, which ought to support the transparency, interoperability, etc. to improve amongst others the mobility of potential employees in Europe and beyond.

As mentioned above there are several ICT frameworks in Europe. The most are rated on the user level (hardly applicable in other companies), some are rated on the reference level (no standard and hardly comparable to each other), but there are none on the meta-level (general description language). The meta-level is very important. At the meta-level ICT frameworks can be analysed and described by a common language (which has to be developed). Amongst others on the long run we want to develop web services to support the introduction of ICT frameworks into companies regardless which ICT framework will be taken as the starting point. The development of an additional ICT framework does not boot, as mentioned above there is a need for standardisation. Further on, a common language for describing ICT frameworks and/or job profiles is necessary and very useful as already mentioned in this paper too.

By the help of our developed and presented conceptual data model as well as our conceptual meta framework ICT competence profiles can be modelled at the meta-, reference and application respective user level. With the assistance of our methodology ICT frameworks, IT models respective qualification systems on all levels can now be analysed and compared to identify similarities and differences, advantages and disadvantages, strengths and lacks, etc.

The consequently further development of our conceptual meta framework as well as conceptual data model and the development of KSC and qualification catalogues and repositories are the next steps ought to be done. These catalogues and repositories are adjusted accordingly to the working area and/or regional respective national location, scilicet among others their advantage is the re-usability of an ICT framework regardless which catalogues and/or repositories are being used within.

Another step is the consideration of portfolios as well as ePortfolios and their closely connection to job and/or individual profiles of employees. [Butler, 06] made the following recommendation for the success of ePortfolios: "Institutions need to be aware of the impact that an electronic portfolio development will have. Electronic portfolios need to be an integral part of a programme of study, not an 'added-on' assessment, which may necessitate the review and restructuring of courses. The type of portfolio required, its purpose and its audience need to be clearly articulated."

One main objective is to develop further web services for human resources (e.g. to attract talented people by respective career guidance, information services and by highlighting flexible entry points to the IT profession), also on the base of our other research results (e.g. performance standards, taxonomy for ICT certifications). At the same time the conceptual models will be implemented to real world systems, considering these catalogues and repositories and in future also web services for a higher transparency, comparability, transferability and interchangeability.

References

[Abbs, Freebairn, 05] Abbs, B.; Freebairn, I.: Longman Dictionary of Contemporary English. Langenscheidt, August 2005.

- [Butler, 06] Butler, P.: A Review Of The Literature On Portfolios And Electronic Portfolios. Creative Commons Attribution-NonCommercial-ShareAlike 2.5 License, October 2006.
- [Cedefop, 05] Cedefop 2005: Typology of knowledge, skills and competences: clarification of the concept and prototype; Authors: Jonathan Winterton, Françoise Delamare - Le Deist, Emma Stringfellow; final draft. January 2005.
- [CEN/ISSS, 05] CEN/ISSS Workshop ICT Skills Phase 2 (2005), CEN Workshop Agreement (CWA); source: http://www.uninfo.polito.it/WS_ICT-Skills/; last visit 01.04.2007.
- [Copenhagen, 02] The Copenhagen Declaration: Declaration of the European Ministers of Vocational Education and Training, and the European Commission, convened in Copenhagen on 29 and 30 November 2002, on enhanced European cooperation in vocational education and training; source: http://ec.europa.eu/education/copenhagen/copenahagen_declaration_en.pdf/; last visit 17.07.2007.
- [eCCO, 05] eCCO 2005: A European e-Skills Meta-Framework. Draft Proposal. AICA, Federcomin, Fondazione Politecnico di Milano. April 2005.
- [ESF, 04] ESF 2004: European e-Skills Forum 2004: e-Skills for Europe: Towards 2010 and Beyond, Final Synthesis Report, September 2004.
- [EUQuaSIT, 04] European Qualification Strategies in Information and Communications Technology (ICT), August 2004; source: <http://www.biat.uni-flensburg.de/euquasit/>; last visit: 19.07.2008.
- [HARMONISE, 07] HARMONISE is a Reference Material Project in the Leonardo da Vinci Programme and a project of CEPIS (Council of European Professional Informatics Societies) co-funded by the European Commission; source: <http://www.cepis-HARMONISE.org/>; last visit: 19.07.2007.
- [HRXML, 04] HR-XML Consortium: Competencies. 2004; source: <http://www.hr-xml.org/>; last visit: 13.04. 2007.
- [Marinoni, 05] Marinoni, C.: eCCO System: Towards a Shared Model from Job Profiles to Competences; source: http://eskills.cedefop.eu.int/eChallenges2005/6_Marinoni_e2005.pdf; last visit: 02.04.2007.
- [Povalej, Weiß, 07] Povalej, R.; Weiß, P.: Conceptual Frameworks for the Modelling of ICT Competence Profiles in eChallenges 2007, to be published October 2007.
- [RCD, 07] IEEE Learning Technology Standards Committee RCD: Draft standard for Reusable Competency Definitions (RCD); source: http://ieeeltsc.org/wg20Comp/Public/IEEE_1484.20.1.D3.pdf/; last visit: 13.04.2007.
- [Weiß, Povalej, Stucky, 05] Weiß, P; Povalej, R; Stucky, W.: Learning across Boundaries: Developments towards a European Framework for ICT Skills in The 5th IBIMA International Conference on Internet & Information Technology in Modern Organizations; December 13-15, 2005 Cairo, Egypt.
- [Weiß, Stucky, Dolan, Bumann, 05] Weiß, P.; Stucky, W.; Dolan, D; Bumann, P.: e-Skills Certification in Europe Towards Harmonisation in Innovation and the Knowledge Economy: Issues, Applications, Case Studies, Cunningham, P. and M. (Eds), IOS Press, Amsterdam, Challenges 2005 Proceedings, 2005.

Analyzing Wiki-based Networks to Improve Knowledge Processes in Organizations

Claudia Müller

(University of Potsdam, Germany
cmueller@wi.uni-potsdam.de)

Benedikt Meuthrath

(University of Potsdam, Germany
bmeuthrath@wi.uni-potsdam.de)

Abstract: Increasingly Wikis are used to support existing corporate knowledge processes. Wikis are an appropriate software solution to support these processes because of their specific design principles. However, it is not yet proven whether Wikis are an adequate knowledge management tool. This paper presents a new approach to analyze existing networks in Wikis. Four perspectives on Wiki networks are introduced to investigate all dynamic processes and their interrelationships in a Wiki information space. The Social Network Analysis (SNA) is used to uncover existing structures and temporal changes. Additionally, a data set of a corporate Wiki is used to present an analysis which has been carried out. Results of analyzes work can be utilized to improve the existing knowledge processes in a company.

Keywords: Social Software, Wiki, Knowledge Work, Network Analysis, Collaboration Network

Categories: A.1, H.0, H.4.3, H.4.m, J.3, K.4.2, K.4.3

1 Introduction

Formerly centralized hierarchical organizations are increasingly transforming to decentralized and network organizational constructs, which promote more openness. In these systems, which are subject to continuous change processes, people and available information are key resources [Tapscott, 93], [Barabási, 03]. Flat hierarchies and flexible structures, team orientation as well as a progressively growing dependency on knowledge are fundamental characteristics of companies today [Cross, 02]. These changes in organizational structures have also impacted corporate knowledge management. To ensure organizational learning, knowledge communities should be fostered [Romhardt, 02]. These knowledge networks within an organization, which are not organizationally authorized, were initially analyzed in the context of Communities of Practices (CoP) [Wenger, 98]. Information Technology (IT) is one important enabler for a successful knowledge community. In this context Wikis are increasingly used for decentralized, collaborative knowledge management of corporate knowledge communities [Klobas, 06], [Tapscott, 07]. The first Wiki, implemented by Ward Cunningham was designed based on various design principles [Leuf, 01]. These principles influence Wiki software, the user behavior and the application of knowledge management. A Wiki is “open”, implying that each

person (employee) can consume and edit every page. Information (externalized knowledge) in a Wiki is mergeable or dividable depending on existing knowledge requirements. Also, information context is arbitrarily adaptable by linking a specific article. When users externalize their knowledge they can link it to existing or non-existing articles (displayed in another link color). In the latter case, a Wiki enables users to formulate their existing knowledge needs. A Wiki information space has initially no predefined structure. The structuring of information objects (articles) is specified by a collaborative self-organized process controlled by the users. All article versions are saved in a history. A critical issue of knowledge work is solved in this way. Usually the final product of knowledge creation is visible but not the actual process of knowledge creation. With the page history the process from the first draft to the final version is traceable. Despite all known advantages it is still not proven whether Wikis are an adequate knowledge management tool to foster knowledge communities in organizations. This paper presents a novel approach to analyze existing networks in Wikis. The Social Network Analysis (SNA) is used to uncover existing structures and its temporal changes. An analysis of a corporate Wiki shows first results of this approach. Two hypotheses constitute the base for the presented research in progress:

- Social Network Analysis is an adequate method to analyze Wiki information spaces and
- The collaboration network shows the nature of cooperation in a Wiki and reveals special roles, like the Wiki-Champion.

This paper is organized as follows: the next section introduces social network analysis and possible application areas in knowledge management. Moreover, basic SNA metrics are explained shortly. In the following paragraph four perspectives of Wiki-specific networks and their characteristics are described. Finally one part of a conducted exploration of a collaboration network in a company is illustrated and results are evaluated. The third section gives an insight in future tasks.

2 Applying Social Network Analysis in Knowledge Management

Social network analysis (SNA) is an enabler to systematically describe and investigate network structures and their principles of order in organizations. Basically, a social network consists of a set of persons and relations defined in them [Wasserman, 97]. The social environment is described by a pattern or regularity among the interacting persons. Social relationships of persons constitute the social capital of an organization. The social capital metaphor is that people who do better are usually better connected. SNA is increasingly applied for analyzing and reorganizing distributed communication and collaboration structures in knowledge management (e.g. [Cross, 02]). Existing (cross-functional) relations are determined in their relation strength and intensity. Expertise within knowledge networks can be located.

The required data for network analysis can be gathered electronically or via interviews. A common approach utilizes user interaction with computer systems (e.g. Email, Blogs, Wikis) as applied in this contribution. Network analysts use matrices and the graph theory to represent information about patterns of ties among social actors. A network can be abstracted as a graph where nodes are described as vertices

and relations as edges. This graph is a model of a network with unidirectional dichotomic relations [Wasserman, 97]. The network analysis offers a range of metrics to analyze social structures in a systematic manner. Measures to investigate the structure of the network are for instance *density*, *average distance*, and *average degree*. These metrics are used here to describe the network structure on a macroscopic level.

Density of a graph is the ratio of relations over number of maximum possible relations [Wasserman, 97]. The highest possible number of edges is reached if each vertex of the graph is connected with each other vertex in the graph. The (geodesic) distance is defined as the length of a geodesic between two vertices. The average distance (average shortest path between all vertices) is calculated for the whole network. This metric is equal to the degree of separation which connects all pairs of nodes in a group. Short distances transmit information accurately and in a timely way. A long distance degree implies slower even distorted information transmission [Cross, 02]. The degree of a node is the number of edges that are connected to this node (adjacent nodes). This measure helps to evaluate the “attractiveness” of a node. The average degree is calculated for all nodes in the network. On a microscopic level, the position of a person in a network can be investigated with the metric *degree centrality*. This metric helps to identify the “most important” actor, meaning those that are extensively involved in relationships with other actors in the social network. A vertex is central in the sense of degree centrality, if a vertex has many relations to adjacent vertices.

All these measures can be utilized to investigate dynamic characteristics of networks as well. The dynamic perspective of network analysis deals with the progression of state changes over time and enables the observation of events and its analysis. To determine changes of any form in the network it is necessary to compare at least two different network states at different times. Dynamic network analysis (DNA) enables analyzing network states, the vertices and their edges as well as changes in structure and configuration of the network [Carley, 04]. The cumulative analysis enables the aggregation of a long time period of the network, including all changes and elements, which are or were part of the network during the period. The calculated measures and their changes are visualized in a diagram.

2.1 Specifying Wiki Networks

A Wiki with all related activities forms a specific information space in a company. This information space consists of different networks. Based on a classification made by Carley [Carley, 04], Wiki-specific networks can be arranged in four categories (Figure 1): social perspective (who knows who), knowledge perspective (who knows what), information perspective (what refers to what), and temporal perspective (what was done before). Relationships in one network usually imply relationships in another [Carley, 04]. Therefore, to understand all knowledge processes in a Wiki an analysis of these different network types is necessary.

The first group contains collaboration networks and discussion networks. Both types are social networks (cp. Figure 1). A *collaboration network* is used to investigate the nature and extend of collaboration between authors of Wiki articles and it enables the analysis of information exchange between communities. Collaboration is based on mutually referred asynchronous generated contributions in a

knowledge process. In a collaboration network vertices are authors, whereas the edges are constructed on common edits of an article. Whenever two or multiple authors worked on the same article in a Wiki, they are connected to each other. The central assumption is that changes of different authors on an article imply collaboration. A strong tie exists if two authors have contributed comparatively more than other authors (in terms of content) or they work collaboratively very often on one or more articles (in terms of frequency). By analyzing collaboration networks the time component has a high impact because cooperation is possible over a long period. Consequently, the effect of aging on network structures has to be considered [Zhu, 03]. With the second network in this perspective - the *discussion network* - a topic specific communication process on a Wiki talk page can be studied. Vertices are authors of specific postings and edges are related postings. The hierarchical order of these postings is defined by the time of edits. An activity level of discussions in a Wiki information space is measurable. The last network in this perspective is the *message exchange network*. Each registered user can exchange messages. This exchange is carried out by an edit on a personal talk page (like an entry in a guest book). Users get informed about new messages on their personal talk pages for instance via email. Nodes are users and relations exist based on exchanged messages in this network. This network supports analyzing and evaluating the amount and characteristics of communication processes in Wikis.

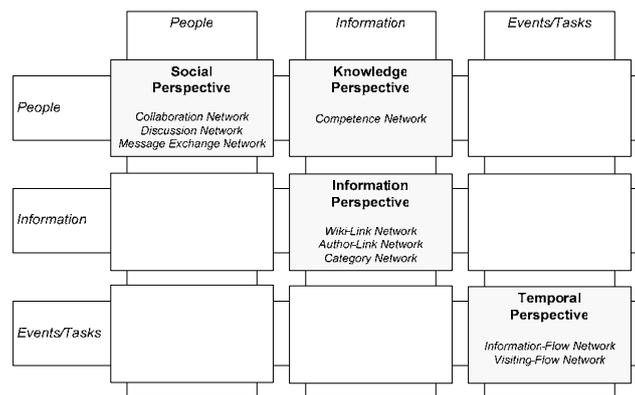


Figure 1: Specification of Wiki networks depending on different perspectives

The second perspective comprises of knowledge networks (cp. Figure 1). Wiki specific realizations are *competence networks*. This type of network shows existing thematic competences of authors. It is a two-mode network which means there are two different types of nodes – authors and articles. The edge between two vertices (article, author) represents the editing activity of a certain author for a given article. An investigation of a competence network gives an impression which author works on what article very often and/or has contributed a lot to one article. This is an index of existing knowledge in a specific topic area. A main assumption in this network is that if authors edit articles they externalize their knowledge as well. Minor changes, e.g. typographical errors, are left out of these considerations. An interpretation of

competence networks is due to their character more complicated than in the networks already introduced.

The information perspective defines information networks (cp. Figure 1). In the Wiki context there are wiki-linked networks, author-linked networks, and category networks. A *wiki-linked network* consists of articles as vertices and wiki-links as edges. It is a directed network since wiki-links are not reciprocal. The temporal analysis of wiki-linked networks enables the investigation of structural developments of themes in Wikis. The evolutionary process of emerging and “dying” topics is visible in this way. A structural analysis based on degree centrality shows highly connected areas in this network. Identifying topic clusters is another structural approach. Existing theme affiliations can be detected by text analysis. The second network in this perspective is the *author-linked network*. It focuses on articles with a high modification rate. The vertices are articles and edges represent changes on both articles by a certain author. The weight of such a relation increases with the number of authors which work on both articles. Opposite to the wiki-linked network, these network centers emerge based on thematic relations of authors and not of articles. Furthermore, this network describes the “attractiveness” of a single article in the information space. One disadvantage of this network is, that authors, who are active with one article only are neglected. Network usefulness decreases, if the number of such authors is high in the investigated information space. The third network is a simple hierarchical network – the *category network*. In Wiki information spaces content is often structured in categories. Besides a hierarchical structure of categorization there are also cross-links between the various category entries. One single category consists of thematic similar topics. The objective of this network is to describe these categories and their utilization (the more articles a category has the higher is the diameter of nodes). The category network comprises category pages as vertices and main-category/sub-category relationship as directed edges. Based on a temporal analysis the development of articles in specific categories can be studied.

The fourth perspective comprises information-flow and visiting-flow networks. Especially in specific topic areas, an analysis of the temporal development of articles and their interdependencies is interesting. The *information-flow network* sequences Wiki articles. The nodes are articles and the relations are defined based on the article history and/or the day of creation. Therefore, these relations show the temporal ordering of article creation or change. The measuring of an information flow from one article to the other is not trivial and time is not a sufficient parameter. Rather the content of articles has to be considered. This means a relation should be defined by transferred text from one article to the other. In the Wiki context it may be interesting to gather information about how the Wiki is used not only by its contributors but also by its readers. The *visiting-flow network* sequences Wiki pages by its temporal order by means of the referrer web server log. A referrer is the page a reader visits directly before the actual page. This information is sent by most browsers and is logged by web and proxy servers in default configuration. Assuming that there are two starting points (external and no referrer) the referrer information is sufficient to generate a directed network with Wiki pages as nodes and their referrer relationship as edges, weighted by its frequency. To clean up results editing and resource access (style sheets, java scripts etc.) should be filtered.

2.2 Analyzing and Evaluating Wiki Networks to Improve Knowledge Work

The predefined Wiki networks are applied to enhance existing network analysis approaches. Using these networks provides a comprehensive insight into existing knowledge processes in a Wiki in terms of people interaction, topic emergence, communication activities and their dynamic changes. Here, the collaboration network is visualized and analyzed to emphasize this new approach. The data set used is based on a Wiki that was launched in the summer of 2005. It serves as a knowledge exchange platform for about 1,000 employees. In April 2007 the corporate Wiki had 185 users and about 8,200 articles. To conduct these analyses a special tool was implemented. The SONIVIS:Tool provides statistical analyses of Social Software. It is based on the eclipse Rich Client Platform (RCP). There are predefined graphical analyses which offer a quick overview on recent developments. Additionally, various SNA metrics are provided on a microscopic and macroscopic level.

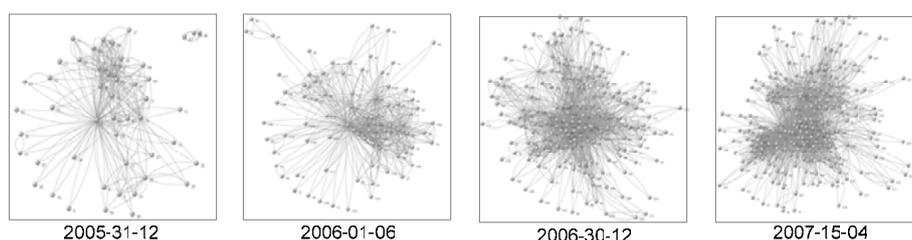


Figure 2: Temporal development of a collaboration network

Figure 2 yields a first insight into the development of the collaboration network. At the beginning of the investigated timeframe (2005-31-12) there is one specific area (sub-network) in which specific vertices are closer than the others. There is also one component with no connection to the main network. This picture changes over the investigated period (2006-01-06). The network is growing. More and more authors join the network and existing collaborations get stronger in certain areas. In the third snapshot (2006-30-12) there is one specific sub-network showing strong relations, but this group has grown compared to the year before. There are still a number of lightly connected nodes in the periphery of the collaboration network. There are maybe only a limited number of authors or they work only on specific articles in the Wiki. The last visualization (2007-15-04) shows three subgroups. A conducted analysis reveals that these subgroups have been formed around specific projects. Therefore, it will be interesting to examine how the Wiki structure is changing because of external, organizational changes.

A pure visualization of a network is not sufficient for understanding the whole evolutionary process. Hence, specific metrics should be used to analyze the network structure on a macroscopic and microscopic level. Figure 3 shows on the left the temporal development of the network's density and on the right side the temporal development of the average distance. Both measures allow an evaluation of the network on a macroscopic level. The density declines continuously during the considered period. However, Figure 2 shows that there are specific zones in the network with a density of almost 1 (so called cliques). The average distance increases

during the evaluated period. During the 2005-10-09 and 2006-10-04 period the network has two components (unconnected sub-graphs). Therefore, the average distance is infinite, in the diagram equal to zero. Overall, the average distance increases which is a result of the growing number of authors and a decrease of the average degree (number of realized connections). Consequently, the spread of information in this network needs more time. A correlation analysis was constitutively conducted based on these results to reveal existing relations between the collected measures.

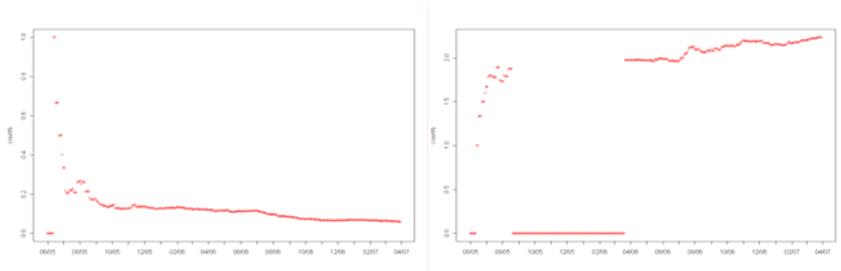


Figure 3: Density (left side) and average distance (right side)

A positive correlation between article count and average degree (0.85) and a negative correlation between density and average distance (-0.34) was determined. The first result is attributed to a minor collaboration in the whole network but a high collaboration in a subgroup. This is in accordance with the results of visualization (cp. Figure 2). A negative correlation implies that the average distance increased due to emerging subgroups. Simultaneously, the density decreases, but not at the same rate.

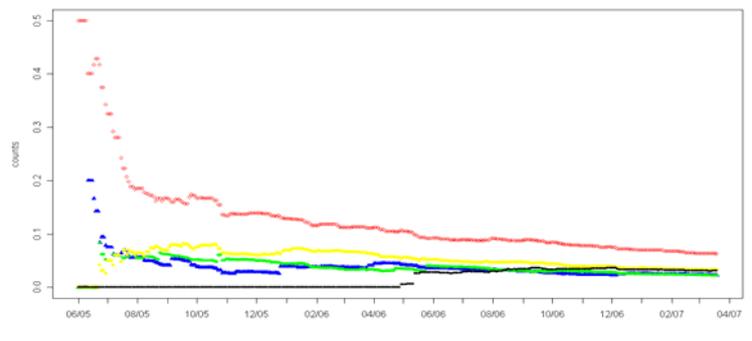


Figure 4: Temporal change of degree centrality on a microscopic level

This result is an indication for existing hubs in the network. Hubs are persons with a high level of activity because they are connected to a lot of persons through different articles. In early stages of a Wiki so called Wiki-Champions have a critical importance for the development of a Wiki information space. They are recognized as early adopters who understand how to use a Wiki very well, encourage others, and get

people involved by informally training them and being available for ongoing support. They serve as a model in using Wikis (<http://www.wikipatterns.com>). In this case one extremely active person can be identified in the beginning. As time moves on this person loses its special position gradually. But there is no reason for concern because the activity level in the Wiki is increasing overall. Also at the same time other persons gain importance in the network (cp. in Figure 4 the top curve). More and more people join the Wiki-Champion and engage more people in the Wiki. Based on these selected results the SNA enables the analysis of existing collaboration in a Wiki and reveals specific Wiki roles such as Wiki Champions.

3 Outlook

This paper shows possibilities how to analyze knowledge activities in terms of collaboration in a Wiki. The network analysis can be used to enable a measurement of an existing network. Due to the complexity of such systems four groups of networks are defined to analyze existing interdependencies. This work was done by a research group - the SONIVIS:Team (<http://www.sonivis.de>). In the future, all defined networks will be implemented to enhance analysis work and to improve our understanding of social processes in a Wiki information space.

References

- [Barabási, 03] A.-L. Barabási: "Linked"; Plume Printing, New York (2003)
- [Carley, 04] K. M. Carley: "Dynamic Network Analysis"; In: R. Brelger K. Carley, P. Pattison, (eds.): "Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers" National Academy Press, (2004), 133-145.
- [Cross, 02] R. L. Cross, A. Parker: "The Hidden Power of Social Networks: Understanding how work really gets done in organizations" Harvard Business School Press, Cambridge (2004)
- [Klobas, 06] J. Klobas: "Wikis: Tools for Information Work and Collaboration" Chandos Publishing, Oxford (2006)
- [Leuf, 01] B. Leuf, W. Cunningham: "The Wiki Way - Quick Collaboration on the Web" Addison-Wesley, New York (2001)
- [Romhardt, 02] K. Romhardt „Wissensgemeinschaften. Orte des lebendigen Wissensmanagements“ (in German), Versus Verlag Zürich (2002)
- [Tapscott, 07] D. Tapscott, A. D. Williams: "Wikinomics" Penguin Books Ltd., London (2007)
- [Tapscott, 93] D. Tapscott, A. Caston: "Paradigm Shift: The new promise of information technology" McGraw-Hill Companies New York (1993)
- [Wasserman, 97] S. Wasserman, K. Faust: "Social network analysis: methods and applications" Cambridge Univ. Press, Cambridge (1997)
- [Wenger, 98] E. Wenger: "Communities of Practice: Learning, Meaning, and Identity" Cambridge University Press, New York (1998)
- [Zhu, 03] H. Zhu; X. Wang, J. Zhu: "Effect of aging on network structure" Phys. Rev. E, American Physical Society, 68:056121 (2003)

Management, Interchange and Reproducible Execution of Sample Preparation Knowledge in Collaborative Research Scenarios

**Christopher H. P. Durst, Frank R. Ihmig, Stephen G. Shirley and Heiko
Zimmermann**

(Department for Cryobiophysics and Cryotechnology, Fraunhofer-Institute for Biomedical
Engineering (IBMT), Ensheimer Straße 48, 66386 St. Ingbert, Germany
{christopher.durst, frank.ihmig, stephen.shirley, heiko.zimmermann}@ibmt.fraunhofer.de)

Abstract: Biobanks store alive cell samples to provide them for a later use in many biomedical applications. Amongst many others, cord blood samples are cryopreserved (long-term stored between -130°C and -196°C) today for future stem cell therapy or tissue engineering. Different cell types need to be treated differently. Therefore, biomedical laboratories and biobanks have to cope with a variety of specialized knowledge for sample preparation, analysis or storage while they may lack appropriate techniques for reliably managing this specific knowledge or for executing preparations reproducibly. Especially in scenarios which require sample exchange or comparing of results (as for example collaborative vaccine studies do), problems can arise. This paper outlines the evolution of a local infrastructure for long-term management of sample preparation knowledge to a system for globally managing preparation protocols and their reproducible execution within a network of collaborating laboratories and biobanks, aiming at the collective process of knowledge-building. The system enhancements are inspired by the requirements of collaborative research and complex sample preparation protocols, derived from the exemplary needs of the GHRC biobank (Global HIV Vaccine Research Cryorepository), being part of the Collaboration for AIDS Vaccine Discovery.

Keywords: biobank; biomedical; cryopreservation; RFID; flash memory; sample preparation; protocol; workflow management; cell sample; low-temperature; HIV; AIDS

Categories: H.3.2, C.3, J.3, H.1.2

1 Introduction

Biological and biomedical science have brought a wide range of biomedical applications into being. Many of them require samples from earlier points of time, for example stem cell therapy, tissue engineering or retrospective diagnostics. Today, most existing biobanks use cryopreservation (freezing samples and storing them at temperatures between -130°C and -196°C , using liquid nitrogen). Knowledge about freezing, analysing or preparing particular cell samples (in the following referred to as “protocols”) is highly specialized expert knowledge and depends on cell type and application. Thus, biobanks and biomedical laboratories have to manage and perform a high number of different protocols. Since the storage of most cell samples is only intended for medical applications like transfusion, transplantation or stem cell therapy and not for deriving comprehensive scientific results, knowledge about samples and their assigned protocols is not shared among many institutions. It only needs to be

provided locally in a long-term save and reliable way. In [Durst 06] we have introduced a method and infrastructure for long-term management of sample preparation knowledge for cryo-biomedical applications, especially tailored to suit the needs for interaction between biobanks and biomedical laboratories or physicians, as described above.

However, there are biomedical research areas which require much more interaction and sharing of samples and knowledge among many collaborating institutions. Those are areas where scientific results need to be derived from comparative evaluations or from distributed and iterative experimental work. Such a collaborative scenario is given at the Collaboration for AIDS Vaccine Discovery (CAVD) project, supporting the GHAVE (Global HIV/AIDS Vaccine Enterprise) [GHAVE 05]. CAVD is a network of eleven HIV vaccine discovery consortia and five central service facilities – the 89 involved institutions are spread over 22 countries worldwide. One of these institutions is the Fraunhofer-IBMT, coordinating a network which is developing the GHRC (Global HIV Vaccine Research Cryorepository), one of the central service facilities [IBMT 06]. GHRC will be the biobank of CAVD for large-scale cryopreservation of HIV infected cell samples (and for providing them to other CAVD institutions), taken from patients at so-called primary sites in countries with high HIV infection rates. Various kinds of knowledge will be created worldwide, and there are different levels of knowledge management to be considered, e.g. between the GHRC and its primary sites or between GHRC and statistical centres or between other consortia and their sub-networks. Although all these levels are worthy of being addressed in detail, this paper focuses on the management, interchange and execution of sample preparation knowledge, being the basis for any knowledge derived from preparation results. But before, creation of sample preparation knowledge will be discussed shortly.

Primary Sites \leftrightarrow GHRC \leftrightarrow Consortia

Figure 1: An exemplary (sample and) knowledge chain

This paper distinguishes between explicit knowledge and tacit knowledge according to the definitions given in [Polanyi 66]. It also refers to the SECI model [Nonaka 95] and its components for transferring explicit and tacit knowledge into each other following the spiral of knowledge. In general, preparation knowledge is created empirically by biomedical experts from iterative experiments. They modify preparation steps of known preparation protocols (explicit knowledge) and/or their parameters. Therefore, they use their individual, tacit knowledge while performing protocols or combine explicit preparation knowledge to obtain potentially useful modifications. By such experiments, the scientist creates individual knowledge which needs to be transferred to the scientific community for enhancing its knowledge base and to allow further protocol optimization or standardization. An important basis for the externalization process is the preparation documentation as prescribed by GMP (“good manufacturing practice”) and GLP (“good laboratory practice”) standards. The documentation contains complex, detailed expert knowledge on performed preparation steps, on parameters, on results and on preparation success. Its

externalization leads to improved preparation protocols. Executing these protocols enhances the individual knowledge base and capabilities of scientist in a learning-by-doing manner (internalization), being starting point for a next iteration in the process for creating sample preparation knowledge.

2 Managing sample-assigned knowledge in collaborative networks

In GHRC, an HIV infected cell sample, for example whole blood, must be split into different products which are results of different preparation protocols, each. A set of diagnostic data is formed by the results of different analytic protocols. There is a much larger variety of knowledge, indeed, regarding the need for optimized freezing [Disis 06] and thawing protocols which directly influence on the proper function (in detail: on the cellular immune response to a vaccine) of a cell sample [Betensky 00], [Weinberg 00], and thus, on the result of vaccine studies. The large number of protocols needed within a collaborative research network and the frequent sample exchange between network members may cause big problems: there is a high risk of mistaking samples and associated protocols. Such a mistake leads to mis-preparation which might falsify scientific results while wasting valuable samples. In biomedical applications different from vaccine research, mis-prepared samples (or simply mistaken samples) even might be lethal for patients. While methods for comparing sample data from different biobank databases already exist [Ölund 07], the laboratories lack techniques for reliable management and interchange of samples and assigned protocols. Mistaking can happen because sample tubes are usually identified by labels (sticker, barcode, RFID [Bettendorf 05]) which do not contain any more information than referring to associated sample data or protocols [Pearson 04]. This information is stored in databases or is paper-written, apart from the sample. Consequently, a sample becomes useless (and the assignment of a protocol becomes unreliable) whenever its label is lost or damaged, or when there is database corruption. Long storage times typical to cryopreservation increase this risk.

In [Durst 06] and [Durst 06a] we presented a technique invented by Fraunhofer-IBMT to ensure reliable sample-knowledge assignment: the physical coupling of sample and electronic memory chip. This method enhances the organizational memory of biobanks by decentral and travelling knowledge containers carrying information relevant to a particular sample. This method allows reliable management and interchange of protocols and other knowledge about the sample. Progress in low-temperature electronics allows flash memory to work reliably even under cryo conditions [Ihmig 03], [Ihmig 04], [Zimmermann 04], [Ihmig 06], [Ihmig 06a]. This technique was used for our presented infrastructure and can be transferred for globally managing preparation protocols within a network of collaborating laboratories and biobanks.

3 Reproducible protocol execution and workflow management

A reliable assignment between cell samples and protocols, as discussed in section 2, will avoid mis-preparations of samples, even in a worldwide network of biobanks and

laboratories. However, differences in execution of identical protocols can occur in different laboratories or with different staff performing the preparation. This is because people might have different tacit knowledge of the how-to of performing particular preparation steps.

In collaborative scenarios this may lead to problems: for instance, in the research for an HIV vaccine the comparative evaluation of vaccine studies is required. Whenever the execution of the same protocol varies between different laboratories, preparations may result in samples having incomparable cellular immune response capabilities and thus, lead to incomparable vaccine studies. Therefore, uniform protocol execution is required among all laboratories within a collaborative network. Precondition for that is the use of standardized, detailed preparation protocols which provide enough explicit preparation knowledge to the lab employees so that people will not need their different tacit knowledge. Since preparation protocols are complex and there are many constraints regarding preparation execution, a workflow management system must process the protocols and support people by supplying the required knowledge for a particular preparation step exactly where and when this knowledge is needed. Such an intelligent knowledge distribution disburdens people and provides uniform protocol execution.

The infrastructure in [Durst 06] provided a simple, high-throughput capable workflow management function. Its principles are reminded here:

- Protocols are stored on a cryo-tolerant electronic memory chip. This chip is attached to the sample (see section 2).
- Cell samples control their preparation themselves by supplying the assigned protocols to the lab infrastructure. A workflow engine controls the execution of each preparation step, covering manual, semi-automated and fully automated preparation steps.
- Sample knowledge (results, documentation) is acquired and stored at-sample on-chip.

Basically, these principles will serve to collaborating networks as well. However, the workflow engine and protocol language used in the presented infrastructure is not sufficient for the needs in collaborative scenarios. The next section will discuss the limitations of the existing system and introduce the improvements which are currently under construction.

4 Limitations and improvements: enhancing the system for collaborative scenarios

In the prototypic implementation shown in [Durst 06], comparative evaluations between different laboratories were not addressed. Biobanks and laboratories were supposed to perform non-identical protocols and to complement one another by division of work. The existing system can only locally provide uniform protocol execution within the same laboratory (or within identically equipped laboratories) since protocols directly address devices or users. They cannot be transferred between differently equipped laboratories without complex reassignments. However, there is

no homogenous laboratory equipment among collaborating laboratories – in contrast, each laboratory has an individual set of devices. This set will change over time: new devices will be added, defective ones removed. Whereas in some laboratories there are devices for particular preparation steps, in other laboratories there are not, and vice versa. Thus, laboratories have their individual set of functionalities and restrictions. The protocols need to be defined independent from any particular laboratory. They must not contain any references to specific devices but to device classes. Therefore, the protocol language must allow abstract device assignments, being mapped by the system to available devices of a particular laboratory just before protocol execution. Then, abstract instructions can be translated into instruction streams for particular devices or humans, depending on different filtering parameters like device availability, load balancing, required range, accuracy, sample volume, and so on. The system architecture needs to be extended by a layer which provides device surveillance and filtering function for dynamic addressing, based on detailed knowledge of the specifications of available devices.

In the requirements analysis for the existing system, representative biomedical preparation protocols for therapy-related sample preparation had been analysed: protocols turned out to be straight, without the need for control structures, conditions or different event-driven branches. Therefore, the existing system can only control a purely sequential protocol execution from start to finish in a high-throughput manner, performing only one preparation step at a time. Analysing typical protocols needed for GHRC shows a much higher complexity: HIV infected cell samples are prepared into different products (mononuclear cells, virus culture, serum and more), requiring multithreading of different preparation branches and parallel execution of preparation steps. Since particular protocol branches depend on results of others, protocol language and workflow management system additionally need to support conditions, global variables and other control structures and thus, finite state machine workflow management.

5 Conclusions and outlook

Basically, the principles of the infrastructure introduced in [Durst 06] can be used for management, interchange and reproducible execution of sample preparation knowledge in collaborative research networks. However, complex protocols found in collaborative research require a more powerful protocol language, exceeding the language capabilities of the existing system by far. For reproducible protocol execution, the language must be laboratory-independent for being compatible with differently equipped laboratories. Therefore, enhancements to the architecture of the workflow management system are required to execute such complex protocols and to map abstract device assignments to particular devices. A different workflow engine needs to be used for supporting this powerful language. A prototype of the enhanced system is planned to be used from end of summer in GHRC which is being established in the “Eurocryo” cellbank of the Fraunhofer Society in Sulzbach/Saar.

Acknowledgements

We thank the Bill&Melinda Gates Foundation for the grant given to Fraunhofer-IBMT and to GHRC, and the Ministry of Economic Affairs of Saarland, Germany.

References

- [Betensky 00] Betensky, R.A., Connick, E., Devers, J., Landay, A.L., Nokta, M., Plaeger, S., Rosenblatt, H., Schmitz, J.L., Valentine, F., Wara, D., Weinberg, A., Lederman, H.M.: "Shipment impairs lymphocyte proliferative responses to microbial antigens"; *Clin. Diagn. Lab. Immunol.* 7 (2000), 759.
- [Bettendorf 05] Bettendorf, E., Malenfant, C. et al.: "RFID Technology and Electronic Tags to Identify Cryopreserved Materials"; *Cell Preservation Technology* 3, 2 (2005), 112-114.
- [Disis 06] Disis, M.L., dela Rosa, C., Goodell, V., Kuan, L.Y., Chang, J.C., Kuus-Reichel, K., Clay, T.M., Kim Lyerly, H., Bhatia, S., Ghanekar, S.A., Maino, V.C., Maecker, H.T.: "Maximizing the retention of antigen specific lymphocyte function after cryopreservation"; *Journal of Immunological Methods*, 308, 1-2 (2006), 13-18.
- [Durst 06] Durst, C.H.P., Ihmig, F.R., Fuchs, C.C., Zimmermann, H.: "A method and infrastructure for long-term management of sample preparation knowledge for cryo-biomedical applications"; *Proc. BPOKI'06*, Graz, Austria, 2006.
- [Durst 06a] Durst, C.H.P., Ihmig, F.R., Biel, M., Daffertshofer, M., Zimmermann, H.: "A method and technology for reliable sample-controlled execution of preparation and freezing protocols in biomedical laboratories and cryobanks"; *Proc. 43rd meeting of the Society for Cryobiology in Association with the Society for Low Temperature Biology*, Hamburg (2006), 152.
- [GHAVE 05] The Global HIV/AIDS Vaccine Enterprise: "Scientific Strategic Plan"; *Coordinating Committee of the Global HIV/AIDS Vaccine Enterprise, PLoS Med.*, 2, 2 (2005), e25.
- [IBMT 06] Fraunhofer IBMT: "Presseinformation: Neue Unterstützung für die europäische und deutsche HIV-Forschung - Zentrale Kryobank finanziert durch die Bill & Melinda Gates-Stiftung geht nach Deutschland"; (2006); (http://www.ibmt.fhg.de/fhg/Images/FraunhoferIBMT_BMGatesFoundation_deutsch_tcm266-70018.pdf)
- [Ihmig 03] Ihmig, F.R., Shirley, S.G., Zimmermann, H.: "Evaluation and Adaption of Flash-Memory for Cryobiophysical Applications"; *Proc. 2nd VDE World Microtechnologies Congress*, Munich, Germany (2003), 643-648.
- [Ihmig 04] Ihmig, F.R., Shirley, S.G., Zimmermann, H.: "Electronic Memory Devices for Cryobiological Applications"; *Proc. 6th European Workshop on Low Temperature Electronics, WOLTE-6, ESTEC, Noordwijk, The Netherlands* (2004), 153-160.
- [Ihmig 06] Ihmig, F.R., Shirley, S.G., Durst, C.H.P., Zimmermann, H.: "Cryogenic electronic memory infrastructure for physically related „continuity of care records” of frozen cells"; *Cryogenics*, 46, 4 (2006), 312-320.
- [Ihmig 06a] Ihmig, F.R.: "Zur Funktion, Integration und Zuverlässigkeit elektronischer Bauelemente und Baugruppen in kryobiologischen Systemen"; *PhD thesis, Universität des Saarlandes*, 2006.

[Nonaka (95)] Nonaka, I., Takeuchi, H.: "The knowledge-creating company"; Oxford, UK, University Press, 1995.

[Ölund 07] Ölund, G., Lindqvist, P., Litton, J-E.: "BIMS: An information management system for biobanking in the 21st century"; IBM Systems Journal, 46, 1 (2007), 171-182.

[Pearson 04] Pearson H.: "Summit calls for clear view of deposits in all biobanks"; Nature, 432, 7016 (2004), 426.

[Polanyi (66)] Polanyi M.: "The tacit dimension"; London, UK, Routledge & Kegan Paul, 1966.

[Weinberg 00] Weinberg, A., Wohl, D.A., Brown, D.G., Pott, G.B., Zhang, L., Ray, M.G., van der Horst, C.: "Effect of cryopreservation on measurement of cytomegalovirus-specific cellular immune responses in HIV-infected patients"; J. Acquir. Immune Defic. Syndr. 25 (2000), 109.

[Zimmermann 04] Zimmermann, H., Ihmig, F.R., Shirley, S.G.: "A low cost stage for testing electronics between room and liquid nitrogen temperatures"; Proc. 6th European Workshop on low Temperature Electronics, WOLTE-6, ESTEC, Noordwijk, The Netherlands (2004), 287-293.

Flexible Office: Assignment of Office Space to Enhance Knowledge Work Productivity

Ronald Maier, Stefan Thalmann

(University of Innsbruck, Austria
ronald.maier@uibk.ac.at, stefan.thalmann@uibk.ac.at)

Florian Bayer

(Martin-Luther-University Halle-Wittenberg, Germany
florian.bayer@wiwi.uni-halle.de)

Michael Krüger, Hendrik Nitz, Alexander Sandow

(GISA GmbH, Halle/Saale, Germany)

Abstract: Even though knowledge management has been around for more than a decade, so far concrete instruments that can be systematically deployed are still rare. This paper presents an optimization solution targeted at flexible management of office space considering knowledge management criteria in order to enhance knowledge work productivity. The paper presents the Flexible Office conceptual model and optimization solution and discusses the theoretical foundation, assumptions and reasoning. A corresponding prototype was field-tested, successfully introduced, evaluated with the help of a series of interviews with users and improved according to their requirements. The paper also reflects on the organisational impact and lessons learned from field test and practical experience.

Keywords: flexibility, hypertext organization, knowledge management, knowledge work, office space, optimization

Categories: Category G.2.3, H1.2, H4.1 H.5.3, K.4.3

1 Introduction

Knowledge management (KM) has been around for about twenty years since its inception by [Sveiby, 87] and [Wiig, 88] and a number of KM instruments have been developed which can be applied for systematic interventions into the way an organization handles knowledge [Maier, 07, 195]. Examples are communities and knowledge networks, competence management, lessons learned, best practices or semantic content management. Many organizations have built an organizational and a technical infrastructure in order to support KM-oriented tasks, such as creating, sharing, applying and retaining knowledge. The organizational infrastructure involves establishing new roles, such as subject matter expert, or fostering informal groups, such as communities. The technical infrastructure provides advanced discovery, publication, collaboration and learning services for a knowledge worker's (KW) work place, sometimes called semantic desktop [Sauermaun, 05].

However advanced these infrastructures are, many of the more human-oriented KM protagonists keep stressing the importance of face-to-face meetings for

exchanging the most relevant knowledge which is a combination of implicit and explicit knowledge that can only be handed on in a process of socialization. Consequently, it seems promising to apply KM-oriented criteria to flexibly assign office space to KWs in order to provide them with opportunities to meet those co-workers who are currently engaged in the same processes or projects, who work on the same themes or who advance the same competences. Building on an existing KM infrastructure that provides easily accessible data on KM-relevant criteria such as KWs' skill levels, organizational affiliation, projects or themes, this paper presents the Flexible Office solution, which has been developed with and implemented at GISA, a medium-sized IT company.

Section 2 discusses the main reasons motivating the solution. Section 3 presents the Flexible Office concept and the optimization solution. Section 4 gives details on implementation and reports on practical experiences and feedback given by KWs affected by the solution. Section 5 finally concludes the paper.

2 Motivation and point of departure

The Flexible Office project was motivated by the following main observations:

Mobility: KWs increasingly work outside their offices, e.g., at their customers' offices, on the road or at home. In the project, the average percentages of time spent outside the company were determined for all organizational units. It turned out that in one unit, KWs spent on average almost 30% of their working time outside the company with a minimum of 14% and a maximum of 55%. If one additionally considers the time spent in meeting rooms or bureaus of colleagues, these quotas are substantially higher. This organizational unit was therefore chosen for the pilot study of the Flexible Office project. Economically, the high portion of time spent outside the company leads to many empty offices and thus to inefficiencies in usage of office space. More efficient use of office space could allow for growth without the need to rent additional office space. From a KM perspective, distribution of KWs over a number of offices inside and outside the company leads to inefficiencies in communication and knowledge sharing.

Project orientation: Office structures at the company reflect the traditional organizational structure and thus are arranged according to the organizational units established in the business system. Typical for an IT company, projects play an important role and therefore they need to be carefully considered. GISA is characterized by a multitude of projects that span organizational units and employees working on a number of projects at the same time. Both, project managers and project team members suffer from the team being spread over a number of offices and would profit from the possibility to reserve a room for team members for a certain amount of time, e.g., for a project kick-off, for preparation of a milestone result or report, for finalizing a project or for documenting lessons learned.

Knowledge management: The increasing velocity, with which new products and services are created, in this case standard and individualized software products as well as consulting services, leads to an also increasing importance of knowledge. This means that KWs improve their competencies, are engaged in learning activities and co-develop themes that span organizational units and project teams. Flexible offices systematically take into account the themes which KWs develop that can hopefully be

turned into successful projects in the future. As a consequence, workplace learning, knowledge transfer between KWs working on the same theme as well as training of KWs new to the job or the theme might be improved with such a solution.

The three observations can be merged by the model of the hypertext organization developed by [Nonaka, 94]. In this perspective, the well-known metaphor of a hypertext document is used to denote the synthesis of traditional hierarchical organizational structures with non-hierarchical, self-organizing structures in order to combine efficiency and stability of the hierarchy with dynamism and flexibility of cross-functional task forces. The hypertext organization consists of three layers:

business system layer: performs business processes and routine operations typically organized in a traditional hierarchy,

knowledge base layer: comprises knowledge resources, themes and competences embedded in people, procedures, documents and the organizational culture,

project team layer: comprises self-organizing project teams pursuing time- and resource-constrained goals.

These three layers form a hyper-linked network. In this view, KWs are engaged in the business systems layer and handle topics as well as use competencies from the knowledge base layer. Furthermore, nearly all KWs are engaged in projects. Office space is typically assigned to KWs using only one of these dimensions, mostly the business system dimension. However, KWs are engaged in the three layers in parallel and shift their attention flexibly between these layers. In order to support this, the Flexible Office solution considers all dimensions of the hypertext organisation for the assignment of office space and thus for the provision of opportunities for interaction for KWs with the help of a set of optimization variables.

3 Flexible Office Concept

The Flexible Office conceptual solution consists of a technical and an organisational component. *Organizational issues* concern selecting employees, designing office space, analysing requirements concerning equipment, defining rules as well as assigning roles. The selection of employees who participate in the Flexible Office solution and thus share their work places is limited by absence quotas and by factors such as sensitivity of the tasks, e.g., daily use of HR records. Also, willingness to participate had to be ensured. This was supported by several meetings communicating the benefits of the approach and identifying prejudices, fears and expected barriers. A pilot group consisting of KWs, project managers and external consultants was set up to test a flexible arrangement of work places. This group also specified the requirements concerning office space, equipment, organisational rules and roles.

The required office space was selected and designed according to the preferences of the pilot group. This included furniture and specific office equipment, e.g., beamers and whiteboards, supporting communicative tasks. Concerning technical equipment, Flexible Office members were set up with mobile phones, blackberries, UMTS network access for laptops as well as a remote access solution and a virtual private network for home office or customer office access to company servers.

Moreover, organisational rules were defined that specify for example the timeframe and deadline in which rooms or single work places can be booked, mechanisms for escalation of urgent bookings or conflicts between projects,

requirements concerning distribution of mail and organization of central project repositories. Additionally, a clean desk policy was established demanding that no personal items are left on the work place in order to ensure unhindered usage by several employees. In order to ensure compliance to the rules, a separate role “Flexible Office manager” was introduced to resolve minor questions and problems, forward problems and escalate conflicts and claims of the participating employees.

The *technical component* is a hotelling software that considers the three main areas of requirements mobility, project orientation and KM described in section 2. Differentials to standard hotelling software are that assignments of work places consider KM-oriented criteria such as project and theme overlappings between KWs, their preferences and reservations by project managers. These criteria supposedly lead to improved communication and coordination in projects, decreased search time, improved knowledge transfer, workplace learning and improved handover of projects between project teams and organizational units responsible for operation and maintenance of the resulting application systems. The hotelling solution was integrated into the B2E (Business to Employee) information infrastructure in the form of an employee self service portal. The software solution consisted of

- user interface for project managers to reserve office space for their projects,
- user interface for KWs to submit their preferences, to apply for home office days and for those work spaces that have not been assigned automatically,
- the core optimization component for the assignment of rooms,
- output components for visualizing the solution in the B2E portal and for notifying KWs per email about the booked rooms.

In the following, the core optimization component is described in some detail. The booking process determines the optimal assignment of work spaces according to the predefined criteria for one work week. Reservations and preferences have to be submitted until Thursday evening in the week proceeding the booking week. The results are forwarded to KWs on Friday noon.

Criteria have been quantified and the optimization problem has been formalized with the help of standard operations research methods. The utility function (score) that is optimized consists of a number of weighted factors (described below).

Factors regarding the project team layer

reservations by project managers: Project managers can reserve a room for one or more KWs of a certain project. In case of two competing reservations, the demanded room will be assigned to the project with the higher priority. Due to hierarchical legitimation, reservations by project managers are treated separately as a kind of “K.O.”-criterion.

attractive rooms for important projects: Rooms are valued according to the attractiveness estimated by KWs on a scale from one, i.e. very unattractive, to ten, i.e. very attractive. A project score consists of a project category reflecting the importance of the project and its customer as well as a time-variant score dependent on the state of activity of the project. These two parts give a project score between 1, i.e. less important project in standard mode, and 9, i.e. very important project in a “hot” phase. Multiplying both leads to results in which attractive rooms are assigned to important, currently highly active projects.

project overlappings: This criterion values the relationships between KWs with respect to their work in projects. Goal is to assign those KWs to the same room who are members of the same projects in as many cases as possible. Also, KWs can submit a project preference stating that it is this project that they will be working on mostly in the booking time frame. This means that overlappings are exclusively considered with respect to the preferred project. If there are no project preferences, the following formula calculates project overlappings po between project team members a and b : ps_{ai} is 0 if KW a is not on project i and is the project's score if a is on project i . $ps_{ai \wedge bi}$ is the project's score if KWs a and b are on project i and 0 otherwise.

$$po_{ab} = \left(\frac{\sum_{i \in P} ps_{ai \wedge bi}}{\sum_{i \in P} ps_{ai}} + \frac{\sum_{i \in P} ps_{ai \wedge bi}}{\sum_{i \in P} ps_{bi}} \right) / 2$$

Formula 1: project overlappings

Project overlappings are only considered if $po_{ab} > 0.6$ because they are only thought to be relevant if there are sufficient and sufficiently important projects that KWs share.

Factors regarding the knowledge base layer:

theme overlappings: Similar to project overlappings, theme overlappings consider the relationship between two KWs according to the themes that they are working on. The assumption behind this is that KWs working on similar themes should be assigned to the same room in order to improve knowledge sharing. Again, KWs can submit theme preferences which in this case means that they would like to sit in a room with a person that has the same or a higher skill level with respect to the preferred theme. Overlappings then are exclusively considered for the preferred theme. In all other cases, theme overlappings to between KWs a and b are calculated according to the following formula:

$$to_{ab} = \left(\frac{\sum_{i \in P} th_{(a \wedge b)i}}{\sum_{i \in P} th_{ai}} + \frac{\sum_{i \in P} th_{(a \wedge b)i}}{\sum_{i \in P} th_{bi}} \right) / 2$$

Formula 2: theme overlappings

th_{ai} is 0 if KW a does not work on theme i and is 1 if a works on theme i . $th_{(a \wedge b)i}$ is 1 if KWs a and b both work on theme i and 0 otherwise. Theme overlappings are only considered if KWs have an equal skill level or if a has a lower and b a higher skill level, but not the other way round.

Factors regarding the business system layer

work group overlappings: KWs can submit a preference for a certain work group meaning they wish to work with members of this group. The score for work

group overlappings wgo reflects the number of KWs in the assigned room that belong to the preferred work group.

moving costs and room preferences: The selected KWs have strong preferences for stability if changes are not too significant. This is why fictive moving costs have been introduced, so that small differences between criteria do not result in a large number of moves between offices without much effect on the utility function. Moving costs also consider room preferences that KWs have submitted. KWs can submit a preference for a type of room, e.g., a single office, a room with specific equipment, e.g., a beamer. If the new solution means a move into a room that the KW prefers, then there are no moving costs calculated. If the KW has to move out of a preferred room, moving costs are higher than in the standard case of no specific preferences for rooms.

The optimization problem is solved in two steps. In a first step, the following utility function is maximized in order to get a quick solution that considers the exclusive reservations by project managers. The mathematical problem can be solved with the simplex algorithm. The indices i and j in the two summarizing functions determine the matrix holding the decision variable X_{ij} meaning that x KWs of project j are assigned to room i . The only criteria considered in the utility function are the weighted multiplication of room attractiveness ra and project score ps , from which moving costs mc are subtracted. Thus, the utility function can be written as follows:

$$U = \sum_{(i \in R)(j \in P)} (X_{ij} \times (\alpha \times ra_i \times ps_j - \beta \times mc_{ij})) \rightarrow MAX$$

Formula 3: Utility function for the first optimization problem

Constraints are as follows: elements of the decision variable have to be positive integers, each room has a limited capacity, no more than the number of KWs that have been ordered by project managers are assigned to rooms and projects requested as exclusive do not have to share rooms with other projects.

The second step considers all KWs and rooms that have not been exclusively assigned in the first step. The weights of the criteria have been refined in a dozen rounds according to the preferences of the KWs participating in the pilot study. The quadratic mathematical problem can be solved with a branch and bound algorithm. The utility function consists of two terms. The first term reflects a matrix of rooms and KWs and the decision variable represents the Boolean assignment of KW j to room i with 1 for assigned and 0 for not assigned. With this term, room attractiveness ra is maximized and moving costs mc are minimized. The second term reflects a three-dimensional matrix of rooms r and the relationships between KWs a and b . Thus, the decision variable is 1 if the corresponding two KWs are assigned to the respective room and 0 otherwise. The term reflects the weighted criteria project overlappings po , theme overlappings to and work group overlappings wgo which have been explained in Table 1. The utility function can be written as follows:

$$U = \sum_{(i \in R)} \sum_{(j \in E)} (X_{ij} \times (\alpha \times r a_i - \beta \times m c_{ij})) + \sum_{(r \in R)} \sum_{(a \in E)} \sum_{(b \in E)} (X_{ra} \times X_{rb} \times (\gamma \times p o_{ab} + \delta \times t o_{ab} + \varepsilon \times w g o_{ab})) \rightarrow MAX$$

Formula 4: Utility function for the second optimization problem

Constraints are as follows: elements of the decision variable have to be Boolean, each KW is only assigned to one room and each room has a limited capacity.

4 Realization and Experiences

The Flexible Office project was carried out in two phases: (1) a feasibility study and a conceptual plan and (2) IT implementation and a pilot study to test the software. In a first step, the situation at GISA was studied in order to determine the sharing ratio, i.e. the number of KWs divided by the number of work places. The investigation included

- literature analysis of relevant case studies,
- analysis of documents, e.g., floor plans, organizational structure diagrams, project management handbook,
- reports on times of absence, e.g., travel, holiday and home office days,
- self-reporting in a more detailed way with five KWs compiling time allocated to projects and customers as well as time spent on the work place, in other offices, meeting rooms, customers' offices etc.,
- personal interviews that helped to refine the information gathered above. The collected data was used to determine the organizational unit that would be the first to profit from Flexible Office (a unit with more than 80% of project work), the sharing ratio (1.2)^[1] as well as several rules, e.g., clean desk policy, or limitations for booking a single work place.

The conceptual plan was then realized as a prototype software solution in the second part of the Flexible Office project. Seven GISA projects, 33 team members and nine rooms were selected for the pilot study. These KWs took over ownership of the Flexible Office project and closely and actively participated in the effort to refine both, the organizational and the technical part of the solution.

From an organizational perspective, success of introducing a flexible assignment of office space requires extensive communication of benefits and strong involvement of the participating employees. Especially identification of their prejudices and fears is important since a flexible assignment goes along with substantial change processes.

During the pilot study, all participating KWs were asked to fill out short online questionnaires and project managers were interviewed on a regular basis. Results of this study show a typical u-shaped curve concerning user satisfaction with the solution. It started out with high hopes, then some problems with the prototype and also the criteria that had not yet been sufficiently refined led to a decline in satisfaction. However, in the last three weeks of the pilot study, the curves reflecting

^[1] This was the most popular sharing ratio found in the literature. This is due to the consideration that it is not cost savings, but KM-related goals that are of primary interest here.

usability, improvements in communication, efficiency, learning and knowledge transfer all showed a positive tendency. In addition to the question about user satisfaction, seven statements had to be answered on a weekly basis over the project's life. Agreement ^[2] refers to the average of all (n= 33) respondents (see Table 1).

Respondents estimated COMM (significance level: 0,0017), NET (0,0127) and KEX (0,0099) significantly higher than SEARCH on a significance level of $\alpha \leq 0.05$. COMM (0,0497) was also estimated significantly higher than PWE. In addition, the main advantage identified by GISA employees was the support of project work. Consequently, reservations by project managers were the type of booking used most prevalently. Due to the optimized assignment of rooms and the possibility to cooperate with colleagues of other organizational units, information and knowledge exchange was supported significantly. Project managers especially mentioned their decreased efforts to communicate with members of their project team.

<i>statement</i>	<i>agreement</i>
COMM: Flexible Office strongly improves communication between KWs.	0.66
NET: Networking with other KWs is strongly improved.	0.63
KEX: Knowledge exchange with other KWs is strongly enhanced.	0.63
LEARN: Learning at the workplace is strongly improved.	0.60
TRAN: Transparency about competencies of other KWs is largely increased.	0.59
PWE: Personal work efficiency is strongly improved.	0.59
SEARCH: Time for searching is strongly reduced.	0.55

Table 1: Empirical results on benefits of the Flexible Office solution

The possibility to use rooms for not project-related, day-to-day-business was seldom used. According to GISA employees, this is due to the fact that there were too few members in the pilot study and consequently too few skills to choose from and cooperate with. Instead, employees preferred work places with their own work group. In order to profit from potential benefits of theme development, managerial systems need to take this dimension into account and define strategic knowledge assets requiring attention and thus providing legitimation or establish communities providing commitment for efforts spent on theme development.

According to project managers, organizational units with a focus on consulting and projects took more advantage of the pilot than their colleagues involved in application support and in-house services. Therefore, after a period of ten months, the project is accepted much more by colleagues involved in project work. However, the solution is supposedly generally feasible for organizational units having a high degree of out of office work like the ones focusing customer projects or sales.

Besides the apparent advantages, there also remain some issues in need of further testing. Initially, there was a considerable amount of doubts at the side of KWs that they would lose efficiency if they could not personalize their work spaces and had to recreate their work environments every time they moved into a new office. This challenge was dealt with by providing mobile personal storage capacities as well as technical equipment that allowed for location-independent office work including home offices. Another serious concern of employees was that they would lose

^[2] 1 = I fully agree – 0 = I fully disagree

influence over the decision of where to be located to a mechanistic solution. This issue was dealt with by weekly feedback solicited from the KWs that was used to evolutionarily develop the set of preferences individual KWs could put forward so that they could keep influence on the decision. During the pilot study, there was a tendency to see the room as “the own” room if projects had the same room for a longer period. Consequently, KWs expected to get the exact room again. Also, when projects had to change offices more than twice a week, moving was seen as a great burden. However, on the one hand, some participants feared that a flexible office would mean losing their personal work space and their relationships with colleagues. On the other hand, more and more KWs not participating in the pilot study claimed their interest in Flexible Office. Longitudinal studies are required to see whether these personal opinions amount to measurable improvements in the dependent variables.

5 Conclusions and Outlook

This paper presented the results of a joint research project on the deployment of a tangible KM instrument in the form of an optimization solution for the flexible assignment of office space. The solution consists of organizational and technical instruments which were field tested in a pilot study. The main difference compared to established approaches, like office 21^[3] focussing on organisational rules and equipment, is the mathematical formalization of factors relevant for KM and the subsequent optimization of office formation.

Surveys and interviews with affected employees show promising results concerning realization of the expected benefits, particularly for the integrative consideration of business system and project team layers whereas the knowledge base layer needs further investigation. Several advancements are planned or currently tested. On the technical side, it is planned to integrate the hotelling software with the company’s MS Exchange server, its MS Project server and the LDAP system to minimize administrative expenses. Organizational requirements have led to the addition of three more factors, i.e. influence of room properties, team centered rooms and a distinction between members of the pilot study and regular workers, which are currently tested. Last but not least, the user interface is improved to make it more intuitive, especially the graphical representation of room assignments.

The more significant organizational changes planned are to expand the project to include the company’s other sites, to integrate work places of other organizational units and consequently improve knowledge sharing between them.

References

[Maier, 07] R. Maier, Knowledge Management Systems. Information and Communication Technologies for Knowledge Management. 3rd edition, Berlin 2007, in print

[Nonaka, 94] I. Nonaka, A Dynamic Theory of Organizational Knowledge Creation, in: Organization Science, Vol. 5, No. 1, 1994, 14-37

^[3] <http://www.office21.de/>

[Sauermann, 05] L. Sauermann, The Semantic Desktop - a Basis for Personal Knowledge Management. In Proceedings of I-KNOW 2005. 5th International Conference on Knowledge Management, 294-301

[Sveiby, 87] K.-E. Sveiby, T. Lloyd, Managing Knowhow, London 1987; based on: Sveiby, K.-E., Risling, A.: Kunskapsföretaget (in Swedish; the Know-How Organization), Malmö 1986

[Wiig, 88] K.M. Wiig, Management of Knowledge: Perspectives of a New Opportunity, in: Bernold, T. (ed.): User Interfaces: Gateway or Bottleneck?, Proceedings of the Technology Assessment and Management Conference of the Gottlieb Duttweiler Institute Rüslikon/Zurich (CH), 20 - 21 October, 1986, Amsterdam 1988, 101-116

[Zinser, 04] S. Zinser, Flexible Arbeitswelten. Handlungsfelder, Erfahrungen und Praxisbeispiele aus dem Flexible-Office-Netzwerk, Zurich (CH) 2004

Retaining Knowledge Management Maturity Models: An Ontology-based Approach

Mark Hefke, Frank Kleiner, Andreas Storckenmaier

(FZI Research Center for Information Technologies at the University of Karlsruhe, Germany
{hefke, kleiner, storckenmaier}@fzi.de)

Oliver Paulzen

(Deutsche Post AG, Germany
Oliver.Paulzen@DeutschePost.de)

Abstract: This paper describes our work on developing an ontology-based software infrastructure for retaining and maintaining theoretical Knowledge Management (KM) Maturity Models (ONTOKNOM) by using a KM Maturity Model Ontology. Moreover ONTOKNOM provides technical means for designing a web-based system that supports the form-based self-evaluation of an organization with regard to its current maturity level, as well as for providing concrete organizational recommendations and measures in order to achieve a higher one. A description of including the existing KPQM model into the system is also provided.

Keywords: Knowledge Management, Maturity Model

Categories: A.1, H.1

1 Introduction

A deeper look at KM literature shows, that there has already been spent a lot of work on developing methods and instruments for supporting the introduction and the steady use of KM [Davenport, 98], [Probst, 99]: KM has to be integrated in daily business in order to handle typical knowledge processes like the acquisition, structuring, development and distribution of knowledge in a more efficient way. Furthermore KM necessarily has to focus on organizational, technical and human aspects in equal measure by considering already existing organizational structures, technical infrastructure and (knowledge intensive) processes. Since many KM projects fail as a result of insufficient know-how about KM strategies, it is furthermore important to have a strategy showing the way how to proceed. But one of the major challenges of bringing KM activities into practice is that organizations have a varying maturity level with regard to already realized KM activities. Therefore it seems to be necessary to identify the current KM maturity level of an organization before starting a KM implementation project in order to avoid KM project failures from the start. For this reason, KM Maturity Models have been developed that help to analyze an organization's maturity level with regard to KM by checking the implementation of different KM-specific activities subject to a maturity level like it has been done before in the context of software process improvement.

Our basic contribution is not to provide yet another theoretical KM maturity model but to make public available a flexible system-based infrastructure for

supporting the easy storage and maintenance of such models that are often only paper based or just supported by simple and inflexible technical means. The developed infrastructure supports the web-based determination of an organization's maturity level. Moreover, it helps organizations in reaching a higher maturity level by providing concrete recommendations and measures based on the underlying KM maturity model. Therefore we define a typical procedure for the identification of an organization's KM Maturity Level and analyze the structure of common KM maturity models.

The paper is organized as follows: Our methodological approach is presented in section 2. Section 3 describes the development of a KM Maturity Model Ontology for the retention of KM maturity models, organization profiles, as well as of an evolving organization's maturity level. An overview of all software infrastructure components, which principally provide the user interface and access to the ontology, is given in section 4. Section 5 presents a selected KM maturity Model, serving as a use case for validating our developed infrastructure. After a discussion of related work in section 6, concluding remarks outline some future work in section 7.

2 Methodological Approach

Our methodological approach has been divided into firstly defining a typical procedure for identifying an evolving organization's maturity level and secondly analyzing the basic structure of KM maturity models from literature, in order to derive requirements for a general KM Maturity Model Ontology and for our software infrastructure.

2.1 Procedure

As a result of defining the typical procedure of identifying an organization's KM maturity level, we obtained the following major steps:

1. If a user is doing an evaluation for the first time, he or she registers with the system by choosing a password and answering some general questions that describes his or her company (i.e. company size, sector, etc.). After that, the system returns a company identifier for logging in. Otherwise the user logs in using a previously created account.
2. If the user continues an evaluation process the system presents questions of the current maturity level to be answered using a web-based form.
3. If the user exits the system during an ongoing evaluation, the state of the evaluation is automatically saved and restored upon continuation.
4. After answering all questions of a maturity level, the system calculates the evaluation results in the form of a maturity level on a scale from m to n (where m represents the maturity model's start level and n represents the model's stop level). After that, the current maturity level is provided to the user combined with recommendations on how to reach a higher one.
5. After e.g. printing out the evaluation results/ recommendations, the user logs out and puts them into practice.

6. Upon completion of all displayed measures, the user logs in again and continues the evaluation process by answering the previously failed questions. As soon as all questions of a certain level are answered correctly, the level is considered as reached and the questions of the next level are presented.

2.2 Structure of a KM maturity model

The basic idea of identifying an organization's KM maturity level was derived from software engineering where the Capability Maturity Model[®] for Software (SW-CMM[®]) has been developed in order to evaluate the quality of software development processes in an organization by the use of a survey-based evaluation method. The model distinguishes five maturity levels in terms of an evolutionary path from ad hoc, chaotic processes (Initial) to mature, disciplined software processes (Optimizing): Except for Level 1, each maturity level is decomposed into several key process areas that indicate the areas an organization should focus on to improve its software process. In the meanwhile, various theoretical KM maturity models were developed that were based on the original idea of SW-CMM[®] (cf. section 6 "Related Work").

3 A conceptual model for representing KM Maturity models

In order to achieve a conceptual data model for representing one or more theoretical KM maturity model(s) consisting of different levels, questions, and linked measures as well as for retaining and managing organization profiles that have to be associated with evolving maturity levels, we decided to model an ontology [Staab, 03] which provides the data layer. The developed ontology can directly be accessed by the later discussed system components of ONTOKNOM (cf. section 4).

3.1 Description of ontology terms

The KM Maturity Model Ontology is modelled in KAON¹ language. Each property has at least one domain concept. Its range may either be a literal or a set of at least one concept. Domain and range concept restrictions are treated conjunctively. Consequently, all of them must be fulfilled for each property instantiation. Furthermore two properties can be inverse to each other. Concepts and properties can basically be arranged in a hierarchy. The hierarchy relation relates directly connected concepts (properties) and it is defined as a transitive relationship. The ontology has an instance pool associated with it. It is constructed by specifying of different concepts and by establishing property instantiations between. Property instantiations must follow the domain and range constraints. A detailed description of the KM Maturity Model including all concepts, attributes and relations has already been presented in [Hefke, 05]. In the following, we only concentrate on relevant entities for retaining KM maturity Models and identifying a company's maturity level.

¹ <http://kaon.semanticweb.org>

3.2 Retaining KM Maturity Models

In order to instantiate different models in one ontology each model is an instantiation of the **“Model”** concept. Each model is assigned attributes indicating the start and stop levels. The **“hasStartLevel”** and **“hasStopLevel”** links a model to the **“Level”** concept. The **“hasTopic”** property linking to the **“Topic”** concept describes the topic of a maturity model (e.g. KM). Every maturity model marks its associated questions with the **“consistsOf”** property. The **“Question”** concept is linked to a level through using the **“hasLevel”** property. Maturity levels are represented by the **“Level”** concept and the different instances indicating which level a specific question belongs to. The **“hasAnswer”** property links a question instance to an instance of the **“Answer”** concept indicating the right question for an answer which is either **“yes”**, **“no”**, or another user-defined value. The property (question) **“hasMeasure”** indicates the connection between the concepts **“Question”** and **“Measure”** whose attribute **“Measure-Text”** is displayed to the evaluating organization in case the question was answered with the wrong answer. Since it is possible that questions depend on each other it is possible to define dependencies between instances of the **“Question”** concept using the inverse property **“dependsOn”**. It is often desirable to have the opportunity of splitting models into sub models in order to structure large maturity models. Each model in the ontology can have sub models and/or be a sub model of other models. There are two properties which control the structuring of models: **“hasSubModel”** and **“Standalone-Model”**. **“hasSubModel”** is an inverse property which links to a model being the sub model of the model with the **“hasSubModel”** property. **“Standalone-Model”** is an attribute taking true or false as values. A model is a standalone model if there is no super model referencing with the **“hasSubModel”** property to the respective model.

3.3 Identification of a Company's Maturity Level

Each registered company is represented as an instance of the **“Company”**. The property **“hasAnswered”** marks already answered questions of a company, while the property **“hasAccomplished”** marks questions that were answered correctly. Defined measures are later determined through looking for questions that are marked with an instantiation of the property **“hasAnswered”** but not with an instantiation of the property **“hasAccomplished”**. In order to verify if a company has reached a particular maturity level, the later described evaluation component accesses the ontology and checks, if all questions of that level have been answered right. This will be done by checking the availability of the property instantiation **“hasAccomplished”** for every answered question. If there exists a question that has not been answered right by the user, the corresponding instance of the concept **“Measure”** is presented to the user. The measure is directly related to the wrong answered question using an instantiation of the property **“(Question) hasMeasure”**. After the realization of the provided measure(s) and the consecutive system login, the system again provides previously wrong answered questions to the user. If the company then answers all questions correctly and therefore reaches a particular maturity level, the level is stored into the ontology by a new drawn instantiation of the property **“hasCompanyLevel”**. To identify the state of evaluation for a company that would like to continue a not yet completed evaluation on a particular level, the ontology provides the property

"hasAnswered". This will guarantee that already answered questions of a particular level are not given again to the user. The property "dependsOn (Question)" which is defined on the "Question" concept guarantees that question hierarchies are observed. If one question which has dependant questions is answered wrong, all sub questions are not asked but are marked as still not answered.

4 System Components of ONTOKNOM

In the following we describe the system components of the developed architecture ONTOKNOM which consists of three horizontal layers and two vertical sets of components. We distinguish between components for the normal use of the system (user components), and components for creating and maintaining models (administration components). They are implemented in JSF and Java and directly connect to the KAON-API, which on this part connects to the Maturity Model Ontology.

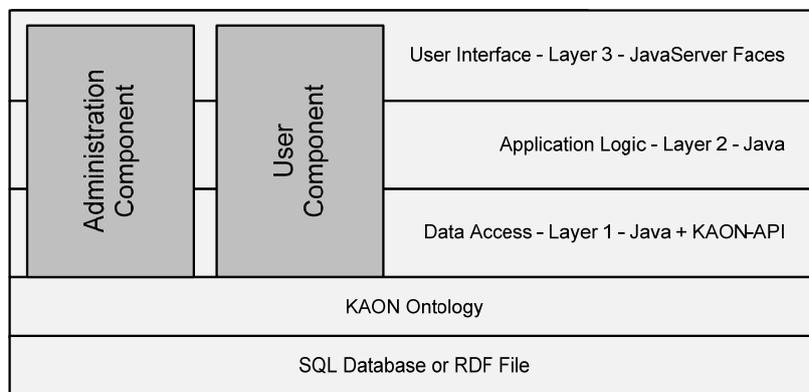


Figure 3: System Components of ONTOKNOM

4.1 Administration Component

The Administration Component comprises the following functions:

- **User Administration:** This function basically supports registering and deleting user accounts, since the registration of new users is done by themselves using the user component.
- **Creating and Managing KM Maturity Models:** A user interface provides means for creating, managing, and deleting maturity models that are stored by the Maturity Model Ontology. It is furthermore possible to hierarchically structure models by applying super/sub model properties as already described above. Interdependences between questions of a model can be handled by the **Dependency Function**, that allows the specification (as well as the deletion) of dependencies among questions. In addition, dependencies

can be checked for cycles to avoid failures since questions which have cyclic dependencies are not presented to the user.

- **Ontology-based Glossary:** In order to provide background knowledge to a created KM Maturity Model, ONTOKNOM provides means for setting up an ontology-based glossary containing terms can later be used to enrich questions automatically.

4.2 User Component

In the following functions of the User Component are described in more detail:

- The **Question Loader** selects questions from the level, which the company is trying to achieve. Each question is displayed on a single page giving the user the opportunity to answer it by choosing “yes”, “no” or any other given option and pressing the continue button. This procedure is repeated until all questions are processed which belong to the level the company wants to achieve. Questions are displayed according to their dependency status. If a model is divided into sub models then the user is asked questions in different areas (corresponding to the sub models) subsequently. Additionally defined glossary terms are highlighted in questions and explanations can be extended.
- **Evaluation Function.** After answering all questions of a level in the form component, the evaluation component is used to display associated measures for failed questions. Measures can include descriptions, external links and external links to graphics explaining the measures.

5 Use Cases

The basic strength of ONTOKNOM is its ability to provide technical means for the retainment of any staged KM Maturity Model. In the following, we shortly sketch one of our use cases, which we have successfully implemented by ONTOKNOM.

Maturity Level	Description
0 - Initial	The quality of KPs is not planned and changes randomly. This state can be best described as one of chaotic processes.
1- Aware	Awareness for knowledge processes has been gained. First structures are implemented to ensure a higher process quality.
2- Managed	This stage focuses on the systematic structure and definition of KPs which includes the clear assignment of responsibilities.
3- Standardised	A standard process scheme guarantees the achievement of a constant quality of results. Processes are tailored to react to special requirements. Standardization aspects can also be transferred to education and assignment of personnel.
4- Quantitatively Managed	To enhance the systematic process management, measures of performance are used to plan and track processes.
5 – Continuous Improvement	The focus on this stage lies on establishing structures for continuous improvement and self-optimisation.

Table 1: KPQM- Model

5.1 The Knowledge Process Quality Model (KPQM)

The KPQM Model by [Paulzen, 02] describes a maturity model for the assessment and systematic improvement of knowledge processes (KPs). It is based on the SPICE (Software Process Improvement and Capability Determination) framework² and consists of the six maturity levels Initial, Aware, Managed, Standardised, Quantitatively Managed and Continuous Improvement (cf. Table 1).

A maturity level aggregates several process attributes (e.g. knowledge process awareness, knowledge process management, standards for education and incentives) which can in turn be audited by related KM activities. Furthermore, the model structures the process attributes by distinguishing the four management areas *process structure*, *personnel assignment and knowledge networks*, *acceptance* as well as *computer-based support* to ensure a holistic approach towards KM implementation.

5.2 Technical Implementation

After setting up the technical infrastructure, the KPQM model had to be transferred into the ONTOKNOM system. To map KPQM to the system, four sub models representing the four management areas and one supermodel consolidating the submodels were build (cf. section 3.1). The questions including dependencies among questions were derived from the descriptions of KM activities [Paulzen, 02] which are needed to fulfil the requirements of a certain maturity level. Measures were generated by compiling the required KM activities of the next higher level. Additionally to including the pure model into the system, a glossary was build to support users in understanding terms unfamiliar to people who are inexperienced in KM or information technology. Apart from KPQM there exists another KM Maturity Model within ONTOKNOM, which is based on [Kochikar, 00]. Both can be accessed at

<http://www.ontoknom.de>

6 Related Work

At this stage, there exist several theoretical but also tool-supported models for identifying the current KM maturity level of an organization (e.g. the KMMM[®] from Siemens [Langen, 04], Kochikar's KM Maturity Model [Kochikar, 00], Berztiss' Capability Maturity for KM [Berztiss, 02], the Knowledge Process Quality Model (KPQM) [Paulzen, 02] and others) which base in the majority of cases on the Capability Maturity Model[®] for Software (SW-CMM[®]) [Paulk, 95]. An overview of existing KM maturity models is also given in [Weerdmeester, 02]. However our endeavour of developing an ontology-based model in order to provide a common ground for retaining KM Maturity Models, as well as the provision of an easily maintainable and therefore sustainable ontology-based software infrastructure for web-based self-evaluation is rather unique in this context.

² Cf. <http://www.sqi.gu.edu.au/spice/>

7 Conclusions and Future Work

In this paper, we have described ONTOKNOM, an ontology-based software infrastructure for technically supporting the retainment, representation and application of any theoretical KM maturity model. This has been realized by providing a KM Maturity Model Ontology as well as a system infrastructure for supporting the creation, modification and enhancement of different maturity models. Based on a selected (multi-part) maturity model, ONTOKNOM also supports the web-based identification of an organization's KM maturity level and provides organizational measures to achieve a higher one. In the future we will perform a further validation of our approach by integrating and combining additional KM maturity models. Furthermore we will evaluate its capability on providing means for the flexible management of ontology-based maturity models as well as their flexible combination. This will be done by e.g. making it available to consultancies in order to accompany organisations more efficiently in the holistic introduction of KM.

References

- [Bertziss, 02] A.T. Bertziss, Capability Maturity for Knowledge Management, 13th International Workshop on Database and Expert Systems Applications (DEXA'02), Aix-en-Provence, France, 2002
- [Davenport, 98] T. Davenport and L. Prusak, Working Knowledge. How Organizations Manage What They Know. Harvard Business School Press, 1998
- [Hefke, 05] M. Hefke, F. Kleiner, An ontology-based software infrastructure for retaining theoretical Knowledge Management Maturity Models, 2005
- [Kochikar, 00] V.P. Kochikar, The Knowledge Management Maturity Model: A Staged Framework for Leveraging Knowledge, KMWorld 2000, Santa Clara, CA, 2000
- [Langen, 04] M. Langen, K. Ehms, KMMM - Knowledge Management Maturity Model, Siemens AG, <http://www.kmmm.org>, 2004
- [Paulk, 95] M.C. Paulk, C.V. Weber, B. Curtis and M.B. Chrissis, The Capability Maturity Model: Guidelines for Improving the Software Process, Carnegie Mellon University, Software Engineering Institute, Addison-Wesley Publishing Company, Reading, MA, 1995
- [Paulzen, 02] O. Paulzen, P. Perc, A Maturity Model for Quality Improvement in Knowledge Management, In Wenn, A.; McGrath, M.; Burstein, F., Enabling Organisations and Society through Information Systems, Proceedings of the 13th Australasian Conference on Information Systems (ACIS 2002), Melbourne, pp. 243-253, 2002
- [Probst, 99] G. Probst, S. Raub, K. Romhardt, Managing Knowledge: Building Blocks for Success, Wiley, London, 1999
- [Staab, 03] S. Staab, R. Studer, Handbook on Ontologies, Springer Series on Handbooks in Information Systems. Berlin, Heidelberg, New York, Tokyo: Springer Verlag, 2003
- [Weerdmeester, 02] R. Weerdmeester, C. Pocaterra, M. Hefke, Knowledge Management Maturity Model, VISION Deliverable D5.2, <http://www.aifb.uni-karlsruhe.de/apollo20080/fzi/vision/deliverables.html>, 2002

Work Process Oriented Introduction of Knowledge Management: Reconsidering the Guidelines for SME

Thomas Herrmann

(Institute for Applied Work Science, Ruhr-University of Bochum, Germany
thomas.herrmann@rub.de)

Gila Brandt-Herrmann

(Centre for Continuing Education, University of Dortmund, Germany
gila.brandt-herrmann@uni-dortmund.de)

Isa Jahnke

(Institute for Applied Work Science, Ruhr-University of Bochum, Germany
isa.jahnke@rub.de)

Abstract: Within a series of practical projects to introduce knowledge management systems in small or medium-sized enterprises (SME), we realized the need to reconsider current guidelines such as the European Guide to good Practice in Knowledge Management. Our basic assumptions leading to this reconsideration are firstly that the introduction of knowledge management is a wicked problem, secondly that a strong relationship to the employees' actual work process has to be established as well as a dense integration of technical and organizational structures, thirdly that it needs at least a series of workshops to develop an understanding of the usage of KM in the workplace, and finally that the introduction is more often triggered by opportunities than by strategic planning. On this basis, we re-orientate the phases of introducing knowledge management.

Keywords: knowledge management, guidelines, project management, process of systems introduction, socio-technical systems; workplace; technology adoption

Categories: C.5, H.1.0, J.3

1 Introduction

Introducing knowledge management (KM) systems into small and medium-sized enterprises (SME) is a particular challenge because of the limited resources of these kinds of companies. The guidelines which help to establish projects for introducing knowledge management, such as the European Guide to good Practice in Knowledge Management - Part 3: SME Implementation [CEN Part 3 04] or others [Herbst 00, Brücher 04, Lehner 06, Belliger and Krieger 07] are more practitioner-orientated than the more comprehensive compendia [Mayer 04, Schwartz 06] in the field of knowledge management. We have contrasted the available guidelines with our experience of a series of cases¹ which we have empirically investigated and / or

¹ Background of 10 cases is the project WInn (BMBF 01HL0018 [Herrmann et al. 03]), of further 2 cases the project Wink (www.wissenmanagen.net, BMWI MM00029) and of 11 cases the project Nova.PE (<http://www.novape.rub.de>).

where we were involved as consultants in an action research² context. We have accompanied knowledge management projects in several business areas with companies of different sizes, mostly SME (IT-consulting, insurance company, chemical and cosmetics industry, development of e-learning software, web-design consulting, PR-agency, quality assurance with respect to steel welding, valve layout, manufacturing planning, communication agencies, public relations management of a non-profit organisation).

According to our experience, the guidelines for introducing knowledge management have to be modified for SME in these business areas, especially for manufacturing branches. In the following sections we focus our considerations on those aspects which indicate the differences between our recommendations and the current literature. The structure of the following description uses the typical differentiation between phases such as starting the project [section 2], analysis/assessment [section 3], planning and developing a solution [section 4], implementation [section 5], and evaluation [section 6]. These phases can be highly interwoven by going back and forth. Our approach can be characterized by the following basic assumptions:

- Establishing knowledge management is a wicked problem [Rittel and Webber 73, Conklin 05] which is not clearly defined at the beginning of the project. The definition of the problem is mostly only completed when the solution of the problem is achieved. People don't have clear ideas about what knowledge management is, how it can help them, and what expectations they have towards these types of methods and tools.
- KM has to be established as a socio-technical solution which integrates technical and organizational structures and prepares the human participants. It is not only about configuring technical components but also about developing commitments and conventions.
- The KM solution has to be deliberately grounded in the actual work processes of those employees who are going to become the KM users. Therefore, the individual tasks, as well as their interplay in the dynamics of work processes, have to be understood.
- KM in SME is de facto more often triggered by opportunities than by endeavours of strategic planning.

2 Starting the Project

2.1 Strengthening the starting basis

The available guidelines usually start at a point (setting up the project) where it is presumed that a sufficient starting basis is already established – or they recommend

² Action research is an iterative process in the context of the focused aspects enabling us to understand a social or sociotechnical phenomenon or to improve its quality. It consists of several phases “analysis” (reflection) and “action” (change) which alternate and are interwoven, which Avison calls “cycle of action”: “Action research involving researchers and practitioners acting together on a particular cycle of activities, including problem diagnosis, action intervention, and reflective learning (...) in real situations, gain feedback from this experience, modify the theory as a result of this feedback, and try it again” [Avison et al. 99, p. 94-95].

starting globally, analyzing the company's situation (processes, organisation, human resource management, information technology, knowledge, assets, competences etc.) according to a top down approach [cf. Treichel 2007]. However, it usually has to be a task preceding the project to check whether an appropriate starting basis is available. This basis should be strong enough to make a sustaining project and a successful implementation possible. The salient triggers which will push the project have to be understood. Typical examples are problems – as kinds of negative triggers – with customer relations, with communication between different participants during manufacturing processes, with the quality of products and services, with information overload, with incomplete documentation of project work, with changes within the personnel etc. We found those problems mainly in the departments of the manufacturing companies where overcoming them was a critical success factor. Positive opportunities are the increased motivation of people with respect to knowledge sharing, ongoing innovation processes, the need for reliable knowledge when one has to react to difficult situations or the availability of new technology.

The starting basis can be strengthened by finding out what the most promising topics or areas of work are, and where the most highly motivated people can be found who are willing to promote knowledge management.

2.2 Creating an understanding of what KM is about

Although some guidelines propose starting with a strategic approach and asking questions such as “Why is KM necessary?”, “What kind of visions and which goals do you connect with KM?”, we do not recommend proceeding in this way. The answers to these questions are not reliable, since the questions are far too abstract and people do not usually really understand what KM is about at the beginning of the project. Our experience shows that the project should begin with examples of prototypes or of applications as they run in other companies. These examples should illustrate potential benefits in relation to the problems and needs employees experience in their every day work life.

2.3 Establishing an orientation towards work processes

It is mostly assumed that strategic thinking and the definition of and a commitment to a set of knowledge related goals is a crucial factor which increases the sustainability of KM. This approach suggests that a knowledge management strategy is introduced by a top-down implementation. However, we suggest that anchoring KM in the work process is more decisive with respect to sustainability. This orientation includes urgent problems occurring during everyday work having to be identified. It has to be carefully checked to establish whether KM can contribute to solving these problems and whether it may lead to quick wins which make the benefits of KM easily comprehensible. We found examples where the work-orientation was neglected but a lot of effort was invested to produce clear and concise stories to document how new products (cosmetic industries) had been brought to the market. However, these stories were not considered useful by the employees to support them in their everyday problems. By contrast, another company (public water supplier [cf. Cichowski 07]) had successfully implemented a reporting system which was directly inspired by lack of information as it occurred during the organization of meetings and decision

making. Therefore, we propose a work process orientated approach. It considers the interdependencies between different organizational units, between different phases of time, between different participants and their resources etc.

2.4 Promoting marketing and involving the management

Even in the starting phase, a continuous process of marketing for the KM-project has to begin. We found, in two larger companies (chemical industry, insurance company), which have invested in more communication to explain the advantages of KM to their personnel, that the usage of the system was more sustaining than in eight other cases which we analyzed during the same time. Furthermore, in another company the management did not support the decision on the proposed area where the solving of knowledge problems was considered most urgent – they tried to push into another direction. In a further case, management was not willing to allow a certain highly motivated employee to be the promoter of the KM-project. In both cases the projects were eventually turned down. Therefore it is definitely reasonable that members of the management join the KM-project as active supporters of proposals which stem from bottom up decision making. Additionally, they should belong to the first adaptors who start to work with the system. The analysis of the starting basis should also take into account that there might be other intervening factors and dependencies such as limited personal motivation or resources, the market and order situation of the company, other ongoing projects of changing the infrastructure etc.

3 Assessment and Analysis

The main aspect of our experience compared to current guidelines is that one should not start with the classification of knowledge or with knowledge maps but should begin with the every day tasks. For example, a KM-solution for the maintenance of machines should not start with classifying the description of these machines but with describing the typical activities of the service engineers and their problems. The work-orientated approach includes going back and forth between:

- the task and the information being used/produced
- abstract information level and concrete examples
- kind of knowledge or types of information which can be made explicit and those employee skills which cannot be documented and therefore remain implicit.

We suggest using a methodological approach which allows one to depict the work activities and to make them a subject of mutual reflection – where those details can be taken under considerations which are related to the employing and production of knowledge. We recommend using a semi-formal modelling notation to represent the work processes and information handling with diagrams. We have developed the modelling method “SeeMe” [Herrmann and Loser 99] to give an example of the characteristics of such a semi-formal notation. SeeMe is comparable and compatible with other modelling notations but has some distinctive features: explicit indicators for incompleteness and uncertainty, rough as well as complete specification of relationships, multi-perspective decomposition of elements, indication of possibilities

for free decision-making³. The advantage of such a semi-formal representation is that it is suitable when expressing the contingent relationships of social structures as well as the formal specifications of a technical solution. SeeMe is employed to focus the communication and documentation within an approach which we call the socio-technical Walkthrough – STWT⁴ [Herrmann et al. 07]. The STWT consists of a series of participatory workshops where the graphical models of work processes, of knowledge structures and/or examples of technical components are documented, discussed and improved by considering their details step-by-step. This in-depth consideration and discussion has to be systematically facilitated. The STWT-workshops can start supporting the analysis and are also feasible to promote the development of a solution and its implementation.

The recommendation to firstly document the knowledge assets doesn't sufficiently take into account that the relevant knowledge is mostly not explicitly documented. Furthermore, it can even be incomprehensible which kind of implicit knowledge is relevant. Often enough it is hard to find traces which indicate the relevant items of knowledge. The focus on documenting the already established tools, methods and documentations of knowledge management neglects the tremendous effort which has to be invested into the process of extracting knowledge and making it explicit. The guidelines should pay more attention to the challenge of knowledge engineering⁵ since it can be so time consuming that it slows down the whole KM-project and makes its success questionable.

The result of the assessment and analysis phase should not be an isolated documentation of types of knowledge, or knowledge maps, which is not anchored into the everyday work, but the kind of documentation which represents the knowledge structure and its close coupling to the work, where information is used and produced.

4 Developing the Solution

From our point of view, the development and planning of a solution have to be closely coupled with the analysis of the work process. This analysis has to be migrated into a concept which specifies the technical tools which may support the information structuring, storing, and distribution, as well as the organizational conventions which accompany these activities.

Current guidelines propose a set of possible methods from which the project leaders can select appropriate tools, such as Balanced Scorecard, Brainstorming, Minutes, Video conference [cf. CEN Part 3 04, p. 22 ff]. With these recommendations CEN does not give details of how one should proceed in order to extract the

³ Herrmann [06] provides a detailed description of the modelling notation.

⁴ For every STWT-workshop, the steps of the walkthrough have to be prepared as well as the questions which accompany the steps and the granularity of details to be discussed. During the workshops, all participants must have a voice and contributions have to be immediately visualized within the SeeMe-diagram. The participants represent varying perspectives such as technology, organization, management etc.

⁵ The term „knowledge engineering” has its origins in the context of designing Knowledge Based Systems [cf. Studer et al. 98]. We use this term with a narrower connotation to circumscribe the effort to construct knowledge with experts to make it explicit.

knowledge. For practitioners it is not sufficiently helpful to give them references to a set of methods which they might not be familiar with. By contrast, they should have a clear description of a procedure to analyse the work processes and to develop a KM solution. We propose combining various methods of analyzing, facilitating, documentation etc. into the integrated approach of the STWT [Herrmann et al. 07].

To enter into the phase of developing a solution, the 2nd or 3rd STWT-workshop can begin with a graphical model, or with the example of a technical component, which are discussed step-by-step to give the participants a feeling of what a KM-solution can look like. This socio-technical concept has to be permanently mirrored to the potential users by employing an appropriate modelling method. It should be possible to enrich graphical models with examples of documents and prototypes which help to make the solution more understandable. The graphical representations can help to build a large picture of the employees' varying knowledge contributions, to make their thoughts visible and to reduce the complexity of their oral descriptions. Early versions of the technical components can give them an idea what the interaction with the KM-tools during their work will look like. It can be reasonable to work with throw-away-prototypes which have the sole purpose of illustrating how technical support and knowledge-based work can be intertwined. It has to become clear from the beginning that technology is a secondary issue while it is more important to find commitments which prepare conventions, e.g. about the structuring of folders, assigning of keywords, rhythm of updating the documentation etc.

5 Implementation

The implementation is mostly about bringing the planned solutions into reality. We found some important aspects which should be emphasized since they are usually neglected in the course of implementation:

- Make sure that the users have had sufficient training and that they have understood the reasons why they should employ new tools and comply with new conventions of knowledge handling
- Provide an easily accessible help desk
- Make sure that the benefits of the system become obvious as fast as possible
- Prepare key-users who are knowledgeable and willing to help others
- Employ graphical models of the work process in a STWT-workshop to train the potential users.
- Make sure that the technology really works before it is used by employees who are absorbed by their actual tasks and not willing to be part of an experiment.
- Make a clear cut by transferring the relevant data from the old to the new system; try to avoid the same data being managed and updated twice on two systems
- The new system should be integrated into the infrastructure of its environment as far as possible and from the beginning to avoid unnecessary maintaining of data.
-

6 Evaluation

During the evaluation, direct personal feedback is more valuable than operationalizing abstract strategic goals (such as increased extent of shared knowledge) and measuring their degree of fulfilment. Gathering personal feedback means that those who were involved or interested in the KM-project should report on whether their task completion has become more effective and efficient, whether their job satisfaction has increased etc. People should be asked whether the quick-wins and other promises which were made at the beginning have been accomplished. The simple question is whether the most urgent problems which lead to the KM-project are really solved. This elicitation of feedback should immediately become part of continuing improvement. The KM-project as well as the users of the KM-solution should be aware that the analyzed work processes and knowledge structures are not stable, but subject to dynamic change which requires continuous adaptation.

7 Conclusion

The European Guide to good Practice in Knowledge Management [CEN Part 3 04] is a valuable orientation for those who are newcomers to the field of KM. However, this guideline is too abstract if clear guidance is required on how to proceed in the case of KM-introduction. For example, the tools which are mentioned, such as lessons learned or debriefing are not an appropriate set of means. They may help to identify the relevant knowledge assets but do not sufficiently describe the procedures of how implicit knowledge can be identified and made explicit.

By contrast, our approach helps the KM-project to run STWT-workshops where the participants make their knowledge explicit for others, classify it with respect to their tasks, and document it in conjunction with the work processes. This conjunction results in increased comprehensibility of the process and their improvement. This improvement may be achieved although it was not a strategic goal at the beginning of the KM-project. The STWT-workshops can accompany all the phases of the KM-project and support the intermingling of these phases, since it may happen that the phase of implementation can even reveal needs for continuing the analyses of work processes and the related knowledge.

The main difference is that the CEN-guideline tries to focus the companies on implementing a strategic KM-approach from the very beginning while we concentrate on the success of the first KM-projects. We suggest that exemplarily establishing a close relationship between work processes and KM will be the best promotion of a strategic KM-approach in SME. Furthermore, assessing the current state of knowledge assets [CEN Part 3 04, p. 8] can become a very time-consuming job because of the huge part of implicit knowledge. Therefore, it is reasonable to identify the need for documented knowledge by analyzing the work processes to focus on those parts of knowledge where the effort of making it explicit is unavoidable.

References

- [Avison 91] Avison, D., Lau, F., Neilsen, P.A. & Myers, M.: "Action Research"; Communications of ACM, 42, (1991), 94-97
- [Belliger and Krieger 07] Belliger, A.; Krieger, D.: „Wissensmanagement für KMU“; Hochschulverlag AG, Zürich (2007)
- [Brücher 04] Brücher, H.: „Leitfaden Wissensmanagement. Von der Anforderungsanalyse bis zur Einführung“; Hochschulverlag AG, Zürich (2004)
- [CEN Part 3 04] CEN Part 3: Wissensmanagement-Einführung in KMUs; European Guide to Good Practice in Knowledge Management; Brüssel (2004), 56-83, <ftp://cenftp1.cenorm.be/PUBLIC/CWAs/e-Europe/KM/German-text-KM-CWAguide.pdf>
- [Cichowski 07] Cichowski, Ch.: „Protokolle von morgen“; Demographischer Wandel: Die Herausforderung von lebenslangem Lernen im 21. Jahrhundert. Wissensmanagement Heft 1, (2007), 16-17
- [Conklin 05] Conklin, J.: „Dialogue Mapping“; Wiley, Chichester (2005)
- [Herbst 00] Herbst, D.: „Erfolgsfaktor Wissensmanagement“; Berlin, Cornelsen (2000)
- [Herrmann et al. 07] Herrmann, Th., Loser, K.-U., Jahnke, I.: "Socio-technical Walkthrough (STWT): a means for Knowledge Integration"; International Journal of Learning Organisation. Special Issue: Solving Problems in Knowledge Sharing with Sociotechnical Approaches, Forthcoming (2007)
- [Herrmann 06] Herrmann, Th.: "SeeMe in a Nutshell. The Semi-Structured Socio-Technical Modeling Method", http://www.imtm-iaw.rub.de/imperia/md/content/seeme/seeme_in_a_nutshell.pdf (2006)
- [Herrmann et al. 03] Herrmann, Th., Mambrey, P., Shire, K.: „Wissensgenese, Wissensteilung und Wissensorganisation in der Arbeitspraxis“; Westdeutscher Verlag, Opladen (2003)
- [Herrmann and Loser 99] Herrmann, Th., Loser, K.-U.: "Vagueness in models of socio-technical systems"; Behavior & Information Technology: Special Issue on Analysis of Cooperation and Communication 18 (5); (1999) 313-323.
- [Lehner 06] Lehner, F. (2006): Wissensmanagement. Grundlagen, Methoden und technische Unterstützung; Hanser, München / Wien (2006)
- [Mayer 04] Mayer, R.: „Knowledge Management Systems“; Springer; Heidelberg / Berlin (2004)
- [Rittel and Webber 73] Rittel, H., Webber, M.: "Dilemmas in a General Theory of Planning"; Policy Sciences, 4, (1973) 155-169
- [Treichel 07] Treichel, D.: „Projektmanagement für Wissensmanagement“; Wissensmanagement für KMU. Hochschulverlag AG; Zürich (2007), 129-145
- [Schwartz 06] Schwartz, D.G.: "Encyclopedia of Knowledge Management"; Idea Group Inc. (2006)
- [Studer et al. 98] Studer, R., Benjamins, R., Fensel, D.: "Knowledge Engineering: Principles and Methods"; Bericht 372, AIFB, Universität Karlsruhe (1998), 161-197

Towards Semantic based Requirements Engineering

Thomas Riechert

(University of Leipzig, Germany,
rieichert@informatik.uni-leipzig.de)

Kim Lauenroth

(University of Duisburg-Essen, Germany,
kim.lauenroth@sse.uni-due.de)

Jens Lehmann

(University of Leipzig, Germany,
lehmann@informatik.uni-leipzig.de)

Sören Auer

(University of Leipzig, Germany,
auer@informatik.uni-leipzig.de)

Abstract: Requirements Engineering is recognized as a crucial part of project and software development processes. This is due to the fact that the different stakeholders involved in a development project have to establish common terminologies as well as goals, scenarios and requirements expressed using these terminologies. Within the Semantic Web initiative various standards emerged for the creation and use of terminologies, expressed in the shape of semantic networks, taxonomies and ontologies. We develop an approach for semantic based Requirements Engineering. We present an ontology for capturing requirements relevant information. Furthermore, we report about a tool for semantic based Requirements Engineering and its application in a real-world development project scenario from the E-Government domain.

Key Words: requirements engineering, semantic web, ontologies, software engineering

Category: D.2.1

1 Introduction

Requirements Engineering (RE) is recognized as a crucial part of the software development process. Recently, RE techniques were also applied to project development and management in a more general context beyond software development. The importance of Requirements Engineering was identified by several studies [Hall et al. 2002] as a critical factor for the success of project development. This is due to the fact that the different stakeholders involved in a development project have to establish common terminologies as well as goals, scenarios and requirements expressed using these terminologies.

Within the Semantic Web initiative, various standards emerged for the creation and use of terminologies, expressed in the shape of semantic networks, taxonomies and ontologies (e.g. RDF, RDF-Schema, OWL). As software and

project development projects become more distributed in the sense that spatially separated stakeholders have to be integrated in the RE and project development cycle [Riechert et al. 2006], these terminological knowledge representation standards can form a solid basis for the elicitation, representation, structuring and management of requirements relevant information. Semantic representations of requirements relevant information, furthermore, can be a crystallization point for the integration of various tools for project development (such as project management software and CASE tools).

In this paper, we present an approach for semantic based Requirements Engineering. After introducing crucial RE concepts (in Section 2) we present a semantic structure (i.e. an ontology) for capturing requirements relevant information (in Section 3). In Section 4, we report about a tool for semantic based RE and its application in a real-world development project scenario from the eGovernment domain. We conclude with directions for future work in Section 5.

2 Requirements Engineering

Requirements engineering is the part of software engineering and project development processes that deals with the requirements for a planned system. It is a cooperative and iterative process that tries to attain three goals (cf. [Pohl 1996], [Pohl 2007]):

- *Elicit and understand all relevant requirements:* RE must elicit the requirements from the relevant stakeholders or develop the requirements together with them, because they are usually not known in advance.
- *Reach agreement:* Agreement on the requirements is a prerequisite for the acceptance of a planned system. It is not given that individual stakeholders agree to all requirements expressed by other stakeholders. Therefore, RE must identify and resolve existing conflicts between the stakeholders, e.g. by discussions, votings, or decisions of distinguished stakeholders.
- *Documentation:* Different types of systems require different types of requirement specifications. For instance, safety critical systems such as a control system of an airplane require a strict and formal definition of requirements in order to ensure that the system cannot endanger human lives. Consequently, it is essential for requirements engineering to document the requirements according to the specific needs of the domain of the planned system.

A common RE approach for achieving these goals, is the use of goals and scenarios as intermediate steps for obtaining requirements. Goals describe the stakeholders intentions for the planned system (cf. [Pohl et al. 1998]). They support for example the resolution of conflicts because it is easier to reach an agreement

be enabled to collect, semantically enrich, classify and aggregate software requirements. The rationale is to provide a semantic structure, which will capture requirements relevant information and enables interlinking of this information with domain and application specific vocabularies.

Figure 1 visualises the core of the SWORE ontology, which we developed in accordance with standards of the Requirements Engineering community [Pohl et al. 1998, Pohl 1996, Pohl 2007, van Lamsweerde 2001]. Central to our approach are the classes *Stakeholder* and *Abstract Requirement* as well as the properties *details* and *defines*. Abstract requirements have the subclasses *Goal*, *Scenario*, and *Requirement*, each of which are defined by stakeholders and can be detailed by other abstract requirements.

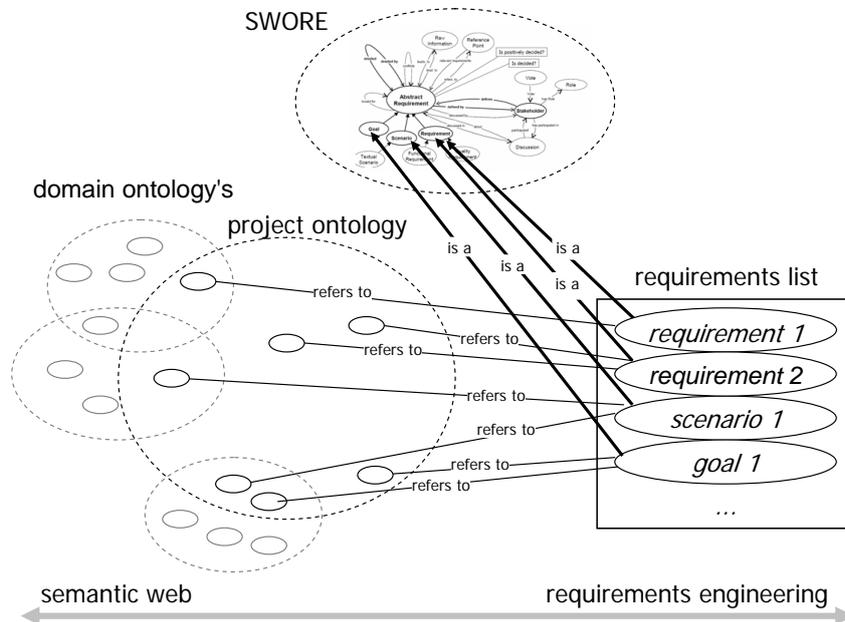


Figure 2: SWORE use and interlinking with domain ontologies.

This enables the specification of abstract requirements at different levels of granularity. (Elicit and understand all relevant requirements cf. Section 2) We emphasise the collaborative aspects of requirements engineering by integrating discussions amongst the stakeholders and voting (with the criteria of agreement and importance) in the model (Reach agreement cf. Section 2). In

the RE process this documentation is often relevant for future decisions (Documentation cf. Section 2). In order to interlink requirements with existing documents or resources, SWORE contains the classes *Raw Information* and *Reference Point* together with suitable properties. SWORE is available for download at: <http://softwiki.de/SWORE>.

The use of SWORE is visualised in Figure 2. A requirements engineering knowledge base will contain instantiations of the SWORE concepts. The innovation of using semantic representations in comparison with traditional RE tools is the possibility to reuse and reference existing information and domain ontologies. We expect the emerging of a multiplicity of domain ontologies, which describe knowledge specific to certain domains. These are very important for the RE process, since they facilitate the understanding between stakeholders and hence the annotation and referencing of requirements with such domain specific information is an integral part of the envisioned SWORE usage.

4 E-Government Use Case

An application and evaluation of the presented semantic based RE approach is currently performed in an E-Government use case. German citizens have to register their current place of habitation with the local community administration. For historic reasons a multiplicity of different applications was developed and is in use by individual administration offices. In order to facilitate automated data exchange between different community administrations and to provide self-service citizen services there are endeavors to homogenize and better integrate these individual applications on a local, regional and federal level.

The software application is implemented using the DotNetNuke Framework². The deployment at local public offices is prepared. We main use is to demonstrate the SoftWiki concept. We identified several data sources to import into the ontology. All requirements are stored in the project management tool in-Step³ in different granularities. The tool provides export functionality, without losing references between the requirements. Stakeholder groups have been extracted from the documentation in particular the manuals of the application. Additional reference points have been linked into the side map of the web application.

The problem of better integrating these different applications, however, is not just a technical one, since different community administrations established different processes and information structures and hence have different requirements regarding information integration and exchange. In order to enable the different stakeholders to collaboratively work on requirements for the standardization, information, and process integration of the registration service, we adopted the semantic collaboration platform OntoWiki [Auer et al. 2006].

² <http://www.dotnetnuke.com/>

³ <http://www.microtool.de/instep/en>

The screenshot displays the OntoWiki interface for a functional requirement instance. The main content area shows the following details:

- Properties:** Map, Calendar, History, Edit
- defined by:**
 - Entscheider
 - Fachadministrator
- description:** Eingabe einer Nutzergruppe des Konzeptes zur Zugriffssicherheit. Ein Administrator kann eine Gruppe anlegen. Es können keine Duplikate angelegt werden. Wenn die Gruppe erfolgreich angelegt wurde, können Zuordnungen angelegt werden.
- detailed by:**
 - Benutzerverwaltung Rollen Gruppen - Zuordnen zu Benutzergruppen
- defined by:** Fachadministrator
- description:** Zuordnen von Benutzergruppen zu einem Benutzer bzw. ...tattfinden und nicht in der DB gespeichert werden.
- detailed by:**
 - Benutzerverwaltung Rollen Gruppen - Aufheben von Zuordnungen von Benutzergruppen
- details:**
 - Benutzerverwaltung Rollen Gruppen - neues Nutzerkonto anlegen
 - Benutzerverwaltung Rollen Gruppen - Zuordnung eines Benutzers zu einer Gruppe
 - Benutzerverwaltung Rollen Gruppen - Eingabe einer Nutzergruppe
- refers to:** Benutzerverwaltung
- comment:** Anforderung aus InStep
- label:** Benutzerverwaltung Rollen Gruppen - Zuordnen zu Benutzergruppen
- refers to:** Benutzerverwaltung
- comment:** Anforderung aus InStep
- label:** Benutzerverwaltung Rollen Gruppen - Eingabe einer Nutzergruppe

The right sidebar contains several sections:

- Actions:** Export: GSV | RDF, Inline: Editing, Add Instance, Visit resource on the web
- Similar Instances:** functional requirement, Organisatorisch...rechnen können | Module zu Fachs...ndigkeitsfinder | Module zu Fachs... Formularserver | Module zu Fachs... Einstellungen | Module zu DL - ...tungsstellung, Show all 134 >>
- Rating:** Average Rating: Your Rating:
- Instances Linking Here:** defines: Fachadministrator, Entscheider, Show all 2 in main window >>; details: Benutzerverwalt...Benutzergruppen, Benutzerverwalt... zu einer Seite, Benutzerverwalt...u einer Methode, Benutzerverwalt... zu einem Modul, Benutzerverwalt...er Nutzergruppe, Show all 5 in main window >>
- relevant requirements:** Benutzerverwaltung, Show in main window >>
- Usage as Property:** Instances, Values
- Search:** Submit, All Knowledge Bases

Figure 3: Functional requirement instance related to user roles and access control of the E-Government application.

OntoWiki is a Semantic Web application enabling a distributed user community to develop ontologies and collect instance data adhering to these ontologies. OntoWiki aims at making the browsing of knowledge bases and user contributions as easy as possible, while trying to provide means for coherent vocabulary development. OntoWiki is open source software and available at <http://ontowiki.net>.

In order to use OntoWiki for the Requirements Elicitation, we pre-loaded the requirements ontology (from Section 3) into OntoWiki. Subsequently, stakeholders of the RE process are enabled to create and interlink requirements and scenarios. OntoWiki also provides means to vote, discuss and comment individual pieces of information and hence to support the requirements agreement. Figure 3 shows a functional requirement instance related to user roles and access control of the E-Government application.

The use of OntoWiki allows to incrementally create a knowledge base of requirements relevant information. In the spirit of adaptive methodologies, On-

toWiki does not presume a rigid RE process, instead requirements can be refined, annotated, discussed and interlinked at any time until the requirements knowledge base reached a stable state which the stakeholder community agrees on.

5 Conclusions and Future Work

The application of Semantic Web technologies is beneficial for RE processes, since a crucial part of RE is the establishment of a common terminology by different (often spatially distributed) stakeholders.

We are currently in the process of gaining more experiences with RE stakeholder communities and plan to adopt the generic knowledge engineering tool OntoWiki more to the RE usage scenario. In particular, we want to develop interfaces to traditional RE and management tools, such as IRqA⁴ and Doors⁵. The information exchange with such traditional tools is especially important, since we focus mainly on the early stages of requirements elicitation and RE processes with many, spatially separated stakeholders. Once an agreement is reached within a stakeholder community the gained information can be used inside traditional tools in order to support the remaining software or project development process.

Acknowledgements

We are grateful to Dirk Fritsch (QA Systems GmbH), who has extensive experience as a consultant in many RE projects, and Steffen Lohmann (University of Essen-Duisburg), for his valuable feedback.

References

- [Auer et al. 2006] Sören Auer, Sebastian Dietzold, and Thomas Riechert. Ontowiki - A tool for social, semantic collaboration. In Isabel F. Cruz, Stefan Decker, Dean Allemang, Chris Preist, Daniel Schwabe, Peter Mika, Michael Uschold, and Lora Aroyo, editors, *The Semantic Web - ISWC 2006*, 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, 2006, Proceedings, volume 4273, pages 736-749. Springer, 2006.
- [Riechert et al. 2006] Sören Auer, Thomas Riechert, Klaus-Peter Fähnrich. *SoftWiki Agiles Requirements-Engineering für Softwareprojekte mit einer großen Anzahl verteilter Stakeholder*. GeNeMe'06 - Gemeinschaft in neuen Medien, 29. Sep. 2006, Dresden, Germany.
- [Pohl et al. 1998] Peter Haumer, Klaus Pohl, and Klaus Weidenhaupt. Requirements elicitation and validation with real world scenes. *IEEE Transactions on Software Engineering*, 24(12):10361054, December 1998.
- [Pohl 1996] Klaus Pohl. *Process-centered requirements engineering*. Research Studies Press, Advanced Software Development, 1996.

⁴ <http://www.irqaonline.com>

⁵ <http://www.telelogic.com/products/doors>

- [Pohl 2007] Klaus Pohl. Requirements Engineering - Grundlagen, Prinzipien, Techniken. Dpunkt Verlag, 2007.
- [Hall et al. 2002] A. Rainer T. Hall, S. Beecham. Requirements problems in twelve companies - an empirical analysis. In Proceedings of the 6th International Conference on Empirical Assessment and Evaluation in Software Engineering (EASE 2002). Keele University, 2002.
- [van Lamsweerde 2001] Axel van Lamsweerde. Goal-oriented requirements engineering: A guided tour. In Fifth IEEE International Symposium on Requirements Engineering (RE01), pages 249-263. IEEE Computer Society Press, August 2001.

Social Network Analysis for Innovation and Coordination

Renato Balancieri

(Federal University of Santa Catarina and Stela Institute, Brazil
renato@stela.org.br)

Roberta Cuel

(University of Trento, Italy
roberta.cuel@economia.unitn.it)

Roberto Carlos dos Santos Pacheco

(Federal University of Santa Catarina and Stela Institute, Brazil
pacheco@stela.org.br)

Abstract: The innovation process is a rhythm of search and selection, exploration and exploitation, cycles of perspectives encountering which allow people to analyze problems from new points of view. In order to enable innovation, a lot of instruments have been developed connecting heterogeneous individuals thinking (e.g. social networks, web portals, wiki systems, organizational yellow-pages, etc.). In this paper we focus on web portals, and how these tools assist the users connections and the innovation processes among them. In particular, we analyze some services implemented in the Innovation Portal of the Brazilian Ministry of Science and Technology geared to stimulating the establishment of strategic partnerships and cooperation projects involving national firms and science institutions. These services are mainly based on social network analysis in order to manage connections (i.e. coordination) and innovation processes among users.

Keywords: Coordination Processes, Social Network Analysis, Innovation Portal

Categories: H.3.2, H.3.3, H.3.5, I.2.1, J.4, K.4.1

1 Introduction

Nowadays, organizations had to deal with dynamic markets, characterized by work specialization, outsourcing processes, just in time and distributed production, etc. In this scenario, innovation and the effective management of knowledge are becoming strategic actions that individuals and organizations have to deal with in order to compete in complex markets. As a consequence, a lot of theories and practices have been developed in both knowledge management¹ and organizational innovation² disciplines. Some of these theories (e.g. communities of practices, sensemaking, etc) argue that individuals, groups, communities and organizations tend to manage

¹ Knowledge management has been analyzed from various perspectives, the more objectivistic ones are discussing about knowledge market and intellectual capital, the more subjective ones are focusing on communities of practices, symbolic knowledge, storytelling, etc.

² Innovation theories are focused on innovation strategy, new product design and development, creative approaches to problem solving, reasoning systems, sensemaking, behavioural theories, etc.

knowledge according to their own perspectives, which satisfy their specific needs. Some authors stress the cognitive nature of interpretation schemas, where a schema is viewed as an individual's perspective of the world (see, for example, the notions of context [McCarthy 93; Ghidini and Giunchiglia 01], mental space [Fauconnier 85], and partitioned representation [Dinsmore 91]), others stress their social nature, where a schema is thought as the outcome of a special form of "agreement" within a community of knowing (see, for example, the notions of paradigm [Kuhn 79], frame [Goffamn 74], and thought world [Dougherty 92]). In this context, cycles of divergent thinking followed by convergence, or the encountering of different perspectives should be enabled to improve innovation.

One of the common instruments used to organize, collect and merge knowledge is the web portal [Davenport et al. 98]. Traditionally, this is composed by the following elements: (i) contribution processes which enable members to manage knowledge portal; (ii) chats and discussion groups, used to satisfy the need of social interaction; (iii) content management tools (text miners, search engines, and so on), which are used to produce a shared view (either implicit – e.g., clusters, neural nets – or explicit – e.g., ontologies, taxonomies) of the entire collection of corporate documents; (iv) common formats (such as HTML, XML, PDF), used to overcome the syntactic heterogeneity of documents composed by different knowledge sources.

The way most web portals are designed embodies a common organization of knowledge, according to which it can be represented in an objective form, which is independent from all those subjective and contextual elements described above. Unfortunately, any approach which disregards the plurality of interpretation schemas is bound to trouble, because the portal outcome will be perceived by users either as irrelevant (there is no deep understanding of the adopted schema) or as oppressive (there is no agreement on the unique schema, which is therefore rejected) [Bowker and Star 99]. The best way to overcome this problem is to develop a distributed knowledge management system which is based on the principles of autonomy and coordination, and it is rooted in a collection of knowledge nodes [Cuel, Bouquet and Bonifacio 05]. Although the distributed knowledge management is, in theory, the best approach, a lot of portals are developed according to a centralized vision. Therefore, innovative solutions should be developed to enable the encountering of users and their perspectives. In this work, we present a tool rooted into social network analysis (SNA) as a mean of innovation and knowledge coordination.

This paper is organized as follows: section 2 provides some basic concepts on innovation, coordination processes, and social network analysis; section 3 presents our case study, the Brazilian Innovation Portal; section 4 describes some previous results deriving from the application of social network analysis in the web portal; and section 5 depicts conclusions and future work.

2 The Innovation and Coordination Processes

It is a commonly accepted opinion that organizations tend to specialize their technologies – according to organizational studies we use the term technology as: the system of activities, processes, values, and all the resources needed to obtain organizational aims and changes [March and Simon 58; Cyert and March 63; Thompson 67] – in a relatively stable environment [March and Simon 58]. In contrast,

in unstable markets, organizations differentiate their technologies answering to the multiple and undetermined requests. Also, the more complex the environment is, the more units (with their technologies) should be managed within the firm [Ashby 56]. In other words, organizations manage external complexity with the proliferation and coordination of autonomous or semiautonomous units which manage knowledge in an autonomous and coordinated way as in distributed knowledge management systems [Cuel, Bouquet and Bonifacio 05]. Furthermore, the continuous and unpredictable encountering of different units (with their knowledge and technologies) are critical factors which enable the creation of unexpected and innovative combination of processes and products [Chandler 62; Brown and Duguid 91].

On the other side, the continuous encountering of different units might increase coordination costs (in terms of time and money). Therefore, coordination processes have been studied since long time, and a lot of dependencies have been unveiled among processes, technologies adopted within firms, and their organizational models [Galbraith 73; Rossignoli 04; Lomi 03].

2.1 Coordination processes

The relevance of coordination processes has constantly increased in organizational studies and a lot of its features have been unveiled and analyzed. For instance, Parson considers coordination processes relevant at any level of the organization: technical, managerial and institutional [Parson 51]. Other authors believe in information asymmetry [Simon 82; Williamson 75] and the encounter of different perspectives [Boland and Tenkasi 95] as two of the main characteristics of coordination processes. Since we are talking about organizations, and thus about a collective level, it is relevant to consider that – without any inter-subjective agreement (or at least believed agreement) – units cannot cooperate, communication cannot take place, coordinated action is impossible, and meaning remains connected just at an individual level [Weick 79]. According to knowledge network theories [Hildreth and Kimble 04; Cross and Parker 04], different and specialized actors which coordinate each others, move beyond the information sharing to the aggregation and creation of new knowledge, and obtain benefits from communications and engagement strategies. Finally, the network of relationships, the local knowledge developed among units, the inner motivation that drives people to share knowledge, and the knowledge they share, lead to the creation of an environment that sustains connections, processes coordination, creativity, and innovation.

2.2 Techniques of Social Network Analysis

Social network analysis (SNA) is an interdisciplinary methodology developed mainly by sociologists and researchers in social psychology during the 1960s and 1970s, further developed in collaboration with mathematicians, statisticians, and computer scientists that led to a rapid development of formal analysis techniques. SNA is rooted in the concepts of:

- “node” which refers to a subject, an office, a group, a community, etc.
- “connections” which refer to channels of communication, coordination, awareness, trust, and processes of decision-making, knowledge sharing, innovation, etc. [Scott 00; Ehrlich and Carboni 05].

Nowadays SNA is an attractive tool for other disciplines such as economics, marketing or industrial engineering, that helps researchers to analyze a network identifying strengths and inefficiencies of connections among nodes [Chan and Liebowitz 06; Krackhardt and Hanson 93; Cross et al. 03]. For instance, SNA applied in an organizational context may show where a specific collaboration is breaking down, where talents and expertises could be better leveraged (connecting them with other network nodes), where decisions are getting bogged down (when a node is overloaded), and where opportunities for diffusion and innovation are being lost. Through the SNA analysis, organizational officers may develop remedial actions to improve productivity, efficiency and innovation. These actions include roles and responsibilities changes to foster more effective patterns of communication, incentive programs to improve trust, etc.

3 The Brazilian Innovation Portal

The Innovation Portal is a Brazilian platform of electronic government components which promotes interaction processes among firms, scientific communities, universities, etc. The portal is based on a technological architecture described by Pacheco [03], that would support cooperation and innovation processes among economic and scientific partners operating in various sectors. Nowadays, the web portal collects information about 980 thousand experts, 3000 firms and 19 thousand registered research groups [Innovation 07].

The SNA tools, implemented in the Innovation Portal, apply algorithms based on Link Analysis techniques [Jensen 99; Lions and Tseytin 98]. They benefit from huge amount of data and are able to unveil new forms of relationships among portal users, industries, research groups, and Brazilian institutions. Portal members are connected by common research interests, papers, projects, and institution employments, but might be interested in setting up new and innovative collaborations. The visualization of a SNA network provides useful insights to researchers which may decide to concretize existing connections (e.g. common interests, aims, and goals) or investigate new ones [Gonçalves et al. 06].

Considering technical, managerial and institutional levels of coordination, SNA might support coordination processes through the cross-visualization of these three levels. For instance, the technical level is represented by a network of authors that wrote papers, the managerial and/or institutional levels may be represented by networks of institutions connected by employees which are involved in common projects or declare common interests. SNA can also enhance meetings among people and the encountering of their perspectives, due to the fact that it can show how specialized actors are connected, enabling them to investigate on what type of connections the SNA describes, and if suitable, to engage new collaborations. These connections show for instance, who works with whom, who works on which project, and who works in one or more research areas. We are aware that the resulting networks visualization doesn't allow us to directly derive that people are motivated to share knowledge and innovate, but at least it shows how people might interact.

4 Results

The Innovation Portal supports two differentiated environments which provide various services and human-computer interfaces for both experts (e.g. researchers, professors), and firms (public or private organizations and in particular institutes of science, technology and innovation).

Deeply analyzing the expert and firm environments, three types of networks can be created:

1. *networks of experts* may visualize co-authorship, co-participation in research projects, thesis co-advisory; board co-membership for thesis discussions. The co-authorship network (Figure 1) results from the analysis of users curricula (stored in the Lattes Platform [Lattes 07]), in which publications are described.
2. *networks of institutions* may visualize portal members which (i) have studied or worked in one organization (Figure 2), (ii) were born or live in the same region, (iii) work in common research areas.
3. *networks of research groups* may visualize connections among firms which work in specific geographical and research areas or projects.

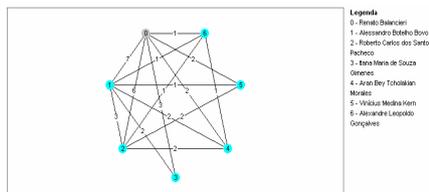


Figure 1: Co-authorship network

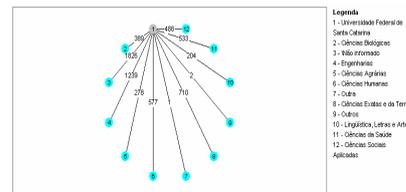


Figure 2: Network of institutions classified by research

Another relevant process supported by the Portal is the visualization of experts curricula. The process is the following one: the user looks at his/her network (for instance the one in Figure 2), then standing a few seconds on a specific tie (which describes the number of people affiliated to an institution), a list of experts pops up. At this stage the user can chose one name and see his/her curriculum vitae (Figure 3).

The Portal also supports processes of localization, contact and correspondence between firms and experts. In this way people can look for researchers (they never meet) that are working in specific institutions, which are involved in particular projects or research areas. First, people can search for experts, groups or firms which have particular competences or job opportunities. Second, they can select some of the resulting experts, groups and firms and save information about them in their favourite lists. Third, people may send proposals to the ones listed in their favourites or receive calls from other actors subscribed at the Portal. At the end, they may engage innovative collaborations. To monitor all the interactions that occur in the system, the Innovation Portal uses the “QualisInteraction System” (Figure 4), a service that allows people to analyze communication processes among experts and firms, focusing

on the list of contacted firms that have answered to the proposals; list of experts that returned the received proposals; list of failed proposals.

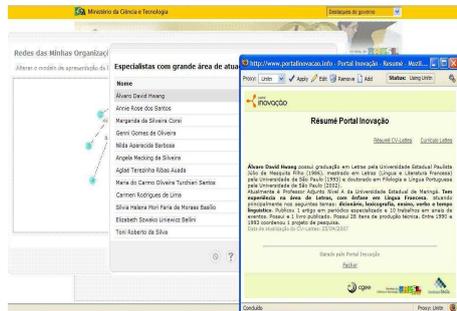


Figure 3: Experts network classified by research area
a) List of experts, b) Expert's curriculum vitae



Figure 4: Interaction between Experts and Firms

5 Conclusions and Future Work

This paper shows how the Brazilian Innovation Portal might support the encountering of different perspectives, enabling users to discover new links and connections with researchers, institutions, research groups and firms. In other words, users are enabled to develop innovative networks looking at SNA graphs from a different perspective. We believe that the continuous creation of unpredictable networks (through the SNA network visualization) might generate unexpected and innovative combination of research projects, processes and products.

The results of visualizing different levels of connectivity among experts and firms can help the Portal members to explore unexpected connections, unveiling new collaborations, knowledge and perspective encounters, and innovations. In the future the Portal will be updated with new services such as:

- an evaluation system that measures the Portal effectiveness on innovation processes;
- semantic networks that will be based on ontologies as semantic backbones for accessing information on the Portal, as well as for developing and maintaining the Portal.

Acknowledgements

The authors gratefully thank the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES) for supporting the investigation which we have documented in this paper.

References

- [Ashby 56] Ashby, W. R.: "An Introduction To Cybernetics"; John Wiley & Sons, NY (1956).
- [Boland and Tenkasi 95] Boland J. R., Tenkasi R. V.: "Perspective Making and Perspective Taking In Communi-Ties of Knowing"; *Organization Science*, 6, 4 (1995), 350-372.
- [Bowker and Star 99] Bowker, G. C., Star, S. L.: "Sorting things out: classification and its consequences"; MIT Press (1999).
- [Brown and Duguid 91] Brown J. S., Duguid P.: "Organizational Learning and Communities-Of-Practice: Toward A Unified View of Working, Learning and Innovation"; *Organization Science*, 2, 1 (1991), 40-57.
- [Chan and Liebowitz 06] Chan, K., Liebowitz, J.: "The synergy of social network analysis and knowledge mapping: A case study"; *International journal of management and decision making*, 7, 1 (2006), 19-35.
- [Chandler 62] Chandler, A. D.: "Strategy and Structure: Chapters in the History of the Industrial Enterprise"; MIT Press, Cambridge (1962).
- [Cross et al. 03] Cross, R., Parker, A. and Sasson, L.: "Networks in the Knowledge Economy"; Oxford University Press, New York (2003).
- [Cross and Parker 04] Cross, R., Parker, A.: "The Hidden Power of Social Networks: Understanding How Work Really Gets Done in Organizations"; HBS PBook (2004).
- [Cuel, Bouquet and Bonifacio 05] Cuel R., Bouquet P. and Bonifacio M. S., "A distributed approach to knowledge management, the concept of knowledge nodes, and their implications"; *Encyclopedia of knowledge management*, Schwartz D., Hershey, Pa: Idea group, 16 (2005).
- [Cyert and March 63] Cyert R. M., March J. G.: "A Behavioral Theory of the Firm"; Prentice-Hall, Engelwood Cliffs, N.J. (1963).
- [Davenport et al. 98] Davenport, T. H., De Long, D. W., and Beers, M. C.: "Successful knowledge management projects"; *Sloan Management Review*, 39, 2 (1998), 43-57.
- [Dinsmore 92] Dinsmore, J.: "Partitioned Representations"; Kluwer Academic Publishers, (1991).
- [Dougherty 92] Dougherty D.: "Interpretative barriers to successful product innovation in large firms"; *Organization Science*, 3, 2 (1992), 179-202.
- [Ehrlich and Carboni 05] Ehrlich, K., Carboni, I.: "Inside Social Network Analysis"; Technical Report 05-10 (2005).
- [Fauconnier 85] Fauconnier, G.: "Mental Spaces: aspects of meaning construction in natural language"; MIT Press (1985).
- [Galbraith 73] Galbraith, J. R.: "Designing Complex Organizations"; Reading, MA, Addison-Wesley (1973).
- [Ghidini and Giunchiglia 01] Ghidini C., Giunchiglia F.: "Local Models Semantics, or Contextual Reasoning = Locality + Compatibility"; *Artificial Intelligence*, 127, 2 (2001), 221-259.
- [Goffman 74] Goffamn, I.: "Frame Analysis"; Harper & Row, New York (1974).
- [Gonçalves et al. 06] Gonçalves, A. L., Beppler, F., Bovo, A. B., Kern, V. and Pacheco, R.: "A Text Mining Approach towards Knowledge Management Applications"; *Proc. of the*

Internacional Workshop on Information Retrieval on Current Research Information Systems, Denmark (2006), 7-28.

[Hildreth and Kimble 04] Hildreth, P., Kimble, C.: "Knowledge Networks: Innovation Through Communities of Practice"; Idea Group Publishing (2004).

[Innovation 07] "Innovation Portal"; May 2007, <http://www.portalinovacao.info/>

[Jensen 99] Jensen, D.: "Statistical Challenges of Inductive Inference in Linked Data"; Preliminary papers of the Seventh International Workshop on Artificial Intelligence and Statistics, Florida, (1999).

[Kuhn 79] Kuhn, T.: "The structure of Scientific Revolutions"; University of Chicago Press (1979).

[Krackhardt and Hanson 93] Krackhardt, D., Hanson, J. R.: "Informal networks: the company behind the chart"; Harvard Business Review, 71, 4 (1993), 104–111.

[Lattes 07] "Lattes Plataforma", May 2007, <http://www.lattes.cnpq.br/>

[Lions and Tseytin 98] Lions, D., Tseytin, G. S.: "Phenomenal Data Mining and Link Analysis"; AAAI Fall Symposium on Artificial Intelligence and Link Analysis (1998), 123–127.

[Lomi 03] Lomi, A.: "Organisation and strategy, dynamics and processing" (2003).

[March and Simon 58] March J. C., Simon H. A.: "Organization"; Wiley, New York (1958).

[McCarthy 93] McCarthy, J.: "Notes on Formalizing Context"; Proc. of the 13th International Joint Conference on Artificial Intelligence, Chambery, France (1993), 555–560.

[Pacheco 03] Pacheco, R. C. S.: "Uma metodologia de desenvolvimento de plataformas de governo para geração e divulgação de informações e de conhecimento"; Report UFSC, Florianópolis (2003).

[Parson 51] Parson, T.: "The Social System"; Free Press, Glencoe (1951).

[Rosignoli 04] Rosignoli C.: "Coordinamento e cambiamento: tecnologia e processi interorganizzativi"; F. Angeli (2004).

[Scott 00] Scott, J.: "Social Network Analysis: A Handbook"; Sage Publications Inc, Newbury Park, CA. (2000).

[Simon 82] Simon H. A.: "Models of Bounded Rationality"; MIT Press, Cambridge (1982).

[Thompson 67] Thompson J. D.: "Organizations in Action"; Mac Graw Hill (1967).

[Weick 79] Weick, E. K.: "The Social Psychology of Organizing"; McGraw-Hill, Inc., 1979.

[Williamson 75] Williamson O. E.: "Markets and Hierarchies: Analysis and Antitrust Implications"; the Free Press, New York (1975).

Information Exploration via Pen, Brush and Text Marker Multi Criteria Knowledge Management by Standard Drawing Tools

Hans L. Trinkaus

(Fraunhofer ITWM, D-67663 Kaiserslautern, Germany
hans.trinkaus@itwm.fraunhofer.de)

Abstract: The paper starts with the wish list for a “perfect” information exploration tool, where the topics of that list are collected from the work of some pioneers and experts in this field, as e.g. from C. Ahlberg, C. Williamson and B. Shneiderman. After that, a novel multi criteria knowledge management technique is introduced, that comes fairly close to the wish list given. This will be demonstrated by looking on some general aspects of information exploration, and how knowCube, a user-friendly software tool supporting graphical decision making, masters such tasks, where user interaction happens via standard drawing tools, like pen, brush or text marker. The paper ends with an outlook on FilmFinder - Version knowCube®.

Keywords: information exploration, multi criteria knowledge management, graphical queries, decision making, interactive navigation, software tool, pen, brush, text marker

Categories: H.3.3, H.3.4, H.4.3, H.5.2

1 Motivation, History and Goal

“Fight information overkill”, “Defeat the deluge”, “Drowning in information, and thirsting for knowledge”, or “Searching for needles in haystacks”. These phrases aren’t new, but nevertheless still adequate for describing the current situation, where about one million terabyte data is generated annually, and within the next years worldwide information will increase more than in man’s total history before [Keim 02]. So, any help for a “quick and easy handling-understanding” of this soaring mountain of data should be welcome.

Already C. Ahlberg, C. Williamson, and B. Shneiderman must have felt like lost in data space, when in 1992 they developed their concept of dynamic queries for information exploration [Ahlberg 92]. They further formulated some characteristics or topics, that an ideal information exploration method or tool should have or address [Williamson 92] [Ahlberg 94]. As an add-on they introduced a software prototype “FilmFinder” to demonstrate the operating principles of their ideas. In the sequel, other application fields were discussed, together with more and more exploration and visualization concepts and techniques [Keim 02] [Alvarado 03] [Teevan 04] [Klein 05] [Reiterer 05] [Jetter 05].

Together with the evolution of information exploration the collection of desires for a “perfect tool” increased accordingly. These “demands” are summarized in a kind of a wish list, which serves as the starting point for this paper. After that, it’s fairly easy to formulate the goal: Develop a method and implement a tool, with characteristics and features coming close to the wish list, passing the “FilmFinder-Test” very well, and above all, being user-friendly.

2 Characteristics of Visual Information Exploration

The following section introduces general statements, observations, demands and hints concerning the field of visual information exploration, as it evolved since the eighties. The collected set of 44 items may be seen (translated accordingly) as a “customer requirement specification” of an utopian software-tool for supporting information exploration. In Section 2.2 three versions of the application example “FilmFinder” show its evolution in time, and serve as a reference for knowCube[®] in chapter 4.

2.1 A Wish List for a Perfect Tool

24 years ago, Ben Shneiderman formulated in [Shneiderman 83] as a must, that direct manipulation interfaces for database retrieval support

- (1) continuous visual representation of objects and actions of interest,
- (2) physical actions or labelled button presses instead of complex syntax,
- (3) rapid, incremental, reversible operations whose impact is immediately visible,
- (4) layered / spiral approach to learning that permits usage with minimal knowledge.

All of these statements have been proved to be true, without any minor change. The same can be said for some other observations of Ben Shneiderman and Chris Rutkowski [Rutkowski 82], that look a bit closer to human cognition and behaviour (in man-machine interaction) – here formulated in short:

- (5) The user is able to apply intellect directly to the task; the tool itself “disappears”.
- (6) Users do not need to decompose tasks into syntactically complex sequences.
- (7) Objects of interest are displayed so that actions are directly in the high level semantic domain.
- (8) The closeness of the command action to the problem domain reduces user problem-solving load and stress.

The wish list for a perfect information exploration system evolved in time, 31 more topics had been added (that will be presented in another paper [Trinkaus 07]) up to a current one [Agne 06], with some further functionalities that should be provided:

- (40) Keyword search: Seeking for documents that contain a specific single term.
- (41) Wildcard search: Keyword search with dummy-characters.
- (42) Phrase search: Like keyword search, additional with phrases.
- (43) Proximity search: Enables for reorganizing and dispersing of phrases.
- (44) Adaptability: Tool assimilates to user behaviour.

2.2 FilmFinder – Version 1, 2, 3

A film database is a very good reference example for testing the principles of tools for supporting visual information. Everybody will be able to understand and discuss pro’s and con’s of different scenarios, and nearly everybody will be interested in it, a bit. (Databases having natural representations as a map – at least for a subset of their parameters – are no good candidates, because they “cloud” the general requirements.)

The first version of such a “FilmFinder” was presented by [Ahlberg 94]. Out of this paper two snapshots are taken, one (an overview on all movies in the data base) is shown here – another one (a zoom into the details) is presented in [Trinkaus 07]:

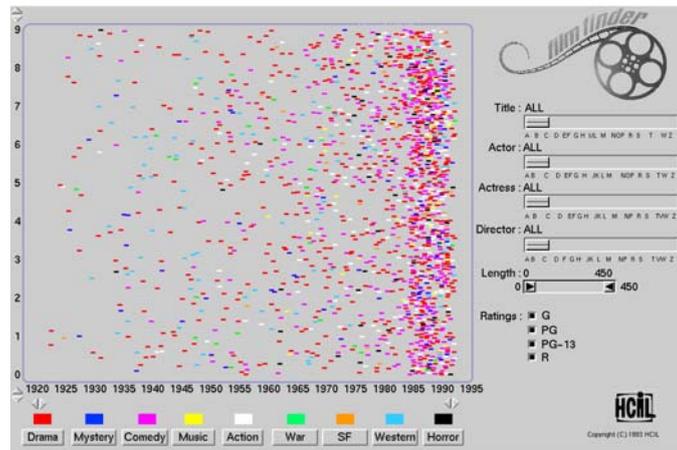


Figure 1: FilmFinder1 [Ahlberg, Shneiderman 94]

This graphic gives a look on the total database content, i.e. all films achievable therein. None of the control elements on the GUI puts some restrictions, so this may be seen as the start-up situation of the tool.

In 2005, two more visualizations were created to support human search behaviour and applied to the film finding scenario, too: the “CircleSegmentView” by [Klein 05], and the “HyperGrid” by [Jetter 05]:

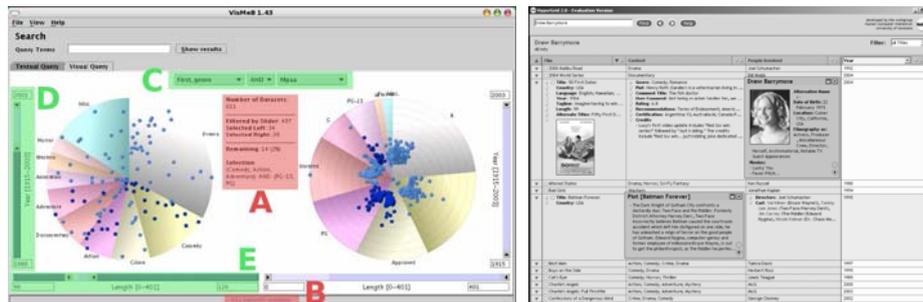


Figure 2: FilmFinder2 [Klein, Reiterer 05] + FilmFinder3 [Jetter et al 05]

FilmFinder2 offers an overview on the information space and enables the user to specify queries, and further to limit and filter the result sets to manageable sizes – but it seems to be a bit too complicated for non-expert users. FilmFinder3 combines well-known table visualizations with zoomable user interface features to provide a two-dimensional grid as an entrance tool to multi-dimensional and relational data. It's easy to use, yet it doesn't provide a graphical scope like the first two tools presented above.

There shall be given no further comments on advantages or shortcomings of these tools, the reader may check by the wish list topics above, what has been done well, and what is still lacking. Anyhow, it ought to be a fine goal to get a much more general tool, that works at least comparable good as these three candidates.

3 Multi Criteria Knowledge Management, Decision Making, and Information Exploration

The concepts discussed in the following fit to any of the disciplines Knowledge Management (KM), Decision Making (DM) or Information Exploration (IE), and the border lines between them are hardly to be seen. So, in the sequel, the terms KM, DM and IE will be used “at random”.

First, an “intermezzo” about the human cognitive system will be presented to motivate the concepts that are introduced in [Section 3.2], where they are explained additionally by their usage in an easily to grip application from the field of diamonds.

3.1 Human Cognition and Visual Data Processing

In [Trinkaus 05] some simple tests concerning human cognition of information are posed, observations are made, and conclusions are drawn. Here an example:

Let A, B, C, D, E be natural persons, sections in business enterprises, departments of universities, states, ...

- *A is positively affected by B and affects B, C and E positively.*
- *B is affected by A and C positively and affects D negatively and A positively.*
- *C is positively affected by A, negatively by E, and affects B positively.*
- *B and E negatively affect D.*
- *E affects C and D negatively and is positively affected by A.*

What’s going on?

Solution (– to be presented interactively with moving arrows):

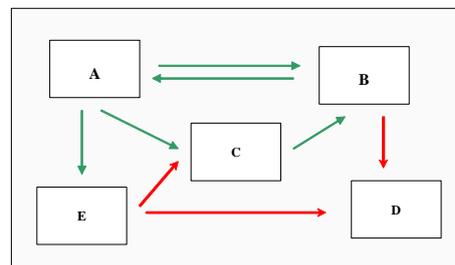


Figure 3: A Little Graphics Helps a Lot

Conclusion: The human visual system has a highly efficient information processing architecture. It allows inferences to be drawn based on visual spatial relationships – extremely fast and with hardly any demands on “working space capacity”, and it is also very sensitive for moving objects and changing shapes.

So, KM, DM, and IE tools should appeal to the human visual system, and transform complex situations into time-animated, spatial presentations.

Furthermore, basic principles of „identifying figures” (i.e. closedness, closeness, and nearly convexity) are discussed in [Trinkaus 05], too, and how to associate dynamic multi criteria situations with moving polygons – the starting point for knowCube®.

3.2 Looking for Nice Diamond Cuts

Diamonds aren't only girls' best friends, they are also a nice playing field to discuss knowCube[®], where in the current case at least three scenarios may come up:

- Everybody's view: Some general diamond knowledge has to be managed.
- Producer's view: Which shape(s) can be cut out of a given rough stone?
- Jeweller's view: How to balance demands, offers, shapes, prizes ...?

Everybody may look into [Trinkaus 07] to get some more information about the world of diamonds. Here, some explanations concerning the producer's/jeweller's view will be given, but formulated as a general setting:

Each property (parameter, criterion, variable, perspective) of an object (here: virtual diamond) is related to a line segment, the line segments are arranged like a "glory", each segment shows a scale (numerical, lexicographic, or something else). The property values of an object are mapped to corresponding points on the "axis", adjacent points are connected by straight lines, all together defining a closed polygon, that visualizes the object of interest. By this little trick the observer comprehends "what belongs together".

DM's scenario corresponds to [Figure 4]. Here, the producer has a rough stone given, i.e. known are its surface geometry e.g. via a point cloud, and its colour and clarity – things determined by nature (and explained in [Trinkaus 07] in detail). The market demands for round shaped diamonds! Then, the producer uses the scales pp1, pp2 and pp3 to restrict or select the parameter proportions of the diamond to be cut. The producer's intuition, a heuristic or some other algorithm calculates width-values, which could be realized for these proportions. Carat values easily are gained by applying Guldin's law, prices are estimated by their exponential dependence on carat weight, and calculating loss is trivial. The producer makes his/her "cut-decision", weighing out the visualized alternatives.

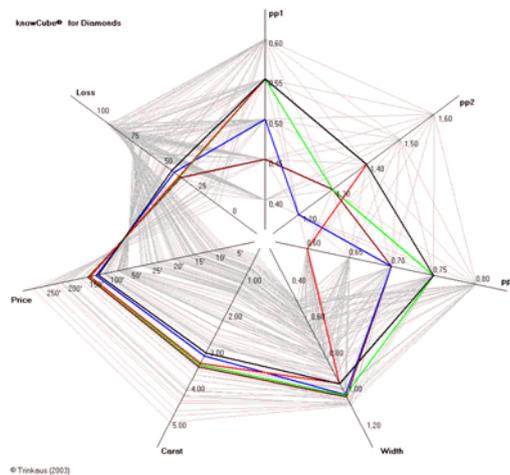


Figure 4: Producer's Point of View

Some "working principles" will be shown in the sequel via a jeweller's point of view.

4 Information Exploration by Graphical Objects and Intuition

Above, all properties of the diamond objects are “numerical” variables, mapped onto their axes in corresponding “naturally ordered” manner. But, the different scales shown there must not necessarily be linear, as e.g. the price axis. And furthermore, the property values on the axes, increasing from the centre towards the outside, also could be taken with opposite orientations. Besides some quantitative items, like *Carat* and *Price*, [Figure 5] presents also non-numerical properties. *Cut*, *Colour*, and *Clarity* are qualitative entities, with their corresponding ranked values ordered increasingly towards the centre, and with equidistantly placed potential “labels” on the axes. It should be remarked, that other orderings or scales also could be handled as well.

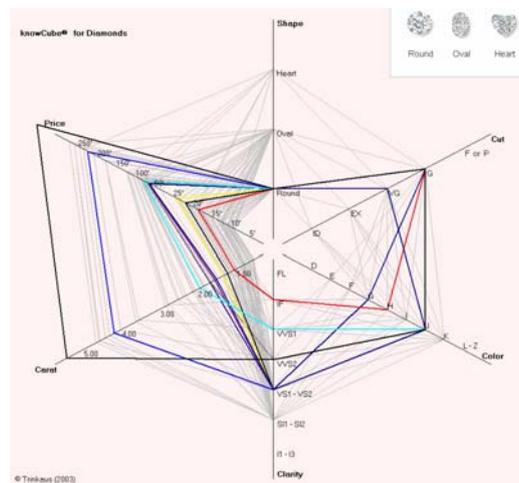


Figure 5: Jeweller's Point of View

Shape belongs to a third type of properties, where the values are abstract entities with arbitrary (not necessarily equidistant) placements, that don't show any kind of pre-defined ordering. Anyhow, such a visualisation method is particularly suited for e.g. distinguishing three clusters of “diamond alternatives”, namely the groups of *Rounds*, *Ovals* and *Hearts*. Later on, a lot of likewise, rather intuitional principles will be introduced, without any restrictions on quantitative variables, on linear scales, on numbers of objects, of criteria, ..., i.e. all those topics addressed by the wish list.

4.1 Brush, Text Marker, and Rubber – and AND and OR queries

Every child knows a *pen* or a *brush*, and (almost) everybody has some experience with a *text marker*, i.e. a German pseudo Anglicism for a felt-tipped pen, or a *rubber*. Then why not use these well known “tools” for information exploration, too?

Imagine, that a right mouse click on an axis in [Figure 5 or 6] opens a context sensitive pop up window, which then offers a selection among *text marker*, *rubber*, *hide*, *scale*, *weight*, etc. Choosing *text marker*, a further selection is presented, from which the *colour* for the *text marker*'s strip may be taken, and where e.g. *red* means

highest, and *blue* stands for lowest priority. (Also other property settings for the *text marker* may be fixed there, e.g. the *strip thickness*, the *opacity* of colour, etc.) Drawing a red colour strip on the *price* axis from 10' to 20' in the jeweller's or customers' point of view means, that diamonds lying within this price range are of highest interest. Likewise, a blue colour strip around 50' indicates, that such diamonds also could be asked for, but with much lower priority. These different colours, i.e. priorities, will get still more interesting with respect to interactive actions, like setting goals or finding k-nearest (weighted) neighbours via the *pen* tool. More details on this, and how to handle with the other selection items offered, like *hide*, *scale* or *weight*, will be explained in [Trinkaus 07], *rubber* follows in short.

By the way, from an abstract point of view this kind of "text marker query" on a database represents a generalized OR-operation. And, as a further comment, people who prefer "more precise queries" may alternatively move pairs of sliders on the axes – or type digits or select symbols from lists in corresponding input/output boxes:

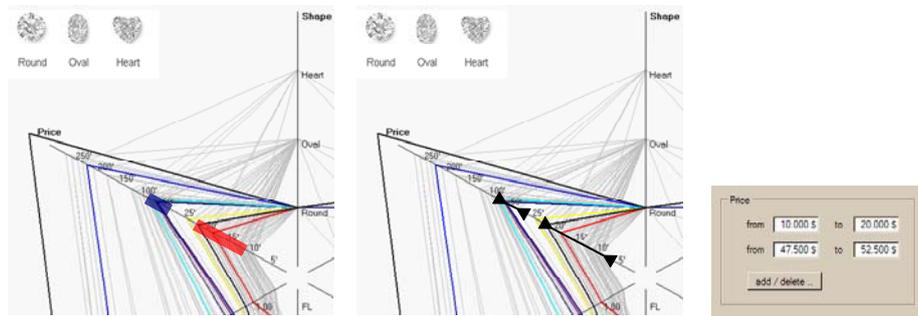


Figure 6: Text Marker Queries, Moving Sliders, or Standard Input/Output Boxes

A right mouse click in between two axes opens another context menu, which now offers a selection among *brush*, *pen*, etc. Choosing *brush*, a *colour* (coding priority), the *brush thickness* (coarse or fine selection purposes), and further properties may be set, according to the user's wishes. Then, for very specific queries, the information explorer brushes in [Figure 7] over such polygon parts, which come fairly close to the criteria looked for, where the jeweller has put special interest on three of the *heart* shaped diamonds, and additionally on some *oval* ones, that are not too expensive.

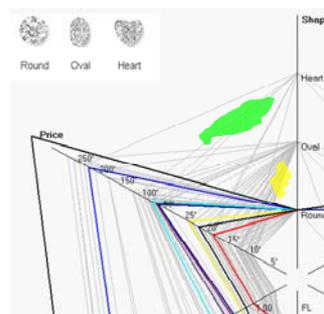


Figure 7: Brush Queries – Combinations of AND and OR

5 Future Work

The knowCube[®] software will be extended to include graphical features for: complex combinations of logical operators, handling criteria of not ordered type, dealing with missing or multiple parameter values, defining similarities and setting goals (via pen), prioritizing, weighting, zooming, hiding, ..., and FilmFinder - Version knowCube[®].

References

- [Agne 06] S. Agne, C. Reuschling, A. Dengel, DynaQ – Dynamic Queries for Electronic Document Management, German Research Center for Artificial Intelligence, Germany, 2006.
- [Ahlberg 92] C. Ahlberg, C. Williamson, B. Shneiderman, Dynamic Queries for Information Exploration: An Implementation and Evaluation, Proc. ACM CHI Conference, 619-626, 1992.
- [Ahlberg 94] C. Ahlberg, B. Shneiderman, Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays, Proc. ACM Conference on Human Factors in Computing Systems, Boston, pp. 313-317, 1994.
- [Alvarado 03] C. Alvarado, J. Teevan, M.S. Ackerman, D. Karger, Surviving the Information Explosion: How People Find Their Electronic Information, Massachusetts Institute of Technology – Artificial Intelligence Laboratory, AI Memo 2003-006, April 2003.
- [Jetter 05] H-C. Jetter, J. Gerken, W. König, C. Grün, H. Reiterer, HyperGrid – Accessing Complex Information Spaces, Proc. HCI, Springer-Verlag, 2005.
- [Keim 02] D.A. Keim, Information Visualization and Visual Data Mining, IEEE Transactions on Visualization and Computer Graphics 8(1), pp. 1-8, 2002.
- [Klein 05] P. Klein, H. Reiterer, The Circle Segment View – A Visualization for Query Preview and Visual Filtering, University of Konstanz, Department of Computer and Information Science, Germany, 2005.
- [Reiterer 05] H. Reiterer, G. Tullius, T.M. Mann, INSYDER: A Content-Based Visual-Information-Seeking System for the Web, Published online, Springer-Verlag, 2005.
- [Rutkowski 82] C. Rutkowski, An Introduction to the Human Applications Standard Computer Interface, Byte 7, pp. 291-310, 1982.
- [Shneiderman 83] B. Shneiderman, Direct Manipulation: A Step Beyond Programming Languages, IEEE Computer, 16, 57-69, 1983.
- [Teevan 04] J. Teevan, C. Alvarado, M.S. Ackerman, D.R. Karger, The Perfect Search Engine is not Enough: A Study of Orienteering Behavior in Directed Search, Proc. ACM CHI Conference, 415-422, 2004.
- [Trinkaus 05] H.L. Trinkaus, T. Hanne, knowCube: A Visual and Interactive Support for Multicriteria Decision Making, Computers and Operations Research 32, 1289-1309, 2005.
- [Trinkaus 07] H.L. Trinkaus, Information Exploration via Pen, Brush and Text Marker: Multi Criteria Knowledge Management by Standard Drawing Tools, DASMODO Cluster of Excellence Proceedings, 35 pages, to appear, 2007.
- [Williamson 92] C. Williamson, B. Shneiderman, The Dynamic Home Finder: Evaluating Dynamic Queries in a Real-Estate Information Exploration System, Proc. ACM SIGIR Conference, 339-346, 1992.

ExpertFinder: Collaborative Expertise Localization

Willem-Olaf Huijsen

(Telematica Instituut, The Netherlands
Wolf.Huijsen@telin.nl)

Samuël Driessen

(Océ-Technologies, The Netherlands
Samuël.Driessen@oce.com)

Dion Slijp

(Océ-Technologies, The Netherlands
Dion.Slijp@oce.com)

Abstract: Straightforward expertise localization is crucial for personal and organizational efficiency and productivity. Common issues with existing approaches are the amount of effort required to build up the database and keeping it up-to-date, and the difficulty of establishing a complete coverage of the organization. The ExpertFinder system employs the concepts of referral chaining, social networks and user-generated data to enable a fast, low-effort and thereby low-cost approach to building an expertise localization database. At the same time, ExpertFinder provides a number of views on this database. We describe the design of the system and discuss in detail two pilots conducted at Océ-Technologies and the Telematica Instituut, the findings of which show high user participation and a good coverage of individual employee expertise and overall organizational expertise.

Keywords: collaborative software, expertise localization, mass collaboration, referral chaining, social networks, user-generated content.

Categories: H.5.2, H.5.3, J.4.

1 Introduction

Finding a person in the organization with a certain expertise is crucial for personal and organizational efficiency and productivity. However a new employee who just started working for a company, does not have a personal network yet. Also, in large companies (more than 100-150 employees), one cannot know everyone. Therefore, supporting the mapping of knowledge is important.

How do employees normally find experts? They tap into their own social network: the people they know. Research has shown that the number of people one knows (meaning we know the person and his/her expertise) is 100-150 people [Cross & Parker, 04]. Some organizations use internal yellow pages systems that help employees find basic information about other employees: name, phone number, email address, office location, etc. Most yellow pages also have (optional) slots in which employees can fill in their hobbies and expertise that can be used for expertise localization. Last but not least, employees use their social network to find the right person [Wenger, 98]. They ask a colleague whether they know someone with the

required expertise. If that employee does not know anyone, usually he/she will point to another employee that can possibly help find the expert.

How do employees decide who is an expert on a certain topic? In practice, this may be based on the fact that one knows a person, that the person exposes him/herself as an expert by talking about the subject and answers questions, and also by whether one likes the person; whether he/she fits in your network. Alternatively, deciding who is an expert can be done more objectively by judging the data and information an employee has produced about a certain topic. This can also be automated. The data and information stores can be mined using an algorithm, which results in a list of experts.

2 ExpertFinder

ExpertFinder is an expertise localization system that uses a collaborative, low-cost and quick method for obtaining an up-to-date overview of the expertise that is present within an organization. It taps into the subjective notions with respect to experts. The system relies on collaborative, quick assessments of employees of their own knowledge levels and on their assessments of the knowledge levels of their colleagues [Cross & Parker, 04, Ichijo & Nonaka, 07].

ExpertFinder can be used to quickly get an overview of the knowledge within an organization with a minimum amount of effort. The system uses the informal social networks people have with a method called referral chaining [Kautz et al., 97, Yimam-Seid & Kobsa, 02]. The system starts with one relevant topic and an ExpertFinder organizer who knows an employee knowledgeable on this topic. This person is sent an e-mail form (Figure 1). This form asks the addressee what his/her knowledge is about the topic and if he/she knows colleagues who might know something about the topic. These people will subsequently be sent a mail too and be asked the same questions (if they have not already received an email for this topic). In this way, a chain reaction is triggered that quickly maps out people and their knowledge level on a specific topic. This cycle is repeated until no new references remain. This process is executed (in parallel) for a large number of subjects. In addition to the subjects that the system is seeded with at the start, employees can also enter new subjects themselves, for which this process will then be executed as well.

In this way, ExpertFinder builds an overall picture of the collective knowledge: What kind of expertise is present in the organization and who is an expert on a certain topic? This knowledge may be technical expertise, but also competence, interpersonal skills, and knowledge of application areas.

Note that the effort required from individual employees is fairly low: It is limited to filling out, for each expertise that they receive an e-mail for, their own level of expertise and the colleagues they know to be knowledgeable about the subject as well.

Subject: Océ Expert Finder Query: LDAP
 From: Expert Finder
 Date: 9-9-2003 12:18
 To:

Expert Finder Query for LDAP

Please help building up the expert database and fill in this form below. Thanks.

How much knowledge do you have on the topic 'LDAP'?

Nothing Little Much Expert

Who else knows anything about this topic? (Enter initials, separated by commas)

If you can't read HTML mail or have any other problem, go to [this page](#)

For more information see [the Expert Finder website](#)

Figure 1: E-mail message from the Océ ExpertFinder system.

Every e-mail also contains a link to the ExpertFinder website, which presents all gathered information from the responses, and gives referral networks per expertise (including referral frequencies). This gives an indication how many employees see this person as an expert. The website allows you to see, for every subject, a ranked list of colleagues knowledgeable on that subject, including their contact details. In addition, a number of related subjects are shown. These subjects are determined on the basis of the similarity of the lists of experts.

Receiving an ExpertFinder e-mail stimulates a receiver's intrinsic motivation and co-operation. Receiving an e-mail works as feedback and signals that someone thinks you have valuable knowledge on a topic, which gives the receiver a feeling of being recognized and competent. This feeling of recognition and competence stimulates the receiver to fill in the form and thereby keep the chain reaction going [Ryan & Deci, 00]. In contrast, other systems require people to fill in seemingly endless lists of questions or systems. Consequently, people do not receive immediate feedback and are therefore not stimulated to cooperate [Macy, 91].

3 ExpertFinder at Océ

Océ has a yellow pages system containing basic information about all its employees: name, phone numbers, picture, office location, etc., and, optionally, hobbies and expertise (10% fill in these two slots, of which 2% keep it up-to-date). With ExpertFinder these slots can be automatically filled and be kept up-to-date.

ExpertFinder was originally developed at Océ R&D. A prototype was created for testing the concept within the Research department. This pilot was set up in an adhoc way and was not controlled. We did not explain to the employees what ExpertFinder is for except for the information in the email they received. No formal management approval was given for ExpertFinder. Running the pilot went well for a few hours, until someone accidentally referred to another department. From then on the pilot escalated. ExpertFinder rolled through the organization as a growing snowball and

within a week almost 500 R&D employees (out of a total of 1100) submitted roughly 2500 email forms on approximately 180 topics. After a month the number of active users stabilized at 680. No form of authentication is required to submit information to the system. In practice this constituted no problem.

Later we integrated ExpertFinder in our yellow pages to automatically fill the expertise slot. ExpertFinder was not integrated into the existing yellow pages system.

The key findings from the Océ pilot are:

- Nearly all reactions to the pilot were positive. The few complaints were about people not liking the number of e-mails. People could opt to be put on a blacklist so as to never receive email from Expert Finder again.
- Employees need support to find experts and are willing to spend a couple of seconds to fill in an email form about their and other's expertise level.
- Employees like to be seen as an expert on certain topics.
- The expertise and the level of expertise is quite accurate. The experts per expertise were recognized as their level of expertise.
- ExpertFinder should be integrated in the existing yellow pages to improve visibility and usability of the tool. E.g. Océ users do not want to go to yellow pages for certain information and to ExpertFinder for expertise information.
- Keeping the database up-to-date with easy-to-use mechanisms is essential for the relevance and usage of the ExpertFinder.
- The referral network graphs can be sensitive. The graph can show that a person sees her/himself as an expert, but that nobody else does.

4 ExpertFinder at the Telematica Instituut

The ExpertFinder pilot at the Telematica Instituut was conducted in October 2006, as part of the Metis project [Metis]. The pilot adapted the original system from Océ-Technologies to TI and ran for a full month. The pilot set-up consists of the interaction with employees and the adaptation, and monitoring of the software.

Communication to launch and promote ExpertFinder comprised e-mails and flyers to all individual TI employees, flyers posted on several bulletin boards, and a news flash on the TI intranet. Managers of various departments were asked to stimulate participation in the pilot. Reminder e-mails were sent out. For every individual employee, the system was seeded with a known subject of interest, and the system subsequently sent out the first e-mails. Throughout the pilot, the TI community was supported with quick, appreciative and helpful responses to questions (both face-to-face and e-mail). We also asked colleagues about their opinion on ExpertFinder. The pilot was aimed at all TI employees, both scientific and support staff, both regular staff and temporary staff such as trainees and graduating students. The pilot did not cover contacts and project members at other organizations. The pilot covers all work-related subjects deemed interesting to share: expertise, competences, and application areas.

The ExpertFinder software was originally developed at Océ. For the TI environment, a number of adaptations were made. The most significant ones are: (1) Integration into TI's knowledge-mapping system, the Knowledge Cockpit [Huijsen

et.al., 04, Driessen et.al., 07]. (2) Instead of a single, flat list of subjects as in the original Océ ExpertFinder, the subjects were divided into three knowledge categories: Expertise, i.e. knowledge on work-related subjects such as human-machine interaction, XBRL, accounting and computer security; competence, i.e. job skills such as giving presentations, writing business cases and programming. A competence is an activity that one can be good at.; application area, i.e. an area in which one has experience such as education, health care and consumer electronics. (3) Referring to colleagues has been extended from just specifying the colleague to also specifying their perceived level of expertise. (4) The inclusion of a FAQ and definitions for knowledge categories (see (2)) and the four knowledge levels used to rate expertise, for greater clarity and consistency.

Expert List for ARCHITECTURES

expertise

the fundamental organisation of a system embodied by its components, their relationships to each other and to the environment and the principles guiding its design and evolution. [IEEE] ([edit description](#))

Related subjects: [CONTEXT AWARE APPLICATIONS](#) - [SOFTWARE ARCHITECTURES](#) - [MOBILE APPLICATIONS](#) - [SERVICE ORIENTED ARCHITECTURES](#) - [MODEL-DRIVEN DEVELOPMENT](#) - [METAMODELS](#)

[Add more experts](#) for this subject

Name	Email	Telephone	Level	Reference levels
[redacted]	[redacted]	[redacted]	very much	4 very much
[redacted]	[redacted]	[redacted]	very much	3 very much
[redacted]	[redacted]	[redacted]	very much	2 very much
[redacted]	[redacted]	[redacted]	very much	1 much, 1 very much
[redacted]	[redacted]	[redacted]	very much	1 much, 1 very much

Figure 2: Employees with expertise on topic "architectures".

User Details for [redacted]

These are the knowledge levels as submitted by the person him/herself

	expertise	competence	application area
much	INFORMATION OVERLOAD KNOWLEDGE MANAGEMENT KNOWLEDGE TRANSFER		
little	CONTRACTUAL IPR		
not entered (yet)	ASP CRM SERVICE ORIENTED ARCHITECTURES KNOWLEDGE MAPPING	PROJECT MANAGEMENT	SPORTS

[Back to the main page](#)

Figure 3: Knowledge profile for an individual user.

The evaluation of this pilot consists of a presentation of the statistics on the one hand and a review of the feedback on the other hand.

For the pilot, we found that after one month, 72% of TI employees had contributed, 524 of a total of 590 ExpertFinder e-mails had been responded to (73%). On average, employees had been associated with 6,6 subjects, and 12,1 references had been made by each employee. At the outset, ExpertFinder was seeded with 74 subjects. Over the course of the pilot, TI employees added another 51 subjects. From this, we conclude that we had a high response rate, and that people actively contributed, also in adding new subjects.

During and after the pilot we received feedback various questions and comments, the most significant of which we will now discuss.

Referring to colleagues: When users want to refer to others, they get a list of all TI colleagues with for each the following interface:

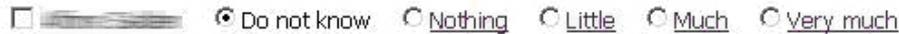


Figure 4: User interface for referring to colleagues.

Users may wonder how what the meaning is of the checkbox as relation to that of the radio buttons. In hindsight, this is a user interface design mistake. The intended use is that the users check the box to indicate that they want to refer to the colleague in question, and then specify their perceived knowledge level for this colleague on the given topic. Observe that the latter implies the former, so that the check box is actually superfluous. In practice, by not checking the check box, any value specified with the radio button will be lost. In this way valuable user input may have been lost.

Confirmation e-mails can be annoying: Confirmation e-mails are sent for every change to users' personally indicated knowledge levels in their knowledge profile. This is necessary to guarantee that the profiles cannot be changed by others without the user knowing, since the current implementation has no user authentication.

Too many e-mails: Users are sent separate e-mails whenever others suggest they are knowledgeable on a given topic. This may add to the prevalent information overload issue. A possible approach to minimize this is to buffer the requests and send only one e-mail a day that combines all of the day's requests.

Interpretation of knowledge levels: Although the knowledge levels have been given explicit definitions and examples are provided, it is likely that different people may have different interpretations.

Few complaints about e-mails. There were almost no complaints about the e-mail messages: We expected to receive a significant amount of feedback objecting to the e-mail messages from ExpertFinder.

A small survey was conducted to get a more objective idea of the quality of the gathered information in the ExpertFinder system. We interviewed one of TI's expertise group leaders and a member of TI's scientific staff. Below we list their observations.

Q: Does ExpertFinder provide a good representation of the expertise at TI?

A: I do recognize the subjects mentioned, and it seems to be a fair overview. However, it is difficult to say for sure. It is hard to get a mental overview of expertise subjects, since this is a list of all kinds of subjects, at different levels of abstraction. Clustering of the subjects would help answering. For my expertise group, ExpertFinder provides a good overview. For example, for the subject "social tagging" all group members are listed that I would expect. The overview for other expertise groups also seems to be fairly good. The overview seems to cover TI's subjects pretty well, although I suspect there to be some areas/subjects missing. It is striking to see that many expertise subjects are listed; the expertise subjects for knowledge management are well covered (ExpertFinder being a KM tool); the competence

"research" is not listed (maybe it is too obvious to mention); and the competence "consortium building" is not listed (maybe too high-level a concept to mention).

Q: Is the list of experts accurate / are the personal expertise profiles accurate?

A: ExpertFinder provides good personal expertise profiles, even for some people for whom only references by others are available. Some have clearly gone to some length to fill out their profiles, while others have not responded at all. Still, for the latter, the personal profile (consisting solely of subjects associated with them by others) is fairly good. It is interesting to find that for the subjects you can find more people than only the well-known experts. For some subjects (such as Web 2.0) very few people are mentioned, and I would expect other people to be in the expert list.

5 Conclusions

As for the pilot at the Océ and Telematica Instituut, the pilot's results show that first, there is a high level of participation, which indicates that the ExpertFinder approach has a good user interaction and good user acceptance. Second, the gathered information on expertise is fairly good (good personal profiles and a good coverage of the overall organizational expertise), obtained quickly and with little effort. A practical drawback is that the overview is a bit rough, but this is inherent to the approach, and this is significantly offset by the quick and low-cost results.

6 Future Work

For actual deployment of ExpertFinder-like functionality in real-life settings, especially in large organizations, the following issues are of great interest.

How can the expertise information be kept up-to-date? ExpertFinder is only as up-to-date as the last time the information has been entered. At any time, people may alter their knowledge levels themselves. However, ExpertFinder does not include a mechanism itself to keep it up-to-date. A naïve approach would be to repeat the ExpertFinder pilot at regular intervals, for example twice a year, sending active users an email form to adjust their knowledge level on all topics they participate in. With this form they can also easily refer to new people who have knowledge about the specific topic. However, this approach is likely to lead to a decreasing participation due to questionnaire fatigue. Alternatively, the system could be adapted to include a decay of information; that is, as soon as entered information reaches a certain threshold age, the original user that entered it is asked to confirm or change that information. This approach minimizes the user burden while guaranteeing the information to be up-to-date. An additional mechanism would be needed to make sure new emerging subjects appear. This could be done by suggesting high-ranking terms from the document analysis of an internal document report system to the employees associated with the corresponding documents.

What options can the user choose from? Extensive research has been done on how to design surveys. Without much research we chose to offer four options. Research should be done if this amount is correctly chosen.

How to deploy ExpertFinder? In order to deploy ExpertFinder, especially in large organizations, it is important to make sure that the same subject name means the same

thing to all employees. This could be addressed by putting a greater emphasis on the subject description by for example making it mandatory or by selecting people as "owner" of the subject and making them responsible for the subject definition.

How should the user interface be enhanced? As ExpertFinder is a pilot system, the current user interface is quite basic. Various improvements can be made, the most significant of which are the following: (1) Improving user authentication through a login mechanism or possibly by checking IP addresses. Currently, ExpertFinder requires users to identify themselves. The risk of deliberate or accidental incorrect identification is mitigated by the fact that users receive a confirmation message for any change to their personally indicated knowledge levels. (2) Support for profile management: Currently, users can manage their profile primarily through the e-mails they received over a period of time, i.e. subject by subject. Users may expect the knowledge profile page to include support for managing their profiles.

Acknowledgements

This research was carried out within the Metis project on knowledge management [Metis]. We thank the following colleagues who were involved in the ExpertFinder pilots: Marcel Bijlsma, Marjan Grootveld, Joost Jager and William van Dieten.

References

- [Cross & Parker, 04] Cross, R., Parker, A.: "The Hidden Power of Social Networks. Understanding How Work Really Gets Done in Organizations"; Harvard Business School Press, Boston, 2004.
- [Driessen et al., 07] Driessen, S.J., Huijsen, W., Grootveld, M.: "A Framework for Evaluating Knowledge-Mapping Tools"; *Journal of Knowledge Management*, 11, 2 (2007), 109-117.
- [Huijsen et al., 04] Huijsen, W., Driessen, S.J., Jacobs, J.W.M.: "Explicit Conceptualizations for Knowledge Mapping", *Proc. Sixth International Conference on Enterprise Information Systems (ICEIS 2004)*, Vol. 3, 231-236, Porto, April 2004.
- [Ichijo & Nonaka, 07] Ichijo, K., Nonaka, I.: "Knowledge Creation and Management"; *New Challenges for Managers*, Oxford University Press, New York, 2007, 37 and 240.
- [Kautz et al., 97] Kautz, H., Selman, B., Shah, M.: "Referral Web: Combining Social Networks and Collaborative Filtering"; *Comms. of the ACM*, 40, 3 (1997), 63-65.
- [Macy 91] Macy, M.W.: "Chains of Cooperation: Threshold Effects in Collective Action"; *American sociological review* [0003-1224], 56, 6 (1991), 730-747.
- [Metis, 07] The Metis project, <http://www.telin.nl/index.cfm?language=en&id=250>.
- [Ryan & Deci, 00] Ryan, R.M., Deci, E.L.: "Self-determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-being"; *The American psychologist*, 55, 1 (2000), 68.
- [Wenger, 98] Wenger, E.: "Communities of practice. Learning, meaning, and identity"; Cambridge University Press, 1998.
- [Yimam-Seid & Kobsa, 02] Dawit Yimam-Seid & Alfred Kobsa: "Expert Finding Systems for Organizations: Problem and Domain Analysis and the DEMOIR Approach"; 2002, citeseer.ist.psu.edu/490565.html.

An Orchestration Model for Knowledge Management Tools in SMEs

Christian Ploder

(University of Innsbruck, Austria
christian.ploder@uibk.ac.at)

Kerstin Fink

(University of Innsbruck, Austria
kerstin.fink@uibk.ac.at)

Abstract: The main idea of this paper is the combination of the following two declarations: (1) There is a variety of different models, tools and techniques in the area of knowledge management, but there is no framework for finding the right tool in combination of the strategic alignment of the company and the technical/cultural infrastructure. (2) A lot of research in the field of knowledge management is done for large companies or international concerns. A very interesting and important kind of business is neglected: the small and medium-sized enterprises. This paper shows a model for describing the flexible process finding the right tools and techniques with the adequate focus on IT support for a SME to efficient implement knowledge management cost-efficient.

Key Words: knowledge management, knowledge process, SME, SOA

Category: A.0, H.1

1 Introduction

Knowledge exists in the mind of people and circulates within organizations [Nonaka and Takeuchi 1995]. For all companies knowledge management (KM) and innovation plays an important role [Porter 2001]. Realizing the organization of knowledge is a strategic corporate asset which has to apply to future organizational problems [Schulttze and Stabell 2004]. Many of KM initiatives are implemented in large companies but in the last years there is a shift towards small and medium-sized enterprises (SMEs) [Fink and Ploder 2006a] [Fink and Ploder 2006b]. SMEs need a simple framework to organize their key knowledge processes: (1) knowledge identification, (2) knowledge acquisition, (3) knowledge distribution and (4) knowledge preservation [Fink and Ploder 2007]. For SMEs it is necessary that the advantages of KM are clear and the implementation is easy otherwise they will continue to focus on their traditional way of working [Nunes et al. 2006]. Based on different research projects in the area of KM [Fink and Ploder 2006a] in SMEs the following lack could be found [Fink and Ploder 2007]: Austrian, German and Swiss SMEs want to use methods for implementing KM and support this with existing software products but

others would base their methods on the organizational culture.

This paper presents results of a study that was conducted at the department of information systems at the University of Innsbruck in winter 2006/2007 with the aim to build a model to find out which tools/techniques are right for a special SME depending on the organizational strategy.

2 Background

The market situation of the European countries is almost identical: More than 95% of the companies are SMEs. SMEs in the European Union employ more than 60% of all employees [ASME 2006, CHSME 2006]. Also the OECD SME Outlook shows a similar situation [OECD 2006]. There are many qualitative and quantitative definitions of SME. For this paper, the definition of the European Council is used: (1) **micro enterprises** are companies with less than 10 employees and an annual turnover less than 2 Mio. EURO or a total asset of less than 2 Mio. EURO, (2) **small enterprises** are companies with less than 50 employees and an annual turnover less than 10 Mio. EURO or a total asset of less than 10 Mio. EURO and (3) **middle enterprises** are companies with less than 250 employees and an annual turnover less than 50 Mio. EURO or a total asset of less than 43 Mio. EURO. For this work SMEs are defined by the number of employees less than 250 employees because it is often not possible to find out the other data. Earlier research was done in the area of finding out the key knowledge processes for SMEs which are shown in figure 1. This four knowledge processes are the most important for SMEs [Fink and Ploder 2007].

All of the four key knowledge processes can be supported by different tools and techniques. Which of the knowledge management tools and techniques are used by SMEs is discussed in [Fink and Ploder 2006a].

For each SME it will be an issue to find the right tools and techniques in order to their knowledge goals and their environment. A further aspect is the development of the SMEs. Finding the first mix of tools and techniques for supporting the four knowledge processes is important but the mix should change over time with the environmental changes - so the use of tools and techniques should be flexible. It is important to react on changes in the culture or the IT-architecture concerning to the changes in the market. The aim of the empirical study was to find out which tools and techniques are supported by the organization culture and which are IT supported.

3 The empirical study

3.1 Framework

Before starting the empirical study it is necessary to define the framework for the research. This paragraph explains the two important concepts of (1) culture

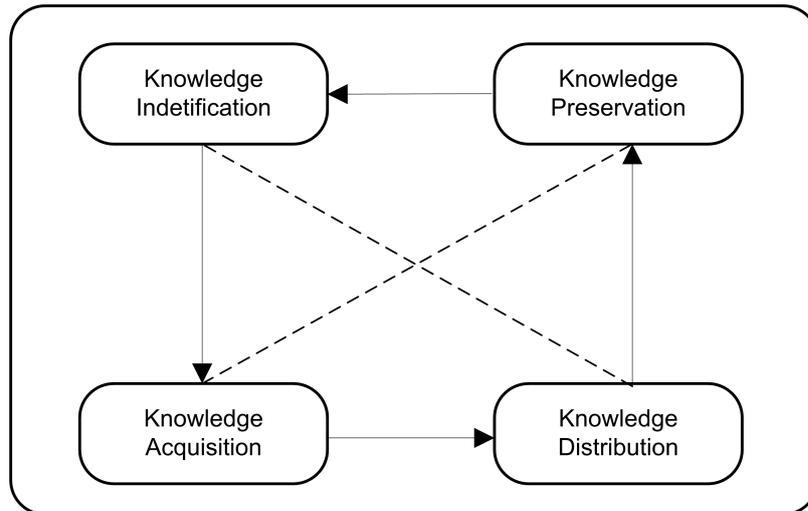


Figure 1: Key Knowledge Processes for SMEs

orientation and IT based (2) service orientation.

Deal and Kennedy define corporate culture as "the way we do things around here" [Deal and Kennedy 1982] or culture has "something to do with the people and unique quality ... of organization" [Kilman et al. 1985]. These factors of culture are important for the association with knowledge management. If the culture is not compatible to knowledge sharing there would not be a possibility to implement knowledge management.

The basis of service oriented architecture (SOA) are the services. A service can be defined as "a repeatable task within a business process" [Marks and Bell 2006] and they can communicate with each other. A service oriented architecture can be characterized by the following three items: (1) Following the principle of **orientation by components** all different systems can communicate with all other systems, (2) a **loose hyphenation** leads to asynchron communication and (3) a **workflow engine** controls the activities [Siedersleben 2006].

The differentiation into culture oriented and service oriented is based on the integral knowledge management model by [Beijerse 2000] which is adapted by the authors.

3.2 Data Sample

The department of Information Systems of the University of Innsbruck conducted 23 expert interviews with scientists (14) and practitioners (9). The experts for the interviews were opted stochastically. Experts from science were

chosen after a Online Research of Research Areas of members of Information Systems departments and Informatics departments in the German speaking countries. Experts of practitioners were chosen out of the pool the department is working with them in different projects because these experts know about the topic. For the empirical study a pool of tools and techniques was presented. The used method was expert interview [Flick 2006] with the intention to assign tools and techniques of knowledge management to hard and soft categories. The experts had to rank each tool/technique on a likert scale (from 1 - culture oriented up to 5 - service oriented). The collected data are analyzed and visualized in figure 2.

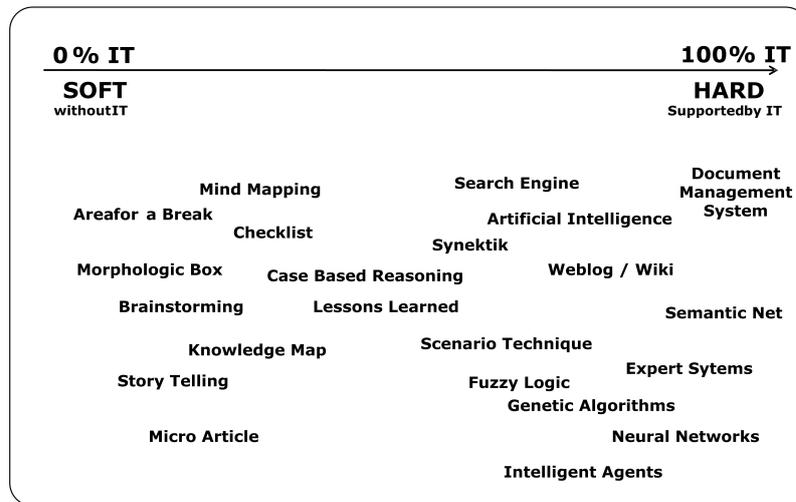


Figure 2: Categorized Tools and Techniques

Figure 2 shows some of the assessed tools and their allocation for use. Not all of the tools are adequate for the use in SMEs but there is no difference of their orientation if they will be used in large companies or in SMEs. For example the tools "brainstorming" and "area of a break" are good examples for tools with less IT support. So these tools are interesting for enterprises with a low level IT infrastructure and can be supported by the corporation culture. "Document Management Systems" (DMS) or "neural networks" are examples for tools with a very high support of IT infrastructure.

4 The Orchestration Model

During the expert interviews it became clear that the requirements to a knowledge management system change over time. So more than 80% of the experts asked for a model which can be adapted over time but introducing strategic knowledge goals of the company. Therefore the ranked methods build the base of the layer concept shown in figure 3.

Developing a model for supporting the choice of the right mix of tools/techniques for efficient knowledge management in SMEs cannot take place in an operationalized view. An important factor for implementing a knowledge management system is the strategic consideration by the knowledge goals [Probst et al. 1999]. By the view of a top down approach the first layer in the model (see figure 3) is the **Strategic Layer**. The SMEs should define their knowledge goals for a medium-term consideration. Looking at current and future potential of knowledge in a company shows a similar situation than looking on the strategic alignment of IT systems. There are four different roles of knowledge in a business (in dependence on [Heinrich and Lehner 2005][Fink 2000]):

1. knowledge as support: There is a low current strategic position of knowledge and in future too.
2. knowledge as manufactory: Knowledge plays an important role at the moment but the importance will decrease in future.
3. knowledge as breakthrough: In future knowledge will play a very important role for the business still it does not.
4. knowledge as weapon: Knowledge is important and the importance will increase in future.

Subjected to the strategic situation (four quadrants at the strategic layer) of knowledge goals of the business and the present and future power of knowledge some of the four key knowledge processes are more important or less important than others. This does not mean that only a few processes will be important but the attention will not focus on all of the four processes. So it is necessary to concentrate on one or two of the processes instead of trying to implement all four key knowledge processes at the same time in a SME. Every quadrant is matched with a priority of the four key knowledge processes and so the way of implementation "process by process" can be described. This is shown in the **Knowledge Process Layer**.

To support the four key knowledge processes with an efficient tool or technique it is necessary to select these by considering the corporation environment. The two limits are (1) culture oriented for tools without any IT support and (2) service oriented for tools with high IT support. The process should be supported

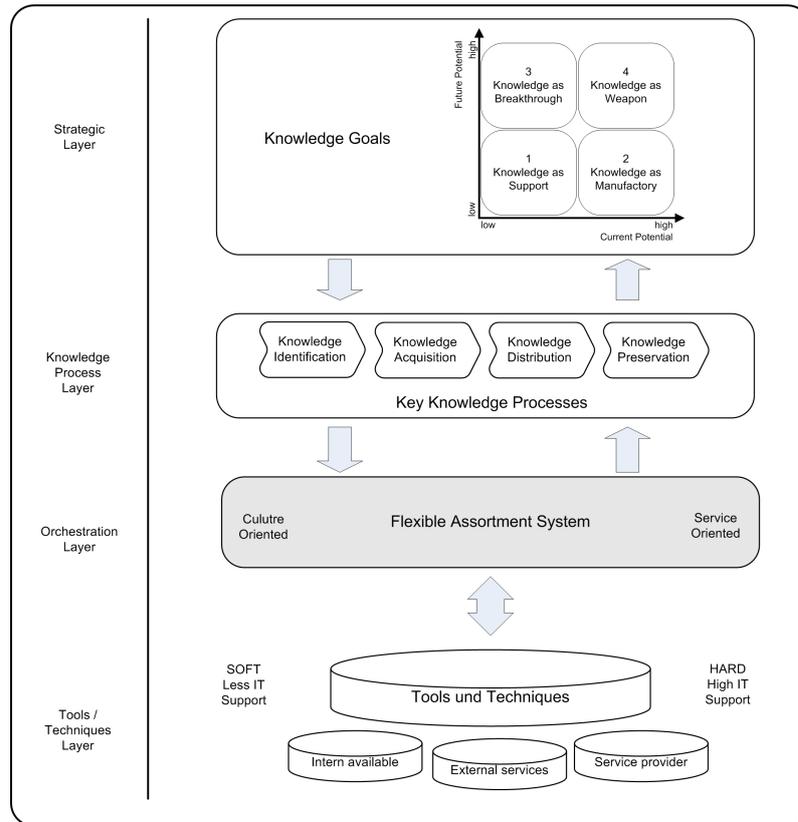


Figure 3: Model for flexible Assortment of the KM tools

by a software tool which should be integrated in the **Orchestration Layer**. This layer has two objectives: (1) Asking the user about the current situation of the corporation environment and assist the user at the classification by culture or service orientation. (2) Give a recommendation of the efficient methods for the special case considering the corporation environment and the strategic knowledge goals. The basis for the orchestration is built by the description of the different processes. Therefore the WS-BPEL (Business Process Execution Language) language can be used for formally expressing the operational characteristics of the orchestration processes [Erl 2006]. Using one of the knowledge processes and the adjustment of the business in direction of culture oriented or service oriented, it would be able to define templates describing different situations with a set of tools/techniques.

The **Tools/Techniques Layer** contains the repository of tools and techniques which are efficient applicable in SMEs. A differentiation can be done: (1) **inter-**

nal available are tools and techniques which can be supported internal and the knowledge already exists in the company, (2) **external services** will be used if tools and techniques are not available internal and must be provided external and can than be implemented internal without any service from externals and (3) services from **service providers** are necessary for implementing tools/techniques. Also a combination of the three differentiated repositories is imaginable. Tools and techniques are stored with all important facts, short introductions to each tool/technique and each tool/technique is labeled with meta data for a better arrangement. The different services should be described in a standardized way like the WSDL (Web Service Description Language) [Erl 2006].

5 Key Results and Future Research

The orchestration model of knowledge management tools shows up a present way for SMEs to find the right mix of tools and techniques to establish a knowledge management system. For SMEs it is necessary to use a simple knowledge process model with the four key knowledge processes and have a look at the current corporation environment. Otherwise they will fail at the implementation in many cases.

Future Research will deal with the design and implementation of a prototype for the orchestration software. Therefore it will be necessary to look on different possible services and their assortment. The assortment should be represented with UDDI (Universal Description, Discovery and Integration) [Marks and Bell 2006]. This will be done on a service oriented approach.

References

- [ASME 2006] Austrian Institut for SME Research, URL: www.kmuforschung.ac.at, Date 12-12-2006.
- [Beijerse 2000] Beijerse, R.: Knowledge management in small and medium-sized companies: knowledge management for entrepreneurs, *Journal of Knowledge Management*, Vol. 4 No. 2, pp. 162-179.
- [CHSME 2006] SME-Portal of Switzerland, URL: www.kmu.admin.ch/index, Date 12-12-2006.
- [Deal and Kennedy 1982] Deal, T., Kennedy, A.: *Corporate Cultures: The Rites and Rituals of Organizational Life*, Addison-Wesley, Reading, MA.
- [Erl 2006] *Service-Oriented Architecture: Concepts, Technology and Design*, Prentice Hall, Upper Saddle River, NJ.
- [Fink 2000] *Know-how Management*, Oldenbourg Verlag, München.
- [Fink and Ploder 2007] Fink, K., Ploder, C.: *Knowledge Process Modeling in SME and Cost-efficient Software Support - Theoretical Framework and Empirical Studies* In Khosrow-Pour, M.: *Managing Worldwide Operations and Communications with Information Technology*, Idea Group Publishing, New Jersey, 152-155.
- [Fink and Ploder 2006a] Fink, K., Ploder, C.: *The Impact of Knowledge Process Modeling on Small and Medium-sized Enterprises*. In Maurer, H.: *Proceedings of the I-Know '06*, Graz, 47-51.

- [Fink and Ploder 2006b] Fink, K., Ploder, C.: A comparative Study of Knowledge Processes and Methods in Austrian and Swiss SMEs In Proceedings of the ECIS 2007, St. Gallen, Date of publication: June 2007.
- [Flick 2006] Flick, U.: An Introduction to Qualitative Research. Second Edition. Sage Publications Ltd., London.
- [Heinrich and Lehner 2005] Heinrich, L., Lehner, F.: Informationsmanagement, Oldenbourg Verlag, München.
- [Kilman et al. 1985] Kilman, R., Saxton, M., Saxton, R.: Gaining Control of the Corporate Culture, Jossey-Bass, San Francisco, CA.
- [Marks and Bell 2006] Marks, E., Bell, M.: Service Oriented Architecture: A Planning and Implementation Guide for Business and Technology, Wiley, New Jersey.
- [Nonaka and Takeuchi 1995] Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company; Oxford University Press, Oxford.
- [Nunes et al. 2006] Nunes, M., Annansingh, F., Eaglestone, B., Wakefield, R.: Knowledge management issues in knowledge-intensive SMEs, Journal of Documentation, Vol. 62 No. 1, pp. 101-119.
- [OECD 2006] OECD: OECD small and medium-sized Enterprises and Entrepreneur Outlook. URL: www.oecd.org/document/, Date 12-12-2006.
- [Porter 2001] Porter, M.: Strategy and Internet, Harvard Business Review, Vol. 79 No. 3, March, pp. 62-78.
- [Probst et al. 1999] Probst, G., Raub, S., Romhardt, K.: Managing Knowledge: Building Blocks for Success. John Wiley & Sons.
- [Schulttze and Stabell 2004] Schulttze, U., Stabell, C.: Knowing what you don't know? Discourses and contradictions in knowledge management research, Journal of Management Studies, Vol. 41 No. 4, pp. 549-573.
- [Siedersleben 2006] Siedersleben, J.: SOA Revisited: Komponentenorientierung bei Systemlandschaften, Wirtschaftsinformatik, No. 29(2007), Sonderheft, pp. 110-117.

Knowledge Discovery Techniques Applied to Knowledge Management in Universities

Mihaela Oprea

(University Petroleum-Gas of Ploiesti, Department of Informatics, Romania
mihaela@upg-ploiesti.ro)

Irina Tudor

(University Petroleum-Gas of Ploiesti, Department of Informatics, Romania
tirinelle@yahoo.com)

Ana Tănăsescu

(University Petroleum-Gas of Ploiesti,
Department of Economy Mathematics and Economy Informatics Romania
atanasescuro@yahoo.com)

Abstract: The evolution of our society to the knowledge based society has raised new challenges for most of the scientific domains that exist. The higher importance given to knowledge extraction instead of getting just information (i.e. data included in a context) has lead to the development of several intelligent techniques for knowledge discovery. This paper shows some examples of using the techniques of case-based reasoning and data mining for knowledge discovery in the knowledge management system of an university. We have taken as example, the educational domain with the particular case of universities as they represent good examples of organizations that acquire, generate, store and use knowledge for various purposes, teaching, learning, and research.

Keywords: Knowledge Management, University Knowledge Management, Knowledge Discovery Techniques, Data Mining, Case Based Reasoning

Categories: H.3.1, H.3.2, H.3.3, H.3.7, H.5.1

1 Introduction

In the last years, knowledge management (KM) became a strategic issue for most of the private and public organizations, due to its highest potential for gaining efficiency and creating value in a knowledge based society.

Knowledge management refers to the ways in which organizations gather, manage, and use the knowledge they need in order to function with higher efficiency and benefits [O'Leary, 98]. The particular case of knowledge management university is very important as the educational domain represent one of the best examples of organizations that acquire, generate, store, and use knowledge for various purposes such as teaching, learning, and research. The higher importance given to knowledge extraction instead of getting just information (i.e. data included in a context) has lead to the development of several intelligent techniques for knowledge discovery [Probst, 00]. The paper shows some examples of applying two intelligent techniques, case-

- a software system for the administrative activities (e.g. an accounting software system, a personnel and salary software system),
- a software system for the admission of students in different programmes of study (undergraduate, MSc, postuniversity, PhD, postdoctoral), as well as for the students activity tracking during their studies, and after graduation (e.g. their work place, the continuous training activity).

As knowledge discovery is one of the most important subtasks of knowledge management we shall present in the next two sections some knowledge discovery techniques, and their applications in university knowledge management.

3 Knowledge Discovery Techniques

The purpose of a knowledge discovery technique is to find patterns, relations, in large databases and in the information flows of the system in which it is applied. Among the current used techniques, we have chosen two of them, case-based reasoning and data mining as being knowledge discovery techniques proper for the university knowledge management. The reason of our choice is that universities collect and store large amount of data and information related to their students, for example, and also accumulate experience from their past teaching and research activities.

3.1 Case-based reasoning

Case-Based Reasoning (CBR) is a model of reasoning which consists in solving new problems by adapting solutions that were used to solve old problems [Kaster, 05]. CBR is both a paradigm for computer-based problem solvers and a model of human cognition. The central idea is that the problem solver reuses the solution from some past case to solve a current problem [Ceccaroni, 01]. As a paradigm for problem solvers, one of the advantages of CBR systems is that they improve their performance, becoming more efficient, by recalling old solutions given to similar problems and adapting them to fit the new problems. As a model of human cognition, CBR is the basic cognitive process by which people solve problems and get about in the world. CBR represents an important technology for designing intelligent systems that is used to retrieve and reuse past solutions for resolving similar problems [Andone, 05]. CBR is ideally suited to the creation of knowledge management systems because of the close match between the phases of the CBR process model and the requirements of a knowledge management system [Watson, 01]. The structure of a CBR system is given in figure 2.

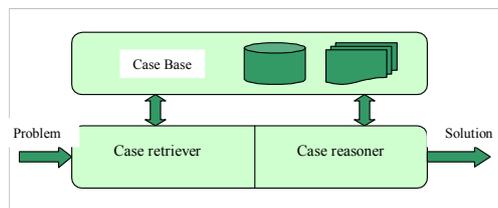


Figure 2: The structure of a CBR system

CBR is a good technique to solve new problems based on previous experience. When working with labelled cases, the retrieval step in CBR process model can be seen as a classification task. The new cases will be classified with the label (class) of the most similar case retrieved from the case base. Most case-based reasoners use a generalized weighted dissimilarity measure such as,

$$diss(C_i, C_j) = \frac{\sum_{k=1}^n w_k * adiss(C_{ik}, C_{jk})}{\sum_{k=1}^n w_k}$$

where: C_i and C_j are two cases; w_k is the weight or importance assigned to attribute k ; $adiss(C_{ik}, C_{jk})$ is the dissimilarity degree between the value of attribute k in the cases i and j .

3.2 Data mining

Data mining (DM) is the automated or convenient extraction of patterns representing knowledge implicitly stored in large databases, data warehouses, and other massive information repositories [Han, 01]. Some researchers view data mining as simply an essential step in the process of knowledge discovery in databases. Data mining is a multidisciplinary field, working in areas that includes database technology, artificial intelligence, machine learning, neural networks, statistics, pattern recognition, knowledge based systems, knowledge acquisition, information retrieval, high performance computing, and data visualization [Hand, 01].

Depending on the DM approach used, techniques from other disciplines may be applied, such as neural networks, fuzzy and/or rough set theory, knowledge representation, inductive logic programming, or high performance computing. Depending on the kinds of data to be mined or on the given DM application, the DM system may also integrate techniques from spatial data analysis, information retrieval, pattern recognition, image analysis, signal processing, computer graphics, Web technology, economics, or psychology [Han, 01]. The most important tasks of DM are representing by classification, prediction and clustering. The most popular DM techniques are decision trees, neural networks, rule induction, and nearest neighbour.

A *decision tree* is a flow-chart-like tree structure, where each node denotes a test on an attribute value, each branch represents an outcome of the test, and tree leaves represent classes or class distributions. At each node, the algorithm chooses the "best" attribute to partition the data into individual classes. Decision tree algorithms, such as ID3 and C4.5, were originally intended for classification.

A *neural network* is a collection of linear threshold units that can be trained to distinguish objects of different classes. Neural nets may be used in classification problems (where the output is a categorical variable) or for regressions (where the output variable is continuous).

Rule induction is one of the major forms of data mining and is perhaps the most common form of knowledge discovery in unsupervised learning systems. Rule induction systems are highly automated and are probably the best of data mining techniques for exposing all possible predictive patterns in a database. In order for the rules to be useful there are two pieces of information that must be supplied as well as the actual rule: accuracy (how often is the rule correct?) and coverage (how often does the rule apply?).

Nearest neighbor is a prediction technique that is quite similar to clustering - its essence is that in order to predict what a prediction value is in one record look for records with similar predictor values in the historical database and use the prediction value from the record that it "nearest" to the unclassified record.

4 Applications

Our current and future research work involves the application of the two techniques, CBR and DM in the knowledge management system of an university, and the analysis of their efficiency in improving the quality of the educational and research activity management. While DM will extract rules from large databases stored by the university, the CBR technique can help in retrieving old good solutions from past experience of the university management system. In this section, we shall present one application of CBR to university knowledge management for distance learning (as it was reported in the literature, and which could be extended by using a DM system), and one application of DM use in the academic environment, that we have developed at the University Petroleum-Gas of Ploiesti.

4.1 Case-based reasoning application

See Yourself Improve (SYIM) [Tsinakos, 02] is a domain independent educational environment for the providing of personalized educational services during asynchronous distance education sessions. Asynchronous distance education is one of the most appealing instructional delivery methods as it combines flexibility to access teaching material with time to reflect, self-study techniques with peer-to-peer collaboration and the use of low-cost technology. But such a system reveals a number of pedagogical drawbacks. To solve these problems it was used a CBR technique. The CBR technique aim was to automate the process of replying to the student's questions, by identifying relevant ones that have been already asked by other students and are stored in system's Educational Knowledge Base (EKB). When a student posts a query on the SYIM system, the posting is not sent directly to the tutor, but instead, the query is recorded in the General Cases Data Base. Instead of having tutors posting replies directly to students, the Process of Identification of Similar Question (PISQ), is triggered to find a relevant case that answers a given question. The PISQ searches among the contents of the system's EKB, where questions of educational value are recorded, stored and retrieved. The PISQ provides relevant answers to questions submitted by students via an easy to use, intuitive key word search feature. Responses to students' key word searches are based on the semantic and pragmatic aspects of the question asked. For these reasons, PISQ employs two search methods: a controlled vocabulary search and free text search among the contents of the Educational Knowledge Base. SYIM, also, automatically displays to students relevant chapter(s) of the instructional material and a list of appropriate keywords.

The CBR application that was briefly presented can be extended (for example, by combining CBR with a data mining technique), and used in the educational management system of an university.

4.2 An application of data mining in the academic environment

We have developed an application of DM in Weka (a Java based DM software [Weka, 02]), for the research activity analysis at the University Petroleum-Gas of Ploiesti (UPG). The algorithm applied upon a data set for estimating the research quality of academic staff is ID3 algorithm. In figure 3 we have presented the training data.

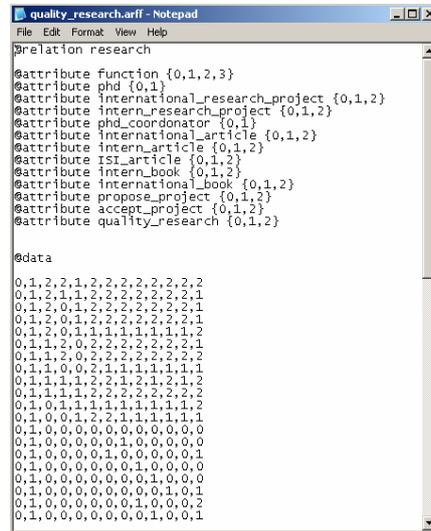


Figure 3: The quality_research.arff file

We have used thirteen attributes for problem modeling presented in Table 1.

Attribute \ Value	0	1	2	3
Function	Professor	Lecturer	University Lecturer	Assistant
Phd	Yes	No	-	-
International_research_project	No projects	Insignificant	Significant	-
intern_research_project	No projects	Insignificant	Significant	-
phd_coordinator	Yes	No	-	-
international_article	No articles	Insignificant	Significant	-
intern_article	No articles	Insignificant	Significant	-
ISI_article	No ISI articles	Insignificant	Significant	-
intern_book	No books	Insignificant	Significant	-
international_book	No books	Insignificant	Significant	-
propose_project	No projects	Insignificant	Significant	-
accept_project	No projects	Insignificant	Significant	-
quality_research	Low	Medium	High	-

Table 1: The correspondent value for the attributes used in problem modeling

The type of these attribute is nominal with the value {0, 1, 2, 3}, {0, 1, 2} or {0, 1} according with the significance of each attribute. The quality_research attribute

represents the target of the application, that expresses a measure of research activities. After compiling the ID3 algorithm for the training data, a batch of statistical results were obtained: confusion matrix, the precision, the measure F, the TP rate, the FP rate, the ROC area and the kappa statistic (figure 4).

The data set representing the input was provided by the research report containing the scientific work during the last three years regarding the Faculty of Letters and Science of UPG.

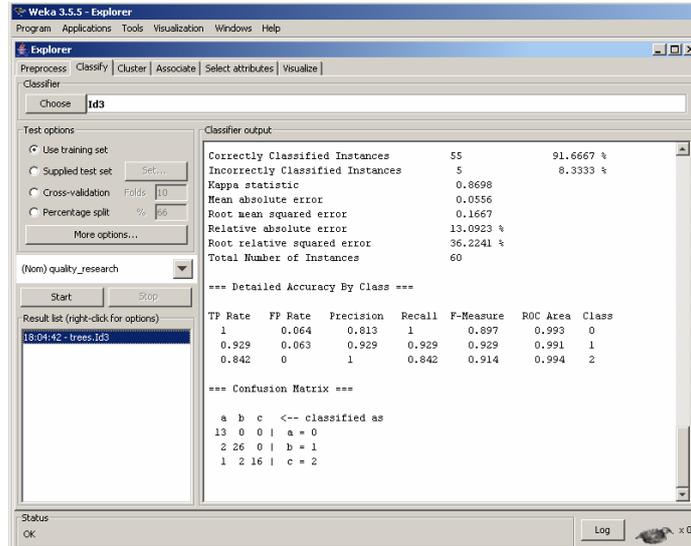


Figure 4: The results of Weka execution algorithm

In this case are three classes, and therefore a 3x3 confusion matrix is generated. The results obtained highlight that a person with a high quality research rate is a person who manages international projects, a person with a significant number of articles published in ISI journals, a person that is a PhD scientific coordinator, a person with a significant number of international scientific communication participation. The most important quality criteria are represented by scientific articles, particularly the articles published in ISI reviews. A quality management procedure was proposed to improve the quality of scientific research activities. On the other hand, the increase number of professors that coordinates the PhD scientific activities implies an increase quality of research. Future work must focused on involving academic staff in much more research activities like participation to international project competitions, publishing in ISI journals, coordinating the PhD degrees, writing books in international publishing houses.

5 Conclusions

The implementation of KM approaches in universities has become more and more a reality. In this paper, we have described two knowledge discovery techniques, CBR and DM techniques, and we have shown two examples of their applications in

university knowledge management. The first one, dedicated to teaching activities management, showed a CBR system applied in distance education, while the second one, dedicated to research activities management showed a DM system based on Weka software and applied in our university. The solution of incorporating both techniques, CBR and DM, in the UKM system can improve the quality of the decisions taken by the university. The two techniques can be combined with other artificial intelligence techniques and methods (such as, expert systems, other machine learning techniques) in order to increase their efficiency. While DM will extract rules from large databases collected and stored by the university, CBR can help in retrieving good solutions from past experience of the UKM. As a future work, we shall analyse the efficiency of using both techniques, CBR and DM in improving the quality of the teaching and research activity management in an university.

Acknowledgements

The research work reported in this paper was funded by the Romanian Education and Research Ministry under the research project CEEEX nr. 95/645/7/2006.

References

- [Andone, 05] I. Andone, Use of CBR and SCBR in Intelligent Systems, The 7th International Conference on Informatics in Economy IE 2005 – Information & Knowledge Age, Bucharest, May 19-21, 2005, 891-897.
- [Ceccaroni, 01] L. Ceccaroni, OntoWEDSS – An Ontology Based Environmental Decision - Support System for the Management of the Wastewater Treatment Plants, Ph.D. Thesis, UPC, Barcelona, Spain, 2001, <http://www.lsi.upc.es/%7Eluigi/thesis.pdf>.
- [Han, 01] J. Han, M. Kamber, Data Mining – Concepts and Techniques, Morgan Kaufmann Publishers, San Francisco, 2001.
- [Hand, 01] D. Hand, H. Mannila, P. Smyth, Principles of Data Mining, MIT Press, 2001.
- [Kaster, 05] D.S. Kaster, C.B. Medeiros, H.V. Rocha, Supporting modeling and problem solving from precedent experiences: the role of workflows and case-base reasoning, Environmental Modelling & Software, **20**(6), June 2005, 689-704.
- [Mikulecká, 00] J. Mikulecká, P. Mikulecký, University Knowledge Management – Issues and Prospects, research report, University of Hradec Králové, 2000.
- [O’Leary, 98] D.E. O’Leary, Knowledge Management Systems: Converting and Connecting, IEEE Intelligent Systems, May/June 1998, 30-33.
- [Probst, 00] G. Probst, S. Raub, K. Romhardt, Managing Knowledge: Building Blocks for Success, John Wiley & Sons, 2000.
- [Tsinakos, 02] A.A. Tsinakos, Distance Teaching using SYIM educational environment, Learning Technology, IEEE Computer Society, 2002.
- [Watson, 01] I. Watson, Case-Based Reasoning and Knowledge Management: a Perfect Match? In, Proc. FLAIRS-2001. 14th Int. FLAIRS Conference, Key West Florida May 21-23 2001, AAAI Press, Menlo Park CA, USA, 118-123.
- [Weka, 02] Weka software <http://www.cs.waikato.ac.nz/ml/weka/>

Utilizing Lifecycle Information for Knowledge Document Management and Retrieval

Lasse Lehmann, Tomas Hildebrandt, Christoph Rensing and Ralf Steinmetz

(KOM Multimedia Communications Lab

Technische Universität Darmstadt

Merckstr. 25, 64283 Darmstadt, Germany

{lasse.lehmann, christoph.rensing, ralf.steinmetz}@kom.tu-darmstadt.de)

Abstract: Classical approaches to document management do not cope with the demands knowledge documents make. Knowledge documents or so-called "living documents" have a far more complex lifecycle than general documents. They are usually used, edited and utilized by several people and many different versions, revisions and variants exist. Today the multitude of information that these processes generate are not captured or used to provide a better management or retrieval for this kind of documents. Our approach shows that the capturing of this lifecycle information can help in the retrieval as well as usage and management of those documents.

Keywords: Document Management, Lifecycle Information, Document Lifecycle, Retrieval

Categories: H.3.1, H.3.2, H.3.4

1 Introduction and Motivation

Imagine the following scenario: A lecturer at a university wants to make slides for his new lecture. Thus he searches for related material and finds lectures from other professors and lecturers that cope with similar topics. He copies some of the retrieved slides directly into his new presentation, while other parts are edited or reviewed before being put into the new lecture slides. He removes parts from the retrieved slides and adds new slides. When he is done he puts an acknowledgement at the end of his slides to give credits to those he got the original slides from. This is a very common scenario, so that it can be assumed that the professors our lecturer got his slides from proceeded in a similar way during the creation of their slides. There are several things the authors of the other slides would appreciate to know. To name a few of them:

- Who is reusing my slides?
- Who has changed the slides in which way?
- In which context are the slides re-used?
- How often are the slides re-used?
- Are there lecturers teaching similar topics and do my slides match theirs?
- How popular are my slides?
- ...

If the lifecycle information for all these documents would have been collected during the aforementioned processes, these questions could be answered easily. On top of that there would be a whole network of information and relations connecting the various versions and variants of the different lectures and slides for the different authors.

Similar scenarios can be found in a corporate context. Here, knowledge workers correspond mainly with other employees of the same enterprise. However, the lifecycle information that can be gained can be of great help here, too. E.g. if one employee has a specific task to do which is new to him, he searches for related material, which he can learn from. Often, the knowledge he is searching for is distributed over a wide range of different documents, some of them in places, where the employee might not have thought of looking at. The lifecycle information helps him to find and structure the information and knowledge he needs and allows him to browse the network of relations connecting the documents.

We will show an approach to collect the lifecycle information for knowledge documents and provide possibilities and concepts for the utilization of the collected information. At first we will define our concept of knowledge documents and analyse their lifecycle and the information that can be collected herein in section 2. Section 3 deals with the capturing and utilization of lifecycle information. Section 4 shows a concept and architecture for a system that implements our approach while section 5 deals with related work. Section 6 concludes this article and gives an outlook at future work.

2 Knowledge Documents and Their Lifecycle

First of all, we need to make clear how we comprehend the concept *knowledge document*. We do not want to constitute an exact definition for knowledge documents. In our understanding it is more like we found a group of documents with certain attributes and tried to find a concept to name this type of documents. Our definition is based on the well known DIKW model - e.g. [Davenport, 99], which states that simple characters, data, information, knowledge and wisdom stand in a defined hierarchical order with accumulating constraints. The exact constraints are perceived differently in existing literature; in our definition, characters become data when syntax is added, data becomes information when context is added and information becomes knowledge when the cross-linking of information is added. Starting there from, a knowledge document is a document with the following attributes:

- It contains knowledge (in contrast to containing data or information only)
- It is mainly text-based but can be a compound of different media and text
- It has a processible format (could be proprietary, like .doc or .ppt, too)
- It is often a result of a collaborative authoring process
- It is re-used or re-purposed (and undergoes the according processes)

Therefore knowledge documents are a subset of documents in general. In our understanding typical knowledge documents are, depending on the context, e.g. lecture slides, technical reports, documentations, papers, scientific works or internal

reports like shown in the aforementioned examples. Existing literature often refers to "living documents" [Berndt, 05] which is a quite similar concept from our point of view, but even vaguer than the term "knowledge document". Learning Objects can be knowledge documents as well, but due to their loose definition it is not possible to state that all learning objects are knowledge documents. We do not consider documents like bills, schedules, calendars or other documents often used in business processes as knowledge documents, since they contain mainly information.

2.1 Context Information

We distinguish two types of lifecycle information. **Context information** is related to one document. It covers the different contexts the document traverses during its history, e.g. applications that opened the document, users who accessed the document, the number of times a document was searched for, downloaded or retrieved respectively as well as the context of creation or usage (in terms of used applications durations of use etc.). Context information can be generated implicitly. This kind of information can be collected automatically by monitoring the used applications. On the other hand context information can also be user-generated, like e.g. feedback users give to the authors of certain documents, annotations to documents or parts of documents or even reviews or ratings.

2.2 Relation Information

The second type of lifecycle information we consider is **relation** information. Relations connect two or more knowledge documents and occur as result of explicit user actions where new instances of that knowledge documents emerge. While capturing these relations, we do not want to solely capture the fact that a relation exists, but also the type of relation connecting the two instances. We defined a specific set of relations for SCORM [ADL, 07] compliant learning objects [Lehmann, 07]. However, these relations can not be transferred to the domain of knowledge documents without modification. Thus, a future step will be the definition of types for relations emerging between knowledge documents in the course of their lifecycle. The lifecycle itself is analysed in the following.

2.3 Lifecycle Model

There are several existing models for the lifecycle of documents in general - most of them from the area of Information Lifecycle Management (ILM). As the name implies, do these models handle documents containing *information*. As stated before, we want to consider documents containing *knowledge*. Such, the prerequisites for lifecycle models are different. In the area of living documents or knowledge documents there are very few approaches for the modelling of the lifecycle. In [Ginsburg, 99] a waterfall model for the "Intranet Document Lifecycle" is presented. However, it does not focus on the systems and phases important for knowledge documents. In the area of e-Learning there are some approaches like [Brooks, 06] or [McCalla, 04] that at least state that the lifecycle has to be taken into account. A specific lifecycle model for learning objects is presented in [Collis, 04]. Since knowledge documents are a subset of documents in general, the lifecycle model for knowledge documents should be compatible with the existing lifecycle models from

ILM or digital libraries. However, it should provide a meaningful view on the relevant phases of the lifecycle from different perspectives. This enables us to analyse the information being generated or utilized during the different phases.

Figure 1 shows a schematic model that is suitable for the lifecycle of knowledge documents and stresses the phases where lifecycle information is relevant. The arrows show where lifecycle information is being generated and where it can be utilized. The dotted arrows represent relation information while the other ones mean context information. The lifecycle in our schema has no defined beginning or end with exception of the creation phase. Here, a **creation** of a new document "from scratch" is meant, where no relations to other documents exist, thus it is the starting point to get into the cycle. Nevertheless, this is quite seldom the case, since most new documents are created by re-using, aggregating, re-purposing or in any other manner modifying existing documents.

Therefore the most common starting point is the **Access / Retrieval** phase, where the documents being re-used are retrieved. This phase holds for a multitude of context information like occurrences of the document in search results, downloads / copies of the document or ratings if available. When a document is copied a relation between the two instances of this document emerges. In turn, this phase is suited well to use the gathered information to support the user in retrieving desired documents.

The actual creation as it takes place in most of the cases is represented by the **Edit / Re-Use / Aggregate / Re-Purpose** phase. While this phase also holds for the emergence of context information, like applications used or persons who worked on a document, this is the main phase for the generation of relation information. Through the different authoring processes different kinds of relations emerge.

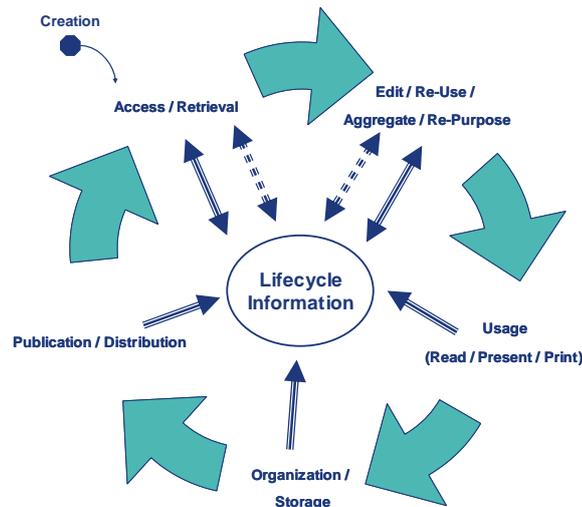


Figure 1: Schematic Knowledge Document Lifecycle

The **Usage** phase depicts the actual use of the document, where the document serves the desired purpose. Context information like e.g. the time a presentation took can be gotten here.

During the **Organization / Storage** phase, the document is stored. Context information can e.g. be gotten from the directory tree. The harvesting of e-mail context would be done in this phase, too.

During the **Publication / Distribution** phase, the document is made available to other persons. Naturally, the labelling of the document would take place during this phase, providing context information. We did not consider a destruction or archiving phase, because we think that this is not relevant for our purpose. The phases shown in Figure 1 may be mostly traversed in the given order; however, in reality interconnections between the different phases do exist but are left out for the sake of clarity.

3 Capturing and Utilization of Lifecycle Information

Our intention is to collect the information at the moment when it emerges. Therefore it can only be captured within the systems and applications that are part of a knowledge document's lifecycle. For the **capturing** there are normally two possibilities: If we have full control over the application, because the application is developed by us or it is an open source application, the capturing can take place from within the source code of the application. Otherwise we have to rely on event-handlers, which are called when different user actions are taken within the application, and accordantly capture the collectable information. Examples for information that can be captured:

- Number of downloads from a document storage
- Explicit feedback information from other users
- Version- and instance relations
- Father-/child relations
- Part-of relations
- Structural relations (permutation, successor, predecessor, etc.)

For the **utilization** of lifecycle information, similar things hold true. We want to present the right parts of processed lifecycle information to the user in moments when he or she can benefit from it the most. Therefore the information has to be integrated into the application the user currently works with. This will require in most cases plug-ins, too. In [Lehmann, 07] we presented a system where we implemented the utilization of lifecycle information for the ranking of search results in an authoring-by-aggregation environment for Learning Resources as well as the recommendation of related resources. For knowledge documents, we plan - among others - the following ways of utilization:

- Ranking (by number of downloads, re-usability, ...) - *retrieval phase*
- Recommendations of related documents - *retrieval/authoring phase*
- Search without query concepts (via relations) - *retrieval phase*

- Visualisation of a the existing document network - *retrieval phase*
- Tracking of a document's influence on other documents - *retrieval/authoring phase*
- Provision of feedback from other users - *authoring phase*
- Notifications (changes, updates, new additions to the document network) - *authoring phase*
- Automatic generation of acknowledgements - *authoring phase*

4 Concept and Architecture

Figure 2 shows the architecture for a system implementing the aforementioned approach. It is a client / server based system. The **Lifecycle Management System (LMS)** on the server side stores and processes the collected lifecycle information. It stores the collected information with reference to the accordant documents stored in the backend. It provides a web service API that the **Capture- and Utilization Plugins (CP & UP)** within the applications make use of to get the processed information from the LMS or to send the collected information back respectively. The **Backend** is hold flexible. Different kinds of document stores can be connected to the LMS that distributes the IDs and holds the lifecycle information for every document.

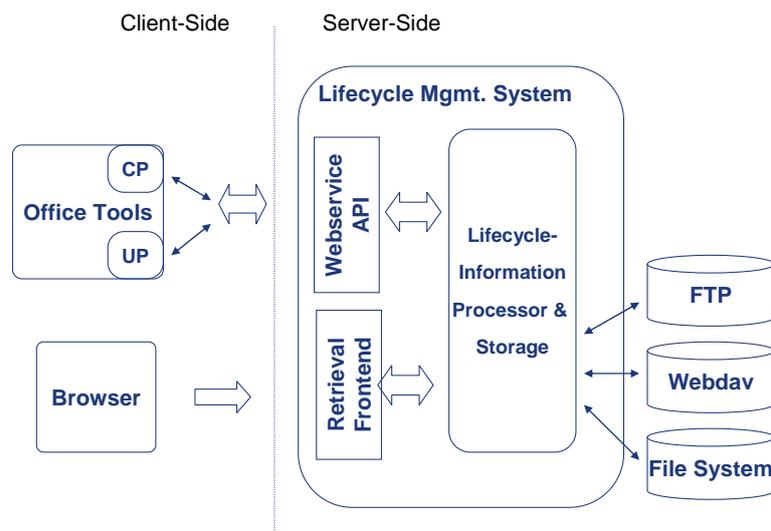


Figure 2: Conceptual Architecture

We plan to integrate a web front-end into the LMS that enables the users to retrieve documents from the connected backend solutions. In such a case we have the full control over the retrieval service and thus can fully support the user with additional information. The processing of the collected lifecycle information includes, among others, the distribution of new information to involved documents, checking of relations to avoid dead links and the enrichment of relations via algebraic rules.

In [Lehmann, 07b] we propose an architecture for the management of lifecycle information for Learning Resources. In contrast to that, this one copes with the special demands, knowledge documents make.

Besides the here described conceptual architecture there are several existing applications and frameworks where lifecycle information would be a nice feature. Examples are so called Semantic Desktops, which are basically desktop search and retrieval application where users can describe and structure their content and documents, like described in [Sauer mann, 05]. Personal Learning Environments [Wilson, 06] might benefit from the capturing and utilization of lifecycle information, too. A P2P version of the approach, where users can share knowledge documents and the collected lifecycle information and thus create huge networks of interconnected documents, is planned, too.

5 Related Work

There are several approaches, where the utilization of lifecycle information is partly conducted. This applies especially for context information. In [Najjar, 06] the Attention Metadata system is described, where context information and user behaviour is tracked to gain information about learning objects. In [Ochoa, 06] an approach is proposed, how this information can be used to rank and recommend search results for learning objects. These approaches mainly deal with learning resources and do not take the relations that connect different instances and variants of knowledge documents or learning resources into account. In [Chirita, 05] context metadata is used to get search results for desktop search. Here the context information is gotten from the e-Mail context, file system information and the browser cache. Again, relation information is neglected. Known commercial platforms like eBay or Amazon utilize special kinds of context information to rank and recommend items. This has been a focus of research several years ago, e.g. in [Good, 99].

Systems and approaches that utilize the relations between knowledge documents are rather sparse. In [Mueller, 06] a system for a management of change for collaborative authoring of structured and unstructured documents is proposed. This is somehow build upon relations but mostly relies on semantic, intra-document relations that are inevitable for a consistent management of change. Nevertheless this might be an interesting addition to the lifecycle information gathered by us.

6 Conclusions and Future Work

This article shows that lifecycle information can be of great importance for the retrieval and management of knowledge documents. A schematic lifecycle model for this kind of documents was proposed and the described architecture has shown how lifecycle information of knowledge documents can be managed. The next step is the refinement of the proposed schema along with the analysis of the different types of context and relation information, before the system can be implemented. In further works, a schema for the storage and management of the gathered information has to be found. We think that lifecycle information can improve the management and retrieval of living documents significantly.

References

- [ADL, 07] Advanced Distributed Learning, SCORM - Shareable Content Object Reference Model, <http://www.adlnet.gov/scorm/index.aspx>
- [Berndt, 05] Berndt, O.; Biffar, J. & Zoeller, B. (2005), DMS – vom elektronischen Archiv zum Enterprise-Content-Management aus VOI: Code of Practice - Dokumenten-Management, VOI, chapter 1.
- [Brooks, 06] Brooks, C. & McCalla, G. (2006), Towards Flexible Learning Object Metadata, *in* 'In Proceedings of Int. J. Cont. Engineering Education and Lifelong Learning, Vol16, Nos 1/2'.
- [Chirita, 05] Chirita, P.; Gavriiloaie, R.; Ghita, S.; Nejdil, W. & Paiu, R. (2005), Activity Based Metadata for Semantic Desktop Search, in 'Proceedings of the 2nd European Semantic Web Conference'.
- [Collis, 04] Collis, B. & Strijker, A. (2004), 'Technology and Human Issues in Reusing Learning Objects', *Journal of Interactive Media in Education* 4.
- [Davenport, 99] Davenport, T.H. & Prusak, L. (1999), *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press.
- [Ginsburg, 99] Ginsburg, M. (1999), 'An Agent Framework for Intranet Document Management', *Autonomous Agents and Multi-Agent Systems* 2, 271 - 286.
- [Good, 99] Good, N., Schafer, J.B., Konstan, J.A., Borchers, A., Sarwar, B., Herlocker, J. & Riedl, J. (1999), Combining Collaborative Filtering with Personal Agents for Better Recommendations, *in Proceedings of the 'AAAI/IAAI*.
- [Lehmann, 07] Lehmann, L.; Rensing, C. & Steinmetz, R. (2007), Capturing, Management and Utilization of Lifecycle Information for Learning Resources, *accepted at 'EC-TEL 2007'*.
- [Lehmann, 07b] Lehmann, L.; Hildebrandt, T.; Rensing, C. & Steinmetz, R. (2007), Lifecycle Information Management and Utilization in an Authoring by Aggregation Environment, *in 'Proceedings of the Edmedia 2007'*.
- [McCalla, 04] McCalla, G. (2004), 'The Ecological Approach to the Design of E-Learning Environments: Purpose-based Capture and Use of Information About Learners', *Journal of Interactive Media in Education* 7.
- [Mueller, 06] Mueller, N. (2006), An Ontology-Driven Management of Change, in 'In Wissens- und Erfahrungsmanagement, LWA (Lernen, Wissensentdeckung, Adaptivitaet) conference proceedings'.
- [Najjar, 06] Najjar, J.; Wolpers, M. & Duval, E. (2006), Attention Metadata: Collection and Management, in 'Proceedings of the WWW2006 workshop on Logging Traces of Web Activity: The Mechanics of Data Collection, Edinburgh, Scotland'.
- [Ochoa, 06] Ochoa, X. & Duval, E. (2006), Use of Contextualized Attention Metadata for Ranking and Recommending Learning Objects, *in 'Proceedings of the CAMA 2006'*.
- [Sauermann, 05] Sauermann, L.; Bernardi, A. & Dengel, A. (2005), Overview and Outlook on the Semantic Desktop, *in 'Proceedings of the 1st Workshop on The Semantic Desktop at the ISWC 2005 Conference'*.
- [Wilson, 06] Wilson, S.; Liber, O.; Johnson, M.; Beauvoir, P.; Sharples, P. & Milligan, C. (2006), Personal Learning Environments: Challenging the dominant design of educational systems, *in 'Proceedings of EC-TEL'*.

Governing Knowledge Risks – Design and Results of an Empirical Study

Florian Bayer

(Martin Luther University of Halle-Wittenberg, Germany
florian.bayer@wiwi.uni-halle.de)

Ronald Maier

(University of Innsbruck, Austria
ronald.maier@uibk.ac.at)

Abstract: Knowledge assets are recognized as primary sources of competitive advantage. Knowledge management typically aims at increasing visibility of knowledge, codifying it and enhancing knowledge sharing in order to improve (re-)use of knowledge assets. However, the often open-minded and decentral approach to knowledge management also bears risks that knowledge assets are negatively affected and competitive advantages potentially diluted. These risks are not considered systematically so far. This paper defines the concept of knowledge risks, discusses knowledge risk management issues and presents a model of hypotheses in relation to their governance. This model builds the basis for the design of an empirical study. Finally, the results of this study on the current state-of-practice of governing knowledge risks are reported based on a stratified random sample comprising 129 German companies.

Keywords: empirical study, knowledge asset, knowledge risk, risk, risk management

Categories: H.0

1 Introduction

Knowledge intensity of activities, products and services has increased substantially over the last decades and has led to a more systematic management of knowledge as resource or production factor. Consequently, knowledge management (KM) approaches focus on efficient deployment of these assets and support their creation, transfer and (re-) application in order to improve productivity of knowledge work. However, corresponding KM activities and instruments do not consider how knowledge can be secured [Desouza, Vanapalli 2005, 76]. Moreover, some authors emphasize a dark side of KM that exemplarily includes misuse of increased visibility or distribution of knowledge [Alter 2006, 3].

Risk management (RM) has long been recognized as integral part of management, but companies have embraced this topic only recently as consequence of e.g., dynamic environments, networked IT infrastructures, prominent bankruptcies and subsequent regulations like Sarbanes-Oxley-Act, EU's 8th Directive, Basel II, HIPAA or KonTraG. Despite the acknowledged importance of knowledge assets, predominantly market, credit and operational risks are targeted, whereas risks that affect knowledge assets (knowledge risks) are considered marginally at most. Consequently, the relevance of governing knowledge risks is motivated by two trends. Firstly, the

importance of intangible and knowledge assets that is more and more considered in accounting and secondly the intensified risk regulation [Jennex 2006, 3].

Based on the definition of the concept of knowledge risk (section 2), the goal of this paper is to systemize the influence of governing measures (section 3) and show the supposed impacts on the concepts knowledge loss, diffusion, transfer and quality (section 4). Finally, the paper reports on findings of an empirical study on governing knowledge risks (section 5) before it is concluded (section 6).

2 The concept of knowledge risk

From a resource-based perspective, knowledge assets as subset of intangible assets build the basis of competitive advantages [Mentzas et al. 2003, 1]. Assets can be defined “as firm-specific resources that are indispensable to create value for firms” [Nonaka et al. 2000, 20] whereas intangible assets lack physical embodiment [Barney 1991, 110f]. Knowledge assets can be operationalized by considering the different media on which knowledge can reside [Nonaka et al. 2000, 20ff; Cummings, Teng 2003, 43f]. The primary “media” are employees who provide skills and experiences [Mentzas et al. 2003, 27]. Knowledge can also be embedded in organizational routines, procedures and structures or it can be incorporated in objects comprising different forms of intellectual property, e.g., patents, as well as products or documents [Lev 2005, 200]. Organizational capabilities bundle knowledge assets in order to contribute directly or indirectly to the creation of value [Grant 2001, 118].

The term risk is discussed heterogeneously in management and economics and focuses either on its causes or its impacts. Risk is related to any business operation and includes possible losses that result from uncertainty concerning the realization of undesired events [Kaplan, Garrick 1981, 12]. Risk can also be defined as a condition in which a deviation from a desired outcome can occur [Gallati 2003, 8]. Deviations can refer to targets, plans or results of a decision. A positive deviation is considered as opportunity and a negative as threat or risk in a narrow sense. Moreover, risks can be characterized by a probability that an undesired event occurs and an extent of loss that goes along with the occurrence of this event [Kaplan, Garrick 1981, 13].

Strategic risks are characterized by long-term impact, more interacting variables, and higher degree of abstraction and are thus harder to identify, assess and manage. Risks on an operational level can be defined as risks “of loss resulting from inadequate or failed internal processes, people and systems or from external events” [Basel 2005, 140]. KM initiatives certainly should be regarded as strategic interventions. However, it is difficult to precisely define and manage strategic knowledge assets due to their intangible nature. Consequently, challenges of risk assessments are even higher compared to the already substantial ones involved in strategic risk management for tangible or financial assets. Thus, in the following the focus is on operational risks being well aware of the fact that an organization’s strategy ultimately should include aspects of knowledge risks. Consequently, knowledge risk is defined as operational risk caused by a loss, unintended diffusion, insufficient transfer and limited quality of knowledge assets and results in a lack or non-exclusivity of these assets (see also Probst and Knaese [1998, 27] and Lindstaedt et al. [2004, 2]).

Knowledge assets as the medium knowledge resides on are the targets that are affected by knowledge risks. This means that knowledge risks can concern knowledge

bound to persons, incorporated in objects or the organization. This definition stresses causes – loss, unintended diffusion, insufficient transfer and limited quality – and effects – lack or non-exclusivity – of knowledge risks. A lack negatively affects designing, planning, monitoring, continuously improving and, in the perspective of operational risks, primarily executing business processes. Considering the requirements of the resource-based view, exclusivity of resources is a necessary condition for competitive advantages [Jordan, Lowe 2004, 243]. In the following, some examples for causes of knowledge risks are briefly discussed.

Loss of knowledge assets is unrecoverable and leads to a lack that negatively affects the execution of business processes. It can be caused e.g., by fluctuation of employees, non-documentation or deletion of documented knowledge [Matusik, Hill 1998, 687; Mohamed et al. 2006, 3]. *Unintended diffusion* means access to sensitive knowledge by unauthorized persons. Contrary to knowledge loss, knowledge is still available, but not exclusively to the organization. It can be caused by insufficient definition or provision of access rights, reverse engineering or attacks on IT systems. *Insufficient transfer* in this case primarily refers to processes in which organizations attempt to get access to external knowledge that they can not create internally for reasons of time or cost [Baughn et al. 1997, 103]. Insufficient transfer results in a lack and can be caused by too rigid rules concerning security, but also vague rules that leave employees hesitant about freely sharing knowledge because they are not aware what is expected from them and what would be considered an act against the interests of the organization. *Limited quality* of knowledge assets can be assessed according to content, i.e. e.g., correctness or timeliness of knowledge, the community in which knowledge is created and used, development and deployment processes that provide the knowledge as well as the IT infrastructures used to provide access to documented knowledge or meta-knowledge about knowledge sources [Eppler 2003, 68]. Moreover, dependencies from knowledge assets limit quality and can result in a lack. Examples are limited correctness, low applicability of knowledge or restricted accessibility.

These four causes are not isolated from each other, but can also interact. For example, fluctuation of employees on the one hand leads to knowledge loss with respect to processes, routines and practices in which the employees participated. On the other hand, fluctuation bears risks that knowledge diffuses, its exclusivity is lost and it is reapplied at a competing organization [Matusik, Hill 1998, 687].

3 Governance of knowledge risks

RM has its roots in the insurance industry and typically comprises identification, assessment, governance and evaluation as basic steps [Archbold 2005, 32]. The knowledge RM process consists of the same steps, but additionally requires identification of knowledge assets (see Figure 1).

Identification of knowledge assets affected by knowledge risks is required since visibility of these assets is a necessary precondition for the identification of knowledge risks. Consequently, established KM initiatives or approaches concerning intellectual capital management are favorable. *Identification of knowledge risks* can use different sources such as review of contracts, policies and their compliance, penetration tests for IT systems or analysis of dependencies on knowledge assets. Identified

knowledge risks have to be assessed concerning probability of occurrence and severity of resulting losses. This assessment has to be based on the value of knowledge assets and also interactions between them. However, valuation of knowledge assets is still in its infancy and thus the assessment of knowledge risks is still problematic.

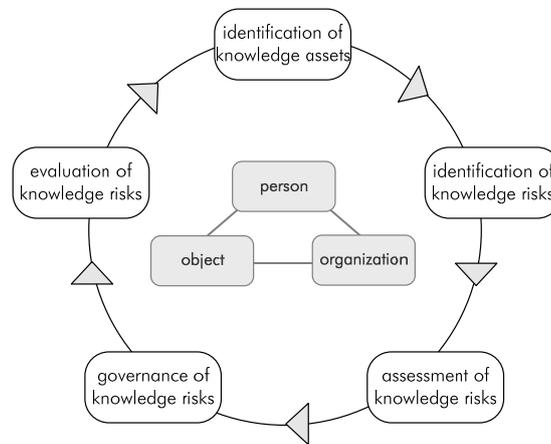


Figure 1: knowledge risk management process

Governance means the set of processes and policies affecting the way the handling of knowledge is directed, administered or controlled. *Evaluation of knowledge risks* concludes the knowledge risk management process, an ongoing process since risks, probabilities, severity and efficiency of governance measures change over time.

Governance is operationalized here by 13 variables and subdivided in organizational, technical and legal measures. Organizational measures comprise a classification of knowledge assets concerning their sensitivity, reduction of dependencies on knowledge assets, deployment of gatekeeper and knowledge transfer policies to control external knowledge flows as well as physical access control and IT security policies to avoid access of unauthorized persons or security breaches that are caused by employees. Moreover, compliance of IT security policies was integrated since policies are only useful, if employees act on them. Technical measures include control of access rights to electronically documented knowledge as well as their provisioning. Legal measures are operationalized by non-disclosure and non-compete agreements that aim at avoiding knowledge diffusion based on misbehavior of employees, alliance agreements that limit the use of knowledge in strategic alliances and intellectual property rights such as patents or copyrights. These governance measures are also summarized in Table 1 together with some empirical results.

4 Hypotheses

The concept of governance of knowledge risks as central element of the knowledge RM process is at the core of the empirical study which investigates influences of governance measures on the four concepts knowledge loss, diffusion, transfer and

quality that also reflect the causes for knowledge risks discussed in section 2. The assumed relations between the concepts are based on an extensive literature survey, depicted in figure 2 and subsequently discussed in detail.

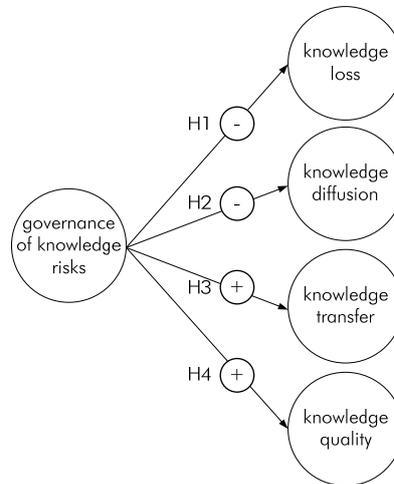


Figure 2: Hypotheses concerning governing knowledge risks

Hypothesis H1. Governance of knowledge risks negatively affects knowledge loss. The concept of knowledge loss can be measured by variables such as non-documentation of knowledge in day-to-day business or in projects or degree of loss caused by job succession or substitution [Desouza, Vanapalli 2005, 84]. It is expected that governance measures negatively affect probability of knowledge loss by rules concerning e.g., email and document retention planning, documentation and reduction of dependencies as well as measures that reduce the probability of fluctuation, e.g., comparable wages, leadership, incentives.

Hypothesis H2. Governance of knowledge risks negatively affects knowledge diffusion. The concept of knowledge diffusion can be measured by variables such as access by unauthorized persons, unfavorable employee fluctuation or reverse engineering activities by competitors [Zander, Kogut 1995, 88f; Norman 2004, 612; Desouza, Vanapalli 2005, 81f]. It is assumed that probability and exposure of knowledge diffusion is reduced by deploying governance measures such as access control, non-disclosure agreements or alliance agreements.

Hypothesis H3. Governance of knowledge risks positively affects knowledge transfer. The concept of knowledge transfer is focused on its success and measured by variables such as contribution of the transferred knowledge to other projects or tasks, extension of the knowledge base or reduction of the dependency or reliance on partner knowledge [Simonin 1999, 621]. It is assumed that companies without clear governance rules are rather restrictive concerning knowledge transfer. Employees might hold back knowledge, if they are in doubt whether it may or must not be transferred. Consequently, clear knowledge transfer policies are assumed to positively influence this concept since they support clear-cut decisions about transferability.

Hypothesis H4. Governance of knowledge risks positively affects knowledge quality. In order to measure knowledge quality, exemplary variables such as accessibility of IT infrastructures, applicability or correctness of documented knowledge are included [Eppler 2003, 74]. It is assumed that governance of knowledge risks positively affects knowledge quality, since companies are sensitized for the importance of knowledge assets and aim at reducing shortcomings concerning the various dimensions of knowledge quality by deploying appropriate measures.

5 Design of the empirical study and preliminary results

These hypotheses are subject to a broad explorative empirical study. For each of the concepts, variables were selected from the literature, operationalized in 5-13 statements and coded according to a 7-point Likert scale ranging from “I strongly agree” to “I disagree”. Additional questions concerned the estimated importance of knowledge, already established RM and KM initiatives as well as organization and background of the interviewee. Seven telephone interviews were carried out as pretests in order to check the questions’ understandability, ambiguity, contribution to measurement as well as length and duration of the telephone interviews.

Of a population of 3.2 million German enterprises, the study comprised 129 organizations that were selected on the basis of a stratified random sample. The stratification of the sample considers the two criteria industry and number of employees. All industries were covered because there has been no evidence of differences between industries in terms of importance of knowledge (risks) prior to this empirical study [Zack 2003, 67ff]. The study covers twelve categories of industries and primarily focuses on organizations with 50 or more employees since it is assumed that relevance of knowledge risks increases with the number of employees. The stratified sample comprises 83 medium-sized companies (50-249 employees) and 22 large companies (>249 employees). Additionally, 12 companies with fewer than 10 employees and 12 companies with between 10 and 49 employees were included.

Appropriate contact persons had to be identified able to answer the questions in a professional way. CEOs, CSOs, CKOs or heads of PR/communication were contacted by telephone. Structured questionnaires were sent out to the identified persons that could be answered in a telephone interview, per e-mail, fax or normal mail.

Based on a sample of about 1,800 addresses and phone numbers drawn randomly from a database of an external provider, 1,216 companies were contacted by phone. 872 interview guidelines were sent out to inform the identified contact persons about the study. Thereafter, the contact persons were called again in order to inform them about the study and check their willingness to participate. 148 questionnaires were sent back or answered during a telephone interview (response rate 12.2%). Questionnaires were predominantly answered by CEOs (53%) or other executive managers with an average work experience within the company of 11.3 years. 26% of the companies state a legal obligation for RM and 53% of the companies have already implemented RM initiatives. KM initiatives exist in 40% of the participating companies.

Table 1 summarizes descriptive statistics concerning the 13 variables that have been used to operationalize the concept of governance of knowledge risks. For each variable, the table includes the median (X_{MED}) and mean (X_{MEAN}) for the entire sample measured as degree of agreement ranging from 1 (low) to 7 (high). The next four

columns contain the means according to the different sizes of the companies. The variables in table 1 are sorted by X_{MEAN} in descending order. IT security policies and IT security compliance are ranked highest with a mean of 4.81 and 4.79. Based on qualitative statements during the telephone interviews and comments on the questionnaires, this tendency could be a result of widely spread standards for IT security and requirements for certification. Moreover, especially here the trend that large companies govern knowledge risks stronger is obvious, since the mean ranges considerably from 3.42 in the case of small companies to 5.73 in the case of large companies. Concerning such an interpretation over all company sizes one has to be careful and primarily focus on the latter two categories, since the sample size in the two categories with fewer than 50 employees is comparably low. Also, non-disclosure agreements are a well established measure. This can be underlined since an open question concerning the three most effective governance measures was included and non-disclosure agreements were mentioned comparably often. Again, large companies deploy non-disclosure agreements to a comparably larger scale.

IT security policies (mean difference MD 0.93, significance: 0.0091), knowledge transfer policies (MD 0.89; significance: 0.0404) and physical access control (MD 0.85; significance: 0.0300) are deployed significantly more widespread in larger companies compared to medium-sized companies on a significance level of $\alpha \leq 0.05$.

variable	type of measure	X_{MED}	X_{MEAN}	number of employees			
				n=12	n=12	n=83	n=22
				-9	-49	-249	>249
IT security policies	organiz. ¹	5	4.81	3.42	4.58	4.80	5.73
IT security compliance	technical	5	4.79	4.08	4.83	4.78	5.18
non-disclosure agreements	legal	5	4.77	4.25	4.25	4.78	5.27
provisioning of access rights	technical	5	4.51	4.33	3.58	4.57	4.91
classification of knowledge	organiz.	5	4.50	4.83	3.75	4.55	4.55
physical access control	organiz.	5	4.46	4.33	4.50	4.29	5.14
electronic access control	technical	4	4.42	4.17	4.00	4.36	5.00
knowledge transfer policies	legal	5	4.33	4.58	4.33	4.11	5.00
gateway employees	organiz.	5	4.32	4.50	4.25	4.24	4.55
reduction of dependencies	organiz.	4	3.90	4.25	3.58	3.87	4.00
intellectual property rights	legal	4	3.65	3.33	3.83	3.51	4.27
non-compete agreements	legal	3	3.57	3.92	3.42	3.49	3.73
alliance agreements	legal	3	3.36	4.08	3.17	3.24	3.55

Table 1: descriptive statistics concerning governance measures (n=129)

Concerning measures with a lower degree of implementation in organizations, alliance agreements are not in place or not relevant for most of the companies participating in this study. This can be explained by the fact that a large part of companies are not involved in strategic alliances and consequently do not require such an agreement. Here, 53 companies (41.1%) disagreed or only weakly agreed. Also, intellectual property rights and non-compete agreements are not as widespread as other measures. This could be explained by the fact that such agreements are particularly of interest for executive managers and not for the entire workforce. Table 1 also shows a tendency that technical and organizational measures predominate, whereas legal measures are considered less important.

¹ organiz. = organizational

6 Conclusion

This paper has taken a different perspective on KM, namely a risk perspective. After defining the knowledge risk management process taking on board experiences in management of operational risks, governing knowledge risks was introduced together with relations to knowledge loss, diffusion, transfer and quality. Empirical findings indicate that large companies govern knowledge risks comparably stronger and are less affected by knowledge risks or more successful in governing them. However, some governance measures negatively affect processes that are intended from a KM perspective (e.g., knowledge transfer, transparency). Consequently, future analysis using multivariate statistics will focus on the influence of specific governance measures and also shed light on possible negative impact.

Moreover, the analysis will focus on differences between industries and the influence of implemented risk and/or knowledge management initiatives. First results indicate that companies with implemented initiatives in both areas are characterized by a comparably higher degree of governance and a lower degree of risk in the corresponding risk causes with the exception of the concept of knowledge diffusion. Here, companies who only have RM initiatives have a lower estimation of knowledge diffusion risks.

These results should feed into supporting identification, assessment and governing knowledge risks. Specifically, the data will be used to develop recommendations concerning the use of governance measures, criteria for classifying knowledge according to the need of protection as well as recommendations for implementing an effective knowledge risk management.

References

- [Alter 2006] Alter, S.: "Goals and Tactics on the Dark Side of Knowledge Management" Proc. Proceedings of the 39th Hawaii International Conference on System Sciences - 2006, (2006), 1-10.
- [Archbold 2005] Archbold, C. A.: "Managing the bottom line: risk management in policing"; *Policing: An International Journal of Police Strategies & Management* 28, 1 (2005), 30-48.
- [Barney 1991] Barney, J. B.: "Firm Resources and Sustained Competitive Advantage"; *Journal of Management* 17, 1 (1991), 99-120.
- [Basel 2005] Basel. "International Convergence of Capital Measurement and Capital Standards". Basel.
- [Baughn et al. 1997] Baughn, C., Denekamp, J., Stevens, J., Osborn, R.: "Protecting intellectual capital in international alliances"; *Journal of World Business* 32, 2 (1997), 103-117.
- [Cummings, Teng 2003] Cummings, J. L., Teng, B. S.: "Transferring R&D knowledge - the key factors affecting knowledge transfer success"; *Journal of Engineering and Technology Management* 20, 1 (2003), 39-68.
- [Desouza, Vanapalli 2005] Desouza, K. C., Vanapalli, G. K.: "Securing Knowledge in Organizations" in: K. C. Desouza: *New frontiers of knowledge management*, Basingstoke, (2005), 76-98.

- [Eppler 2003] Eppler, M. J.: "Managing Information Quality. Increasing the Value of Information in Knowledge-intensive Products and Processes"; Springer, Berlin (2003).
- [Gallati 2003] Gallati, R. R.: "Risk management and capital adequacy"; McGraw-Hill, New York (2003).
- [Grant 2001] Grant, R. M.: "Contemporary strategy analysis"; Blackwell Publ., Malden, Mass. (2001).
- [Jennex 2006] Jennex, M. E.: "Security and Knowledge Management Success" Proc. The Second Secure Knowledge Management Workshop (SKM) 2006, Brooklyn, (2006).
- [Jordan, Lowe 2004] Jordan, J., Lowe, J.: "Protecting Strategic Knowledge: Insights from Collaborative Agreements in the Aerospace Sector"; *Technology Analysis and Strategic Management* 16, 2 (2004), 241-259.
- [Kaplan, Garrick 1981] Kaplan, S., Garrick, B. J.: "On the quantitative definition of risk"; *Risk Analysis* 1, 1 (1981), 11-27.
- [Lev 2005] Lev, B.: "Intangible Assets: Concepts and Measurement" in: K. Kempf-Leonard: *Encyclopedia of Social Measurement Vol. 2*, Amsterdam, (2005), 299-305.
- [Lindstaedt et al. 2004] Lindstaedt, S., Koller, S., Krämer, T.: "Eine Wissensinfrastruktur für Projektrisikomanagement - Identifikation und Management von Wissensrisiken" Proc. Tagungsband zur KnowTech 2004, 6. Konferenz zum Einsatz von Knowledge Management in Wirtschaft und Verwaltung, 367-375, Munich, Germany, (2004).
- [Matusik, Hill 1998] Matusik, S. F., Hill, C. W.: "The utilization of contingent work, knowledge creation, and competitive advantage"; *Academy of Management Review* 23, 4 (1998), 680-697.
- [Mentzas et al. 2003] Mentzas, G., Apostolou, D., Young, R.: "Knowledge asset management: beyond the process-centred and product-centred approaches"; Springer, London (2003).
- [Mohamed et al. 2006] Mohamed, S., Mynors, D., Grantham, A., Walsh, K., Chan, P.: "Understanding One Aspect of the Knowledge Leakage Concept: People" Proc. European and Mediterranean Conference on Information Systems (EMCIS), July 6th-7th 2006, Alicante, Spain, (2006).
- [Nonaka et al. 2000] Nonaka, I., Toyama, R., Konno, N.: "SECI, ba, and leadership: a unified model of dynamic knowledge creation"; *Long Range Planning* 33, 1 (2000), 5-34.
- [Norman 2004] Norman, P. M.: "Knowledge acquisition, knowledge loss, and satisfaction in high technology alliances"; *Journal of Business Research* 57, 6 (2004), 610-619.
- [Probst, Knaese 1998] Probst, G., Knaese, B.: "Risikofaktor Wissen: wie Banken sich vor Wissensverlusten schützen"; Gabler, Wiesbaden (1998).
- [Simonin 1999] Simonin, B. L.: "Ambiguity and the Process of Knowledge Transfer in Strategic Alliances"; *Strategic Management Journal* 20, 7 (1999), 595-623.
- [Zack 2003] Zack, M. H.: "Rethinking the Knowledge-Based Organization"; *MIT Sloan Management Review* 44, 4 (2003), 67-71.
- [Zander, Kogut 1995] Zander, U., Kogut, B.: "Knowledge and the Speed of the Transfer and Imitation of Organizational Capabilities: An Empirical Test"; *Organization Science* 5, 1 (1995), 76-92.

Knowledge Fusion: From Management to Mobilization

Peter Keen

(University of Nebraska, USA
petergwkeen@yahoo.com)

Margaret Tan

(Nanyang Technological University, Singapore
mtan@ntu.edu.sg)

Abstract: Leading researchers in the fields of business strategy, organizational theory, institutional economics and economic development routinely emphasize the growth of the knowledge economy and by implication, there is a vital need for organizations to manage, share and leverage knowledge assets. Yet, it has been increasingly clear from previous research that the field of knowledge management (KM) has not in general met its goals of transforming information sharing and collaborations in organisations. Lucier and Torsilera [97] claim that 84 percent of KM programs fail to have any real impact. Despite these problems, KM addresses a clearly perceived need for innovation, whether expressed in terms of the knowledge economy, knowledge workers, or intellectual capital management. This paper points to some gaps in the current body of knowledge about KM, and suggests areas of investigation that build on its strength. We propose a simple framework for what we term Knowledge Fusion to provide rigor and relevance to KM

Keywords: knowledge management, knowledge mobilization, knowledge fusion

Categories: H.m, H.5.2

1 Introduction

The Knowledge Life Cycle [Birkinshaw and Sheehan, 02] states that knowledge progresses through four stages (creation, mobilization, diffusion and commoditization) and as it develops overtime, knowledge becomes accessible to more and more people – first in the organisation and finally, to the general public – therefore, companies must use different strategies to realise the maximum value of the knowledge flows. In this context, KM is therefore axiomatically a mission-driven, corporatist field. Its focus is on management processes that use information resources and its related corporate “assets” to enhance innovation and collaboration: creation, sharing and dissemination. There are, however, many valid and powerful alternatives to the axioms of KM, explicated by [Ekbja and Hara, 04; Wilson, 02], but they basically reject KM for its mission as much as its methods and intellectual base. To a large degree, “membership” in the KM field of research and practice involves accepting the corporatist mission. We choose the word “corporatist” carefully, since it captures the view of knowledge as organizational assets, the aggressive goal of innovations and the purposive intentions of generating a high return on investment (ROI). It largely views knowledge as fungible organisational assets to be organized and made accessible through the many information and communication tools and

coordinated through changes in work practices that facilitate collaboration and information-sharing. Such view generates conflict for many thinkers who do not believe that knowledge is to be valued mainly for its contribution to organizational payoff especially in the context of 'ROI'.

KM as a corporatist practice is in many ways an announcement by the information systems community that it has positioned to move beyond information organization to information deployment; that shift is signaled by the choice of "knowledge" as the target for "management." A constant tension in the KM field is the difference between information and knowledge, yet at its core, KM has been very much information-centric. It aims at connecting innovation and growth, the core goals of the enterprise, back to information-based capabilities, and in doing so, it aims to raise its own centrality as a strategic force in and of itself rather than as a support base for change management, process innovation, and business capability development.

A major limitation to progress in KM application and impact is that there is a clear difference between the fundamental dynamics of knowledge management and of knowledge *mobilization*. Knowledge management addresses the supply side of information organization, creation of environments for communication, collaboration, leveraging of intellectual capital, and providing incentives for shifts in work practices, especially those that either impede or facilitate knowledge-sharing, with "knowledge" largely being independent of the individual. Indeed, it is viewed as a corporate asset. Leonard's [89] assertion is representative here: "Just as organizations are financial institutions, so they are knowledge institutions." Knowledge mobilization, by contrast, reflects the demand side that is dominated by knowledge being part of individual identity and hence personal choice of whether, where, why and with whom to share knowledge and expertise [Qureshi and Keen, 05]. Knowledge mobilization views information and knowledge in terms of situational needs – "what do *I* need to know *now*?" – while knowledge management tends to focus more on "what knowledge can we provide to our employees and what mechanisms can we put in place for them to make most effective use of it?" The push-pull tension between management and mobilization is captured in a comment by a manager that "The organization does not understand how knowledge is shared here and I tend to ignore the knowledge management initiatives wherever I can" [von Krogh et al, 94].

We propose a framework for what we term *Knowledge Fusion* to mobilize knowledge based on the argument that captures what KM is as a field, rather than what many of its critics feel it should *not* be as a domain of intellectual study and social action. KM should not get stuck in definitional debates, but it does need some shaping framework that encourages intellectual and pragmatic diversity and a balance between the thought leadership priorities of the pragmatists, often consultancy firms, and the research and scholarship excellence of the intellectual disciplines, mostly but not entirely in the academic communities. Our proposal is to "partition" the wider field of "knowledge" into four areas: knowledge *management*, the goal; knowledge *mobilization*, the enabler; knowledge *embodiment*, the study of what it means to "know"; and knowledge *regimes*, the organizational, political and sociological factors that shape how knowledge is focused, authenticated, legitimized and validated in the organizational and professional context. Each of these is a distinctive arena, in terms of its fields of research and scholarship, axiomatic base, mode of investigation and

professional communities. We see three links: (1) knowledge management and knowledge mobilization; (2) knowledge regimes and knowledge management; and (3) knowledge embodiment and knowledge mobilization. The Knowledge Fusion framework rests on the logic that contributions from these communities will determine how they link their specific body of theory and practice to the KM mission.

2 The Grounded Theory Development of Knowledge Fusion

Our formulation of the Knowledge Fusion framework is based on the analysis of the KM field and a grounded theory of investigation that began from a wide-ranging scan of an almost unsurveyable field. Scholar Google lists 220,000 references to “knowledge management”; the fragmentation and breadth of the field is indicated by the fact that only a tiny fraction of these are cited in even twenty of the other close to a quarter of a million publications. A Google search on the term generates 57 million results, and to put that in context, “business process reengineering” produces 1.3 million and “electronic commerce” 28 million (February 2006).

Such proliferation eliminates any practical possibility of a grand theory of KM or a unified definition of “knowledge.” Our approach to generating a grounded theory base for KM extension is to identify salient themes in KM, such as communities of practice, knowledge-sharing, knowledge creation, tacit knowledge, and intellectual property, and then to test how well conceptually and in practice how they hold up. We saw a need for new “codings” on the processes we identified. For example, it became clear from our analysis that the widely-used distinction between explicit and tacit knowledge [Nonaka and Takeuchi, 95] is not robust and has generated tautologies and challengeable conclusions, such as the claim that tacit knowledge is knowledge that cannot be made explicit and structured, followed by the statement that a goal for knowledge management is to make tacit knowledge structured and explicit [Gourlay, 00; Haldin-Herrgard, 00]. [Gray, 01] provides a succinct summary of the tacit-explicit knowledge distinction and states that “most organizations want to transfer tacit knowledge to explicit knowledge.”

It is a substantive matter for the effectiveness of KM whether or not this transfer is possible and even conceptually meaningful. We propose, via [Wilson, 02], a simple extension of the distinction to include implicit knowledge as the bridge between tacit and explicit. Implicit knowledge is what we take for granted, rarely think about and are surprised to find that others do not share; many faux pas that we make when we travel abroad reflect the fact that a national culture has many areas of implicit knowledge concerning etiquette and social norms. The red-faced blunderer asks “why didn’t you tell me about that?” “You didn’t ask and it’s obvious anyway.” We suggest that tacit knowledge be accepted as inherently tacit and that, using our coding distinctions, knowledge management should structure explicit knowledge, which is information-centric, explicate implicit knowledge through dialog and leverage tacit knowledge through respectful collaboration.

Our grounded theory approach is more than taxonomic in its goals and less than ontological. It is a search for a parsimonious addition to the distinctions in the KM field which will help resolve conceptual contradictions and reported problems of application, such as the tacit-explicit contrast. This approach helps us to avoid getting caught in the definitional debate. Clearly, new distinctions are needed for KM to

achieve its impacts. We did not “invent” knowledge *mobilization* nor redefine “knowledge management” but instead added it as a new coding and then looked at where and how it contributes to the KM mission. This in turn points to the value of a new distinction in knowledge mobilization of three levels of personal knowledge identity in how individuals assess their own knowledge and how and when to share it: accountable, discretionary and autonomous knowledge [Qureshi and Keen, 05].

The final stage in the development of the Knowledge Fusion framework is to narrow down the broad range of KM topics, distinctions and concerns into a parsimonious set of “partitions” that link to and from traditional KM research and practice. We propose that the Knowledge Fusion framework has four main partitions.

2.1 Knowledge Management

Perceived as the organizational mission for continuing the evolution of information management to become a core factor in business innovation; the supply and dissemination of knowledge-relevant information, communication and process capabilities, and the development of change management initiatives in order to develop new knowledge-building and knowledge-sharing practices. Given the mission, issues of technology options and methods are highly germane to this partition, even though they are a distraction - for commentators who oppose the main KM axioms. These critics stress that technology in and of itself is not relevant to knowledge. But it is highly relevant to KM and indeed, there are emerging developments in technology that are promising enablers of the new knowledge work, especially in the library sciences, where the exploration of and the expertise in archiving or designing Semantic Web - are adding the missing dimension to the mainstream focus on database management systems, data repositories and Web portals.

2.2 Knowledge Mobilization

Conceptualized as the dynamics of the processes by which individuals make their personal choices about information-seeking, knowledge creation and knowledge-sharing. This demand side of knowledge must be synchronized with the supply side for effective joint benefit. The discretionary and personal nature of knowledge activation and identity leads to many gaps in practice between corporate supply and individual demand or use of knowledge, between push and pull. In fact, more elements of personal knowledge creation and sharing lie outside corporate ownership and control. For instance, blogs are becoming a significant force in professionals' knowledge-sharing that their impacts on corporate reputation can be significant. In recent times a number of companies are harnessing blogs to create two-way communication links with their constituencies, in effect mobilizing - company and stakeholder knowledge-sharing. Blogging illustrates the potential value of the partitioning approach to Knowledge Fusion. It is as yet little studied in the context of corporate KM, though there is a growing body of work on the role of blogs in e-knowledge and distributed knowledge creation in professional communities. Given the rapid growth in blogs and its increasing corporate focus, this is a topic that merits study and organizations should make them as part of their knowledge mobilization if not part of their knowledge management.

2.3 Knowledge Embodiment

Conceptualized as the deep processes of “knowing” in the widest sense of the term. To debate the pragmatic and axiomatic KM concept of knowledge as an organizational asset is obviously partial at best, as many commentators claim that it is largely invalid and little more than a re-labeling of information. Many KM researchers are thus exploring reference disciplines, most obvious philosophy and epistemology, which address such topics as the social and political nature of knowing, speech act theory and the linguistic nature of knowledge, and the nature of tacit knowledge. Their goal is to enrich but not to attack KM. Connecting such lines of investigation to the pragmatics of KM concepts and applications is best handled through viewing knowledge embodiment as a partition in and of itself, but one whose findings and frameworks can be brought into focus through being linked to the knowledge mobilization partition. The logic is that any effort to generate a consensual concept of knowledge will fail and that much of the research in this partition will have less impact on KM than it merits because it is positioned as a new approach to KM *as a whole*. We suggest that its power will come from rigorous scholarship and research made relevant by showing how and where it helps in increasing knowledge mobilization.

2.4 Knowledge Regimes

The term refers to the contextual rules, controls and processes that directly shape and constrain KM. These include political, cultural and sociological factors. Our identification of knowledge regimes as a partition of Knowledge Fusion was prompted by work in political science [Sowell, 96], philosophy [Foucault, 80] and the wide literature that links knowledge management to questions about capitalism and post-capitalism. The legitimacy, verification, use and control of information are an integral part of what Foucault calls “Regimes of Truth”. Knowledge regimes are the socio-political forces that strongly affect the specific legitimacy, meaning and effective rights of ownership of “knowledge.” These include organizational design, information systems, professional associations, and “culture.” Knowledge regimes vary widely between countries, with history, censorship and social norms often creating bounds on knowledge embodiment, knowledge management and knowledge mobilization.

Our proposal of knowledge regimes as a partition is intended to resolve the rigor-relevance tension in Knowledge Fusion. The mainstream of KM is driven by relevance to business and organizational innovation and collaboration. It maintains that focus sometimes at the expense of rigor; the extreme instance of this are vendor claims that, for instance, document-management software is “knowledge” or the casual comment in a leading book on KM that begins “Because of the human element in knowledge ...” [Davenport et al, 97]. Surely, any scholar in fields that address knowledge embodiment and knowledge mobilization would scream in reply that humans are not an element in knowledge but *are* knowledge. The remark makes more sense in its context of the conceptualization of knowledge as a corporate asset, much of which is embodied in information resources, not people. That said, this is certainly not a rigorous statement and it is typical of ones that critics zero in on very quickly.

Conversely, discussions of knowledge regimes are highly abstract and formalistic. They adopt different axioms of “knowledge” than does KM. For example, Day’s history of KM highlights “the European documentalist, critical modernist and Italian Autonomous Marxist influenced Post-Fordist traditions” [Day, 01]. Fuller’s blandly titled *Knowledge Management Foundations* is anything but that; it is a resonant and complex exploration of “civic republicanism” and social epistemology [Fuller, 01]. His KM manifesto includes discussion of “pseudo solutions” such as cyber platonism, and academic bullionism – the “scourge of KM.” Both Day and Fuller offer a counterview to just about every assumption, goal and application of KM. An obvious question then is why they self-classify their work as KM, virtually guaranteeing that it will have no impact on the communities within the field? The logic of the Knowledge Fusion framework is that such work is a valuable potential contribution to KM if, and only if, it can build linkages to the KM mainstream.

Our framework of partitions aims to help improve both the rigor and relevance of the KM field through appropriate use of reference disciplines. These are fields of scholarship and research – the two are not always equivalent – that offer insights, theory and findings that are relevant to but not within the immediate academic and professional purview of KM. The KM field as a whole will be enriched through such diverse and unconnected reference disciplines as epistemology, library sciences and education, sociology (Berger and Luckman’s, [66] work on the social construction of reality that is relevant to knowledge embodiment), political science (knowledge regimes), critical theory, hermeneutics, phenomenology and computer science.

Figure 1 summarizes the Knowledge Fusion framework. To be of value, it must pass tests of parsimony, that is usefulness in helping provide a coherent and comprehensive high-level mapping of a very complex and fragmented field, and originality in pointing to new lines of investigation and lessons from existing research. That judgment will be made by our readers.

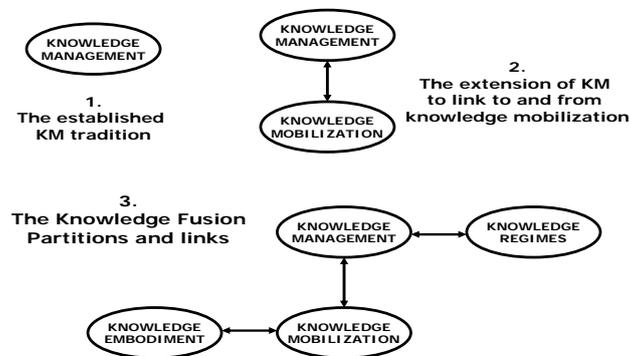


Figure 1: From Knowledge Management to Mobilization to Fusion

The recommended agenda for Knowledge Fusion is thus: (1) maintain the mainstream focus in KM on harnessing organizational resources for the purpose of innovation, knowledge-creation and collaboration, (2) sharpen the focus on linking individual demand and use of peoples’ own and others’ knowledge (knowledge

mobilization) to organizational supply and encouragement of new practices and processes (knowledge management), (3) enrich the discussion of “knowledge” and knowledge embodiment and thus of opportunities for knowledge mobilization and (4) investigate the impact of knowledge regimes on knowledge management assumptions and practices, including the role of dissent, and adapt those practices to meet the changing demands of other regimes, including those of customers, competitors, pressure groups, and social movements.

3 Conclusion

At the core of our framework is a single distinction that we view as fundamental to the effective development and impact of KM as both a field and an area of management practice: knowledge mobilization. Most of the criticisms address weaknesses in the “knowledge” component of KM. But KM is not about the topic of knowledge as such but the mission of management. KM is in fact part of a corporatist regime of truth, in Foucault’s sense of the phrase. It is fully valid to attack that regime on social, political or moral grounds — but criticisms made on the same terms about the validity of KM itself largely miss their mark. Also, the criticism that it uses fuzzy, inconsistent and contradictory definitions of knowledge is misleading in that it implies that elsewhere there is a body of consistent and reliable definitions of knowledge. No one has successfully generated an accepted and universal theory of knowledge after 2,500 years of continuous effort. The irresolvable debate or often conflict between science and religion is a difference of belief about what is knowledge and hence “truth.”

As we moved consciously to adopt a grounded theory approach to position our concept of knowledge mobilization and activation, we acknowledge the extent to which KM is axiomatic rather than definitional in its very varied domains of theory and practice and that in many ways the axioms drive the theory and practice. The heterogeneity of these domains is both the opportunity and the problem for KM. They are an opportunity, in that the work of Habermas, Rorty and Searle in the field of philosophy appears more and more as of direct practical relevance to our understanding of knowledge mobilization. They are a problem in that discussions of their work at the theoretical level and from the perspective of their own axioms do not connect well if at all to the mainstream of KM.

We have no intent in promoting Knowledge Fusion as a new theory or “model”. We developed it to guide our own investigation and collaboration with colleagues, as a vehicle for knowledge mobilization in the KM field. The diversity and quantity in many ways motivated our study; there is too much of it, it does not seem to be generating a cumulative tradition of study on which results build on each other, and at times KM is almost a haystack in which almost any needle can be found just by digging around. Something has to be done to frame KM, not to homogenize it, but to give it more shape. We hope that our framework offers a useful starting point.

References

- [Birkinshaw and Sheehan, 02] Birkinshaw, J and T. Sheehan, "Managing the Knowledge Life Cycle", MIT Sloan Management Review, Fall 2002, 75-83.
- [Berger and Luckman, 66] Berger, P. and T. Luckman, "The Social Construction of Reality: A Treatise on the Sociology of Knowledge", New York, Anchor Books, 1966.
- [Day, 01] Day, R., "Totality & Representation: A History of Knowledge Management through European Documentation, Critical Modernity & Post Fordism", Journal of the American Society for Information Science and Technology, 52, 9, 2001, 725-735.
- [Davenport et al, 97] Davenport, T., D. DeLong and M. Beers, "Building Successful Knowledge Management Projects", Center for Business Innovation Working Paper, 1997, 1-16.
- [Ekbia and Hara, 04] Ekbia, H. and N. Hara, "The Quality of Evidence in the Knowledge Management Literature: The Guru Version", 2004.
- [Foucault, 80] Foucault, M., "Power/Knowledge: Selective Interviews and Other Writings, 1972-1977", New York: Pantheon Books, 1980.
- [Fuller, 01] Fuller, S., "Knowledge Management Foundations", New York: Elsevier, 2001
- [Gourlay, 00] Gourlay, S., "On Some Cracks in the "Engine" of Knowledge-Creation: A Conceptual Critique of Nonaka & Takeuchi's Model, Proc of British Acad of Mgt Conf, 2000.
- [Gray, 01] Gray, P., "Knowledge Management Overview", Center for Research on Information Technology & Organizations, University of California at Berkeley, 2001
- [Haldin-Herrgard, 00] Haldin-Herrgard, T., "Difficulties in Diffusion of Tacit Knowledge in Organizations", Journal of Intellectual Capital, 1, 4, 2000, 357-365.
- [Leonard, 89] Leonard, D., "The Wellsprings of Knowledge: Building and Sustaining the Sources of Innovation", Cambridge, Mass: Harvard Business School Press, 1989.
- [Lucier and Torsilieri, 97] Lucier, C. and J. D. Torsilieri, "Why Knowledge Programs Fail: A CEO's Guide to Managing Learning", Strategy & Business, 4th Qtr, 1997, 14-28.
- [Nonaka and Takeuchi, 95] Nonaka, I. and H. Takeuchi, "The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation", NY: Oxford Uni. Press, 1995.
- [Qureshi and Keen, 05] Qureshi, S., and P. Keen, "Activating Knowledge through Electronic Collaboration: Vanquishing the Knowledge Paradox", Proc. of the 40th HICCS, Hawaii, 2005.
- [Sowell, 96] Sowell, T., "Knowledge and Decisions", New York: Harper Collins, 1996.
- [Von Krogh et al, 94] Von Krogh, G., J. Roos and K. Slocum, "An Essay on Corporate Epistemology", Strategic Management Journal, 15, 1994.
- [Wilson, 02] Wilson, T., "The Nonsense of Knowledge Management", Info Research, 8, 1, 2002.

Knowledge Management from an Ancient Chinese Point of View, or the Knowledge Flow through an Organisation

Werner Riegler
(memeticor@, Austria
office@memeticor.com)

Abstract: The ancient Chinese kind of thinking, as best known from TCM (Traditional Chinese Medicine) is very different from the modern western approach of scientific medicine. But there is powerful knowledge in both systems and the knowledge is complementary. It is obvious that our methods and techniques of knowledge management do not support the foreign kind of thinking. Though many of our trends have striking similarity to this ancient Chinese behaviour. Holistic approaches, synthesis versus analysis, networking, social web, empathy, mirror neurons, memetics, graphical versus textual communication. Many of the things we cannot understand are complex systems and for these systems our analytic behaviour in thinking does not seem to work. The straight rule of cause and effect does not work for complex systems, so our western philosophy based on causality fails. Our attempt in managing knowledge using analytical models based on causality will probably fail too. To break up information into sentences of words and looking for grammar (taxonomy), syntax and relations (ontology) cannot work. Because it is the wrong way of asking questions compared to the ancient Chinese philosophy. So what could be a useful approach in managing complex systems like knowledge in organisations from the ancient Chinese point of view?

Keywords: Artificial Intelligence, Knowledge Management, Semantic Web, Multimedia

Categories: I.2.4, I.2.10, I.2.6, K.6

1 Introduction

Practising TaiJi-Quan [Olson, 95] for many years and starting to teach gave me a totally new picture of knowledge and learning, and it finally gave me a clue of what the Chinese idea of Qi means. On the other hand I have been working for many years in document and knowledge management projects in big companies and I have started to develop a collaboration framework using a holistic approach with graphical, web based technologies [memeticor, 03]. Some upcoming scientific validations of how our brain works [Bauer, 05][Rizzolatti, 06] and many publications about how these effect our culture, encourage me to try to make a comparison about what it could mean to use the ancient Chinese methods of TCM to heal organisations. And this paper is a part of it, dealing with ancient Chinese methods for KM.

However, TCM is not yet accepted as a science, and an organisation is not a human being and does not have a body, a mind and a soul. But it is the analogy that is interesting to me, like the approach of memetics as an analogy to genetics [Blackmore, 99], which is helping us to get some very different points of view of our culture and how it, and consequently our collective knowledge, is changing.

Nevertheless an organisation consists of human beings and it shows behaviour akin to a human being [Bridges, 98]. If we hear just the brand name of an organisation or company we get a whole and complex picture of the nature of this organisation. And there are a lot of new forms of organisations using new methods and techniques that are more close to the ancient Chinese kind of thinking, than our western philosophy based on causality, such as virtual organisations like business platforms, XING [Lamprecht, 07] for instance and on-line communities like Wikipedia [wikipedia, 01].

2 Existing Methods of Managing Knowledge in Organisations

The very first approach of thinking about knowledge in organisation in my opinion was started in the US when Japanese companies became successful in the 1970s and 1980s. “The Knowledge-Creating Company” is the classic book in knowledge management until today [Takeuchi, 95]. The oneness of body and mind is mentioned there. But what happened in practice? Technology has changed, causing culture to change too. This nature of change brings us back to the ancient Chinese point of view.

2.1 Balanced Scorecard and Strategy Maps

A common method of trying to manage knowledge is a tool known from accounting and bookkeeping. The balanced scorecard is the allegory of the western kind of thinking. Knowledge is expressed in figures and numbers. We can calculate with them. The total is exactly the sum of the parts. Emergence is ignored.

There is an explicit approach to find the missing link and fit it into the existing schema: converting intangible assets into tangible outcomes – published in Strategy Maps [Norton, 04]. Norton is using graphical representations - maps and a holistic approach. Nevertheless is he breaking up information into numbers and the result is a highly complex system based on quantity and very difficult to handle.

2.2 Evolution and Self-Organisation

Human beings and their brains are not working this way. Though it is easier to handle numbers and figures in computer programs that are used to control knowledge processes, knowledge creation and knowledge flow do happen very different.

Some of us have had the guts to show that the approach in managing knowledge by figures and numbers ends up in highly complex systems and that these systems may not be controlled using analytical methods. Findings from quantum physics were used to demonstrate this: the Bremerman’s Limit [Malik, 03]. Malik shows that we need a holistic approach – he is using cybernetics – and that we have to “hope for” an effect of self-organisation. This is what we can learn from nature by evolution – what is also called bionic. “Not ignorance, but ignorance of ignorance is the death of knowledge” (A.N. Whitehead) is quoted in Malik’s book. Nevertheless Malik uses a causality based approach – cybernetics and quantum physics – to get to this conclusion.

2.3 Social Web

In the meantime, nearly unnoticed, we have got a lot of organisations that have changed the world. Prophets have been unheard, like Rick Levine with his “cluetrain manifesto” [Levine, 01]. But people are looking into Wikipedia [wikipedia, 01] first if they do not know the meaning of something or would like to know more about certain surroundings of something special. They are asking their contacts at XING [Lamprecht, 07] if they need some help or ideas before they ask their colleague or manager. Some standards have been ready to use for Web 2.0 for a long time. Taxonomy and Ontology may easily be solved using RDF and OWL [Birkenbihl, 05]. But is this technology the key for the flow of knowledge? In my opinion no, it isn't. It isn't used in the leading on-line organisations either. May be it will be used in future as e.g. by the chamber of commerce in Austria – the Austria Pro working group “Semantic Web”, where memeticor@ is also a member now. But here we use it for B2B E-Commerce [Austria Pro, 07]. It does not have anything to do with the flow of knowledge, it is all about the flow of numbers and figures. And the existing standards do what western kind of thinking is used to do: breaking up information into numbers.

3 The ancient Chinese kind of thinking and the holistic approach

The ancient Chinese philosophy is part of a culture which does not know a kind of Genesis. Therefore it is not necessary to ask for a cause to understand an effect. In the Taoist understanding there is an everlasting cycle of recreation by the effect of changing Yin and Yang by the rules of the 5 phases (wood, fire, earth, metal, water). It is a little bit strange for those who are not familiar with this kind of thinking. But it is very natural and also pragmatic for those who are. But why is this kind of thinking important for us “longnoses“ too? Because we are also part of nature and, believe it or not, we act by these rules too. Most important, it is a natural way of how our brain works and we can call this “thinking” too.

The essence is that we can synthesize single parts of information to a whole that is bigger than its parts. And we can imagine this whole just by small parts of it. These abilities we have from our mirror neurons [Bauer, 05][Rizzolatti, 06]. This means that evolution gave us the ability to detect a tiger by its stripes within a clump of bushes, generally speaking [Urban, 04]. This is also the key to empathy and it turns out to also be the key to our culture and the way we change(d) it by the flow of knowledge.

3.1 Pattern recognition – Quality instead of Quantity

Information may be described by numbers and figures or by using pictures and emotions. TCM uses mainly Yin and Yang to describe symptoms. There is either Yin (cold) or Yang (hot) and there is Yin within Yang and Yang within Yin as well. And Yin or Yang mean totally different things in a different context. So TCM uses the quality (importance for our behaviour in our environment) and not the quantity (comparing to standardized values) to describe symptoms. Even the organs (heart, liver, lung, ...) are not perceived as physical parts in our body but as the meaning of these parts for our health. These parts are working together and each of it has a certain meaning for the well being of a human.

And a single human has a meaning in the society (organisation) as well. Health is when the flow between organs and the flow between humans is harmonic. Lack or excess of either Yin or Yang in a single organ is just a single pattern. Not the quantity of the pulse rate but the quality – its pattern - is important to make the diagnosis in TCM. The heart gives us the ability to calm our Shen (spirit). This art of healing is very difficult and it takes a lot of experience to become an expert. The best way to equip others with this knowledge is to use pictures. Not only in a clinical picture and disease pattern but also in how to see the different meanings of one and the same pattern in different contexts. The whole thing could affect the single pattern in a very subtle way.

3.2 Flow and Disharmony

As knowledge is more than information, Qi [Kaptchuk, 83] is more than blood flowing in our body. Although it is almost impossible to translate Qi, it is commonly translated to something like “energy of our life”. Interesting is that this Qi flows in certain pathways through our body, the acupuncture meridians. But Qi is also in the air and in our food, so there are again different qualities of Qi.

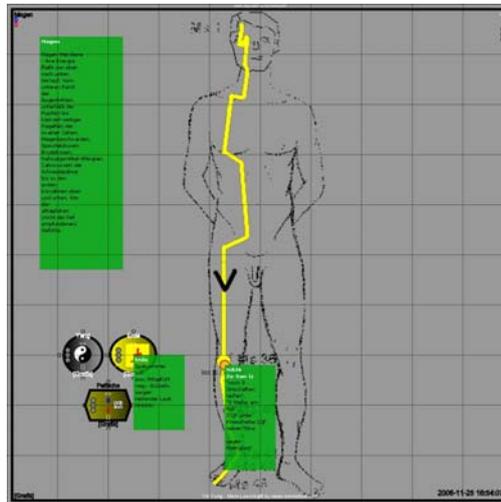


Figure 1: Acupuncture meridian, a pathway where Qi flows through our body

If the flow of Qi is not working properly the result is disharmony and illness. The way of forcing Qi to flow is ruled by the 5 phases, the cycle of mutual creation: wood (spring) creates fire and fire (summer) creates earth and so on. And there is a second cycle which is the cycle of mutual controlling: wood (liver) controls earth (spleen) controls water and so on. An example of a disharmony in creation is: less storage (water) in cause of less harvest (metal). An example of a disharmony in controlling is: excess storage (water) in cause of less control of growth (fire).

4 The analogy to the flow of knowledge in an organisation

As knowledge is getting experience by using certain information, the quality of knowledge may be described by Yin and Yang and knowledge creation and flow may be described using the principles of the 5 phases:

- birth (attraction, wood) – innovation, a new idea
- growth (interest, fire) – some interested parties adding information
- ripening (desire, earth) – someone want to use it for some purpose
- harvest (action, metal) – using it is useful in some cases
- storage (retention, water) – results stored for later use, creation of new ideas

4.1 Picture of an organisation (analogy to the 5 phases)

An organisation is a cooperation of human beings. It is essential that knowledge is flowing through the organisation harmonically for a healthy organisation.

The 5 elements controlling the flow of knowledge in an organisation could be:

- targets (wood) – vision, motivation for creating new ideas.
- ideas (water) – appraising ideas and gathering ideas to new innovations
- problems (fire) – mastering problems together
- communication (metal) – distilling the essence and share it with others
- learning (earth) – using information and experience

All these elements have their own subsystems in quality and cycles. There are typical patterns of quality for each context, because the same thing and quality may mean something different in an other context.

In creating ideas we need different patterns than in appraising them, or we need different patterns in solving a problem than in communicating it or in learning about. This perception does exactly match the mechanism of our graphical collaboration management system mem-topics©. And this was one of the reasons (patterns) why I was thinking about this analogy in depth (whole picture).

As a first guiding principle of the quality of the 5 elements seen by people in their organisation we can use the results of an on-line survey called “Fish Test” [Fish Test, 06]. This survey was automatically created by using our graphical collaboration management system mem-topics©. There were 10 chapters with 5 questions each to answer with a quality of -100% to +100% (similar to Yin and Yang).

targets	ideas	problems	communication	learning
51,67%	39,16%	40,21%	32,34%	37,69%

Table 1 : Results of the “Fish Test” – quality of the 5 elements in organisations shows that communication (flow of information) has the lowest overall quality

4.1.1 Flow of knowledge (analogy to acupuncture meridian)

Another interesting analogy is the flow of knowledge in an organisation compared with the flow of Qi in a human being.

There are also some certain pathways as we can see by using different case studies within the patterns of the big picture. There are planned pathways between the

elements resulting from the design patterns. And pathways are used, which are detectable by available data of the used design patterns in different use cases. The difference in the quality of the pathways may be seen in what planned pathways are less or excessively used (Yin and Yang). Any disharmony may be seen in certain use cases if there are missing pathways or excessively used ones.

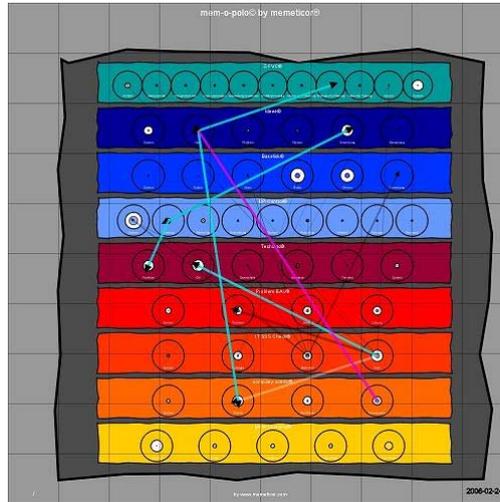


Figure 3 : mem-o-polo© - a pathway where knowledge flows through an organisation

4.1.2 Disharmony of the flow of knowledge in an organisation

Everybody knows what happens if the flow of knowledge in an organisation is in disharmony. But thinking in cause and effect again leads to breaking up information into numbers. And this is not useful if we want to harmonize the flow again. Often the disharmony is based on our intrinsic behaviour and again the ancient Chinese kind of thinking promises to be a more natural and powerful solution finding method.

Disharmony in the flow of knowledge in an organisation always stems from 2 points, one having a stagnation and another having a deficiency. The quality of stagnation and deficiency always depends on the context. But there are some well known pictures of typical diseases in my opinion (see Figure 4).

The key questions are how to detect the right clinical picture by looking at the symptoms we can see. And what is the right therapy. And to see and act in time.

4.2 Using TCM for an organisation

An analogy for TCM could be very useful for healing organisations, because organisations are highly complex systems, especially when they are knowledge intensive. Precondition to be useful is that dealing with information in this context is changed from causality and quantity to holistic and quality. When an organisation is treated in a TCM-like model, the parts must not be seen physically but in their meaning for the whole organisation.

To detect disharmony in the flow of knowledge we can identify boundaries, blockades and fear (see Figure 4). As therapy we use and suggest special training for healing wood based boundaries – wrong organisations [mem-training, 07], systematic graphical communication for healing fire based blockades – wrong tools [mem-topics, 06], a special form of benchmarking and PR for healing water based fear – wrong culture [mem-power, 06].

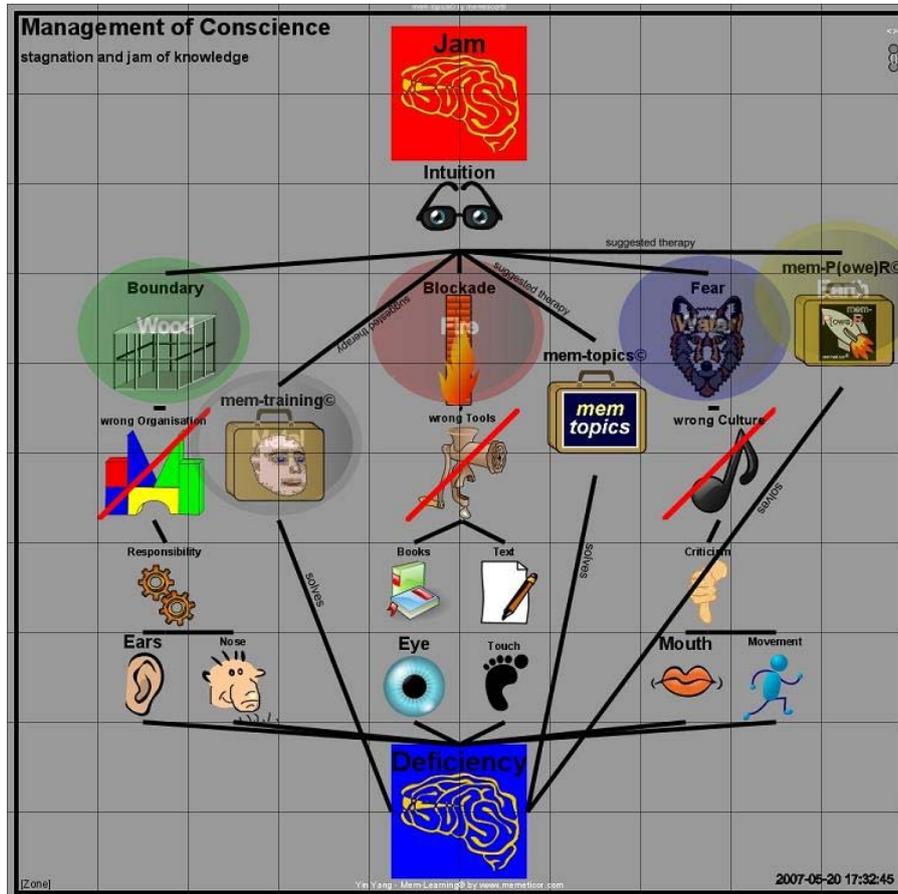


Figure 4: Disharmony of the flow of knowledge in an organisation and therapy

5 Conclusions

The current work with the analogy of TCM for KM demonstrates that this could be a very efficient and pragmatic way in “healing organisations”. To know the right pathway of knowledge flow and to see the quality of the meanings in the right context may help a lot for getting the right therapy in time.

6 Future Work

In the near future we plan to identify and standardize “hot spots” in analogy of the acupuncture points located on the pathways.

Therapies are very individual and we need more experience from different use cases to get therapies that are more easier to apply. The Chinese culture needed several 1000 years to pool its expert knowledge. We hope to have some advantages to accomplish the analogy for organisations in a smaller amount of time:

1. computer and web based graphical communication
2. web based surveys and
3. social on-line communities (Web 2.0)

Future interfaces will support RDF and OWL. XML and SVG are already in use in mem-topics©. Making taxonomy and ontology exchangeable along with the patterns and pictures may speed up the information gathering process though it is just machine interpretable information.

References

- [Austria Pro, 07] Austria Pro Working Group “Semantic Web”, 2007, <http://www.austriapro.at/>
- [Bauer, 05] Joachim Bauer, Warum ich fühle was du fühlst, Hoffmann und Campe, 2005.
- [Birkenbihl, 05] W3C-Standards und Semantic Web - Tutorial, <http://www.w3c.de/PubPraes/>
- [Blackmore, 99] Susan Blackmore, The Meme Machine, Oxford, 1999
- [Bridges, 98] The character of organisations, 1998.
- [Fish Test, 06] FishTest, Test the soft-facts of your organisation, www.memeticor.com/fischtest
- [Kaptchuk, 83] Ted J. Kaptchuk. Chinese Medicine : The Web that has no Weaver, 1983.
- [Lamprecht, 07] Stephan Lamprecht, XING Das Buch, Heise, 2007, <http://www.xing.com/>
- [Levine, 01] Rick Levine: The Cluetrain Manifesto, Perseus Pubkliching, 2001
- [Malik, 03] Fredmund Malik, Systemisches Management, Evolution, Selbstorganisation, 2003.
- [Norton, 04] , David P. Norton, Robert S. Kaplan, Strategy Maps, 2004.
- [Olson, 95] S. Olson. The Intrinsic Energies of Tai Chi Chuan, Dragon Publications, 1995.
- [Rizzolatti, 06] G. Rizzolatti, L. Fogassi, V. Gallese, Mirrors in the Mind, Sf. American, 2006.
- [memeticor, 03] W. Riegler, The best Ideas ‘from Scheibbs to Nebraska’, www.memeticor.com
- [mem-training, 07] mem-training, here meme gets trained, <http://www.memeticor.com/training/>
- [mem-topics, 06] mem-topics, takes your ideas home, <http://www.mem-topics.info.com/>
- [mem-power, 06] mem-power, let your knowledge work for you, <http://www.mem-power.info/>
- [Takeuchi, 95] Ikujiro Nonaka, Hirotaka Takeuchi, The Knowledge-Creating Company, 1995.
- [Urban, 04] Martin Urban, Wie die Welt im Kopf entsteht, Lübbe, 2004.
- [wikipedia, 01] Wikipedia, The free Encyclopedia, 2001, <http://www.wikipedia.org/>

Emerging Data Mining Applications: Advantages and Threats

Narayanan Kulathuramaiyer

(Institute for Information Systems and Computer Media (IICM)
Graz University of Technology, Austria
nara@iicm.edu)

Hermann Maurer

(Institute for Information Systems and Computer Media (IICM)
Graz University of Technology, Austria
hmaurer@iicm.edu)

Abstract: Data Mining describes a technology that discovers non-trivial hidden patterns in a large collection of data. Although this technology has a tremendous impact on our lives, the invaluable contributions of this invisible technology often go unnoticed. This paper addresses the various forms of data mining while providing insights into its expanding role in enriching our life. Emerging forms of data mining are able to perform multidimensional mining on a wide variety of heterogeneous data sources, providing solutions to many problems. This paper highlights the advantages and disadvantages arising from the ever-expanding scope of data mining. Data Mining augments human intelligence by equipping us with a wealth of knowledge and by empowering us to perform our daily task better. As the mining scope and capacity increases, users and organisations become more willing to compromise privacy. The huge data stores of the 'master miners' allow them to gain deep insights into individual lifestyles and their social and behavioural patterns. The data on business and financial trends together with the ability to deterministically track market changes will allow an unprecedented manipulation of the stock market. Is it then possible to constrain the scope of mining while delivering the promise of better life?

Keywords: Data Mining, Knowledge Discovery, Social Aspects

Categories: H.4.0, I.2.6, K.4.2

1 Introduction

As we become overwhelmed by an influx of data, Data Mining presents a refreshing means to deal with this onslaught. Data Mining thus holds the key to many unresolved age-old problems. Having access to data thus becomes a powerful capability which can be effectively be harnessed by sophisticated mining software. Data at the hands of credit card companies will allow them to profile customers according to lifestyles,

spending patterns and brand loyalty. Political parties on the other hand are able to predict with reasonable accuracy how voters are likely to vote. [Rash, 2006]

According to [Han and Kamber, 2006] data mining is defined as the extraction of interesting (non trivial, implicit, previously unknown and potentially useful) information or patterns from data in large databases. We take a broad view of data mining, where we also include other related machine based discoveries such as deductive query processing and visual data mining. Databases may include both structured data (in relational databases), semi structured data (e.g. metadata in XML documents) as well as unstructured documents such as text documents and multimedia content.

Despite the success stories in areas such as customer relationship modelling, fraud detection, banking, [KDD, 2005], the majority of applications tend to employ generic approaches and lack due integration with workflow systems. As such, Data Mining is currently at a chasm state and has yet to become widely adopted by the large majority [Han and Kamber, 2006].

2 Data Mining Process

Data Mining typically describes an automated acquisition of knowledge from a large collection of data. This traditional view corresponds to the knowledge creation phase in knowledge management. Current developments of data mining have expanded this to also cover support for knowledge organization and consolidation with respect to existing domain knowledge, and data visualization for iterative learning. Data Mining can thus be seen as complementing and supplementing knowledge management in a variety of phases.

In order to describe the processes involved in performing data mining we divide it into 3 phases: domain focusing, model construction (actual mining using machine learning algorithms), and decision making (applying the model to unseen instances).

Data focusing phase involves the application of some form of clustering or may incorporate intensive knowledge engineering for complex applications. At the model construction phase, a model of generalized patterns is constructed to capture the intrinsic patterns stored in the data.

The model generated is then employed for decision-making. Simplistic applications of data mining tend to merely employ the model to predict the likelihood of events and occurrences, based largely on past patterns. Amazon, for example, is able to recommend books according to a user's profile. Similarly, network operators are able to track fraudulent activities in the usage of phone lines by tracking deviation patterns as compared to standard usage characterization.

3 Applications of Data Mining

3.1 Web Search As Data Mining

Search engines have turned the Web into a massive data warehouse as well as a playground for automated discovery of hidden treasures. Web Search is thus viewed as an extensive form of multidimensional heterogeneous mining of a largely

unstructured database for uncovering an unlimited number of mind-boggling facts. The scale of data available is in the range of peta bytes, and it much greater than the terra bytes of data available at the hands of large global corporations such as Walmart.

Search engines can either simultaneously or incrementally mine these datasets to provide a variety of search results which include phone contacts, street addresses, news feeds, dynamic web content, images, video, audio, speech, books, artifacts. In performing domain focusing, a model allowing to characterize aggregated user search behaviour is used [Coole et al, 1997]. This phase could involve associational subject link analysis, requiring a contextual domain analysis (for mobile users). This Mining phase involves the derivation of aggregated usage profiles based on a multidimensional mining of usage patterns according to clustered characterization. By analyzing search history over a period of time, search engines have access to a great deal of insights into lives of presumably 'anonymous' searchers. A search query can indicate the intent of a user to acquire particular information to accomplish a task at hand. Search engines can thus track patterns and drifts in global user intentions, moods, and thoughts.

3.2 Environmental Modelling Application

There are complex problems for which data mining could be used to provide answers by uncovering patterns hidden beneath layers of data. In many cases, domain focusing has in the past has been the biggest challenge. Data mining could be employed for the modeling of environmental conditions in the development of an early warning system to address a wide range of natural disasters such as avalanches, landslides, tsunami and other environment events such as global warming. The main challenge in addressing such a problem is in the lack of understanding of structural patterns characterizing various parameters which may currently not be known.

As highlighted by [Maurer et al], although a large variety of computer-based methods have been used for the prediction of natural disasters, the ideal instrument for forecasting has not been found yet. As highlighted in their paper, there are also situations whereby novel techniques have been employed but only to a narrow domain of limited circumstances.

Integration of multiple databases and the compilation of new sources of data are required in the development of full-scale environmental solutions. As advances in technology allow the construction of massive databases through the availability of new modes of input such as multimedia data and other forms of sensory data, data mining could well provide a solution. In order to shed insights on a complex problem such as this, massive databases that were not previously available need to be constructed e.g. data about after event situations of the past [Maurer et al, 2007]. Such data on past events could be useful in highlighting patterns related to potentially in-danger sites. Data to be employed in this mining will thus comprise of both of weather and terrestrial parameters together with other human induced parameters such as vegetation or deforestation over a period of time. [Maurer et al, 2007]

Domain focusing will be concerned with discovery of causal relationships (e.g. using Bayes networks) as a modeling step. Multiple sources of data which include new sources of data need to be incorporated in the discovery of likely causal relationship patterns. A complex form of data mining is required even at the phase of domain focusing. This will involve an iterative process whereby hypothesis

generation could be employed to narrow the scope of the problem to allow for a constrained but meaningful data collection. For complex domains such as this, unconstrained data collection may not always be the best solution. Domain focusing would thus perform problem detection, finding deterministic factors and to hypothesize relationships that will be applied in the model. [Beulens et al, 2006] describe a similarly complex representation for an early warning system for food supply networks.

Subsequently, the model construction phase will employ a variety of learning algorithms, to profile events or entities being modeled. As this stage may negate model relationships, domain focusing will need to be repeated and iteratively performed. The model construction phase will allow the incremental development of a model, based on a complex representation of the causal networks. [Beulens et al, 2006]

Mining methods such as clustering, associational rule mining, neural networks will be used to verify the validity of causal associations. Once a potential causal link is hypothesized, verification can be done via the application of data mining methods. [Beulens et al, 2006], have proposed a combination of approaches which include deviation detection, classification, dependence model and causal model generation.

The Decision Making phase will then apply the validated causal relationship model in exploring life case studies. An environment for an interactive explorative visual domain focusing is crucial, to highlight directions for further research. Data mining could serve as a means of characterization of profiles for both areas prone to disasters or those that are safe.

3.3 Medical Application

We will briefly discuss another form of mining that has a high impact. In the medical domain, data mining can be applied to discover unknown causes to diseases such as 'sudden death' syndrome or heart attacks which remain unresolved in the medical domain. The main difficulty in performing such discoveries is also in collecting the data necessary to make rational judgments. Large databases need to be developed to provide the modeling capabilities. These databases will comprise of clinical data on patients found to have the disease, and those who are free of it. Additionally non-traditional data such as including retail sales data to determine the purchase of drugs, and calls to emergency rooms together with auxiliary data such as micro array data in genomic databases and environmental data would also be required. [Li, 2007]

Non traditional data could also incorporate major emotional states of patients by analyzing and clustering the magnetic field of human brains which can be measured non invasively using electrodes to a persons' heads. [Maurer et al, 2007] Social patterns can also be determined through profile mining as described in the previous section to augment the findings of this system. Findings of functional behavior of humans via the genomic database mining would also serve as a meaningful input.

The development of large databases for medical explorations will also open possibilities for other discoveries such as mining family medical history and survival analysis to predict life spans. [Han and Kamber, 2006]

4 The Advantages of Data Mining

Data mining has crept into our lives in a variety of forms. It has empowered individuals across the world to vastly improve the capacity of decision making in focussed areas. Powerful mining tools are going to become available for a large number of people in the near future.

The benefits of data mining will include preserving domestic security through a number of surveillance systems, providing better health through medical mining applications, protection against many other forms of intriguing dangers, and access to just-in-time technology to address specific needs. Mining will provide companies effective means of managing and utilizing resources. People and organizations will acquire the ability to perform well-informed (and possibly well-researched) decision-making. Data mining also provides answers through sifting through multiple sources of information which were never known to exist, or could not be conceivably acquired to provide enlightening answers. Data Mining could be combined with collaborative tools to further facilitate and enhance decision-making in a variety of ways. Data mining is thus able to explicate personal or organizational knowledge which may be locked in the heads of individuals (tacit knowledge) or in legacy databases, to become available. Many more new benefits will emerge as technology advances.

5 Disadvantages of Data Mining

A great deal of knowledge about users is also being maintained by governments, airlines, medical profiles or shopping consortiums. Mining applications with dramatic privacy infringement implications include search history, real-time outbreak and disease surveillance program, early warning for bio-terrorism [Spice, 2005] and Total Information Awareness program.[Anderson, 2004] For example, search history data represents an extremely personal flow of thought patterns of users that reflects ones quest for knowledge, curiosity, desires, aspirations, as well as social inclinations and tendencies. Such logs can reveal a large amount of psychographic data such as user's attitudes towards topics, interests, lifestyles, intents and beliefs. A valid concern would be that the slightest leak could be disastrous. The extent of possible discoveries has been clearly illustrated by the incidence where AOL released personal data of 658,000 subscribers [Jones, 2006].

Another common danger is profiling where there is a possibility of drastic implications such as a conviction being made based on the incriminating evidences of mining results. There is also a danger of over-generalization based on factors such as race, ethnicity, or gender. This could result in false positives, where an entirely innocent individual or group is targeted for investigation based on a poor decision making process. For the domestic security application, a reasonable data mining success rate of 80% implies that 20% of all US citizens (or 48 million people) would be considered false positives [Manjoo, 2002].

Data mining will further empower mining kingpins to be able to go beyond the ability to PREDICT what is going to happen in a number of areas of economic importance, but actually have the power to KNOW what will happen, hence can e.g.

exploiting the stock market in an unprecedented way. They also have the capacity to make judgments on issues and persons with scary accuracy.

6 What can we do?

A majority of related works are more concerned about privacy, but it is no longer the main issue of concern. [Kovatcheva, 2002] has proposed a means of protecting the anonymity by the use of anonymity agents and pseudonym agents to avoid users from being identified. Their paper also proposed the use of negotiation and trust agents to assist users in reviewing a request from a service before making a rational decision of allowing the use of personal data.

A similar agent-based approach is described by [Taipale, 2003] via rule-based processing. An "intelligent agent" is used for dispatching a query to distributed databases. The agent will negotiate access and permitted uses for each database. Data items are labeled with meta-data describing how that item must be processed. Thus, a data item is protected as it retains relevant rules by which it describes the way it has to be processed. The main challenge then lies in coming up with guidelines and rules such that site administrators or software agents can use to direct various analyses on data without compromising the identity of an individual user. This approach however is not applicable for areas such as Web search, where a standard framework for conscientious mining is far from sight. Furthermore, the concern with the emergence of extensive mining is no longer solved by addressing privacy issues only. As more and more people are willing to compromise privacy, the questions that we pose are: Who do we trust as the gatekeeper of all our data? Do we then trust all our private data at the hands of a commercial global company?

One approach to overcome the concerns mentioned above is by employing a distributed data mining approach, where separate agencies will maintain and become responsible and accountable for different (independent segments of) data repositories. This proposal further ensures that no central agency will have an exclusive control over the powerful mining technology and all resources. [Kulathuramaiyer, Maurer, 2007] In order to realize this solution, governments have to start playing a more proactive role in maintaining and preserving national or regional data sources. We are also beginning to see partnerships between global search engines and governments in this respect. Such partnerships should however be built upon a balanced distribution of earnings. Such a solution can be complemented by the nurturing of a larger number of autonomous context-specific or task specific miners.

7 Conclusion

As data mining matures and becomes widely deployed in even more encompassing ways, we need to learn to effectively apply it to enrich our lives. At the same time, the dangers associated with this technology needs to be minimized by deliberate efforts on the part of the enforcement agencies, data mining agencies and the users of the system. There should be strict regulations to prevent the abuse or misuse of data. Users should also be made aware of the privacy policies in order to make an informed

decision about revealing their personal data. The success of such regulations and guidelines can only be guaranteed if they are backed up by a legal framework.

References

[Anderson, 2004], Anderson, S.R., Total Information Awareness and Beyond, Bill of Rights Defense Committee; White paper. The Dangers of Using Data Mining Technology to Prevent Terrorism, July 2004

[Beulens et al, 2006] Beulens, A., Li, Y., Kramer, M., van der Vorst, J., Possibilities for applying data mining for early Warning in Food Supply Networks, CSM'06, 20th Workshop on Methodologies and Tools for Complex System Modeling and Integrated Policy Assessment, August, 2006
http://www.iiasa.ac.at/~marek/ftppub/Pubs/csm06/beulens_pap.pdf

[Coole et al, 1997] Coole, R. Mobasher, B., Srivastava, J., Grouping Web Page References into Transactions for Mining World Wide Web Browsing Patterns, Proceedings of the 1997 IEEE Knowledge and Data Engineering Exchange Workshop Page: 2, 1997 ISBN:0-8186-8230-2 IEEE Computer Society

[Han and Kamber, 2006] Han, J., and Kamber, M., Data Mining: Concepts and Techniques, 2nd ed., The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor , Morgan Kaufmann Publishers, March 2006.

[Jenssen, 2002] Jenssen, D., Data mining in networks. Invited talk to the Roundtable on Social and Behavior Sciences and Terrorism. National Research Council, Division of Behavioral and Social Sciences and Education, Committee on Law and Justice. Washington, DC. December 11, 2002

[Jones, 2007] Jones, K .C., Fallout From AOL's Data Leak Is Just Beginning ,
<http://www.informationweek.com/news/showArticle.jhtml?articleID=191900935>,
 accessed 2007

[KDD, 2005] KDDnuggets : Polls: Successful Data Mining Applications, July 2005
http://www.kdnuggets.com/polls/2005/successful_data_mining_applications.htm

[Kovatcheva, 2002] Kovatcheva, E., Tainen, H., The technological and social aspects of data mining by means of web server access logs
<http://www.pafis.shh.fi/~elikov02/SFISWS2/SFIS2.html> 18 January 2002

[Kulathuramaiyer, Balke, 2006]Kulathuramaiyer, N., Balke, W.-T., Restricting the View and Connecting the Dots — Dangers of a Web Search Engine Monopoly, J,UCS Vol. 12 , Issue 12, pp.1731 – 1740, 2006

[Kulathuramaiyer N., Maurer, H., 2007], Kulathuramaiyer, N., Maurer, H., "Why is Fighting Plagiarism and IPR Violation Suddenly of Paramount Importance?" Proc. of International Conference on Knowledge Management, Vienna, August 2007.

[Li, 2007] Li, C. S., Survey of Early Warning Systems for Environmental and Public Health Applications, in Wong, S., Li, C. S.,(eds.), Life Science Data Mining, Science, Engineering, and Biology Informatics- Vol. 2, 2007

http://www.worldscibooks.com/compsci/etextbook/6268/6268_chap01.pdf

[Manjoo, 2002] Manjoo, F., Is Big Brother Our Only Hope Against Bin Laden?, Dec. 3, 2002 http://www.salon.com/tech/feature/2002/12/03/tia/index_np.html

[Maurer et al., 2007] Maurer, L., Klingler, C., Pachauri, R. K., and Tochtermann, K., Data Mining as Tool for Protection against Avalanches and Landslides, Proc. Environmental Informatics Conference, Warsaw, 2007

[Milne, 2000] Milne, G. R., Privacy and ethical issues in database/interactive marketing and public policy: A research framework and overview of the special issue, *Journal of Public Policy & Marketing*, Spring 2000

[Mobasher, 2005] Mobasher, B., Web Usage Mining and Personalisation, in Singh, M. P. (ed.) *Practical Handbook of Internet Computing*, Chapman & Hall/ CRC Press, 2005 <http://maya.cs.depaul.edu/~mobasher/papers/IC-Handbook-04.pdf>

[Rash, 2006] Rash, W., Political Parties Reap Data Mining Benefits, *eWeek.com enterprise News and reviews*, November 16, 2006
<http://www.eWeek.com/article2/0,1895,2060543,00.asp>

[Spice, 2003] Spice, B., Privacy in age of data mining topic of workshop at CMU, March 28, 2003 <http://www.post-gazette.com/nation/20030328snoopingnat4p4.asp>

[Taipale, 2003] Taipale, K.A. "Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data". *Colum. Sci. & Tech. L. Rev.* 5 (2). SSRN 546782 / OCLC 45263753, December 15, 2003.

Front End Decision Making: How Initiators Decide to Reveal Their Knowledge about Radical Opportunities

Dietfried Globocnik

(University of Graz, Austria
dietfried.globocnik@domino.uni-graz.at)

Stefan Vorbach

(University of Graz, Austria
stefan.vorbach@uni-graz.at)

Abstract: Radical innovation becomes a more important capability of today's companies. It involves high resource commitment, a long range perspective and the acceptance of risk and uncertainty. As this paper will show, this is not only a question of resources but about organizational system architecture and individual characteristics. We focus the Front End of the innovation process and the individual behaviour of the initiator, who is the key person in this phase. He first recognizes the radical opportunity in terms of a potential product-market combination which is new to the world. The information processing behaviour differs from the usual bottom up initiatives driven by suggestion systems, continuous improvement processes or kaizen. The information resp. knowledge about radical opportunities is not visible for anybody else in the organization until the initiator starts its transition at the organizational level. Otherwise it stays hidden inside the initiator.

The transition needs high personal initiative and overcoming internal and external barriers. The tasks include predevelopment activities and issue selling. Radical innovations comprehend risk in respect of uncertain feasibility, costs and customer acceptance. Besides this organisational perspective the initiator has additional personal risks, concerning career, and as a consequence payment level and even job security. That is why prior revealing his information he has to take all potential consequences of the initiative into account. In this paper we focus this specific decision and determinants of the opportunity recognizer's behaviour to take charge of the radical opportunity - an essential question not addressed in innovation research so far.

Unlike entrepreneurs, initiators act inside an existing organizational context with specific strategies, procedures, rules and norms. Moreover, every initiator has certain characteristics like traits, experiences and knowledge. We suggest that the personal characteristics and organizational environment influence the individual risk perception of the radical opportunity and the following decision in a specific way. The organizational variables can be configured as such that risk perception decreases and a higher motivation and target level are set. Hence, the top management will soon get guidelines of organizational design to indirectly influence this individual decision and to get to know about more radical opportunities. Then they can evaluate them which is especially important for building a balanced innovation portfolio.

The article helps getting a deeper understanding of the information processing behaviour of initiators in front end innovation processes and therefore has practical relevance for all organisations which try to intensify their innovation activities.

Keywords: front end, radical innovation, information processing, discontinuity, risky choice, cognition

Categories: A.0, A.1, A.m.

1 Introduction

Innovation and renewal are important organizational abilities of today's companies to survive. At the *macro level* technological progress and convergence, competitors, customers and environment exert pressure on creating new products, processes and structures to meet the stakeholder's expectations [e.g. Hauschildt 2004, Prahalad 1998]. For example on the technological level, competence destroying discontinuities disrupt the existing market organization of an industry because of the necessity of completely new skills, knowledge, substitution of product classes and the fall of entry barriers [Tushman/Anderson 1986]. Thus, even large established firms need to create and implement game-changing innovations – both as frontiers or adopters – to survive, since incremental next-level product development is not sufficient anymore [e.g. Leifer et al. 2001] and innovation is no longer an option [Grochla 1980].

When analyzing firms as *social systems*, we observe norms, values, routines and structures – all built to stabilize and align the perception resp. behaviour of its members. The arising organizational inertia leads to the paradox, that "... every social system tries to avoid innovation while it is in fact in need of it." [Pohlmann 2005:11]. Markets and technologies which are not familiar to the company are neglected because they are seen as risk for the operating business and legitimized theories in use [e.g. Sharma 1999, Levinthal/March 1993].

While past research primary focused activities and conditions for already organizationally legitimized ideas and their creation of ideas, another interesting question is about the intermediate stage, which contains activities for the transition at the organizational level. This needs the attention to the *individual* in the innovation process. Individuals suffer from a bounded rationality because of difficulties with imperfect information and anticipating future events. Organizational mechanisms help to focus on relevant environments to overcome the limited information processing capabilities [March/Simon 1958, Simon 1976]. Thus, cognition depends from individual capabilities and the specific organizational context, which both determine what information is labelled as relevant.

2 Focus of Interest

This paper addresses the above mentioned paradox in the front end of the innovation process. The leading questions are: what are the characteristics of this phase, who is responsible for its tasks, and how can the decision situation be described from the viewpoint of information processing?

(i) *Front End Innovation Process (FEIP)*: The FEIP is characterized by little resource commitment, experimental nature of work which is difficult to formalize, speculative revenue expectations and low top management involvement [Kim/Wilemon 2002, Koen et al. 2001, Murphy/Kumar 1997, Doll/Zhang 2001]. It includes the following activities: idea generation, opportunity recognition, initiative decision, predevelopment, issue selling activities, and finally the top management decision about resuming the initiative with suitable resources [e.g. Khurana/Rosenthal 1997, 1998, Burgelman 1983, Dutton/Ashford 1993], or on other words, activities of

information gathering, processing and diffusion. This paper focuses solely on the initiative decision process. A positive decision leads to above-named activities. In case of a negative decision there is no further information gathering, exploration and communication of the opportunity to convince top management of the feasibility and potential profitability of the new direction.

(ii) Initiator: This person is the main driver of the FEIP and is usually situated in the lower/middle management in a boundary spanning position with marketing and technical capabilities [e.g. Floyd/Wooldridge 1997, O'Connor/Rice 2000]. Not necessarily being the creator of the idea, he recognizes the potential commercialization of a product-market combination. Depending on personal and context specific characteristics he decides about the transition of the opportunity at the organizational level which would be associated with high personal involvement [Hornsby et al. 2002, Frese et al. 1997, Frohmann 1999, Hauschildt 2004, Quinn 1985].

(iii) Organizational Context: Unlike entrepreneurs, initiators operate within existing organizational contexts. The Initiator benefits among others from risk spreading, operational and administrative support, but suffers from limited control and is in need of sponsorship from a superior [e.g. Luchsinger/Bagby 1987]. Work organization, corporate culture, management support, reward system and resource endowment exogenously limit the scope of perception and action [e.g. Zhara et al. 2004, Kuratko et al. 2005, Hornsby et al. 1999, 2000, Antoncic/Hisrich 2001].

(iv) Decision setting: Radical innovation projects exhibit very critical characteristics, including nonlinear long term projects with sporadic stops and revivals, unpredictability because of multidimensional uncertainty and equivocality, instable markets, cannibalization risk and high impact on strategy and organization [Leifer et al. 2000, Rice 1999, Mensel 2004, Daneels/Kleinschmidt 2001]. The radical opportunity in terms of a vague vision about a product-market combination contains additional uncertainty because of the little information available [Brauner/Becker 2006], which cannot be improved at that point. If this is taken into account, this early decision of the initiator is characterized by (1) potential individual and organizational consequences of the decision (not) to start a radical initiative, (2) the significance of these consequences and (3) the uncertainty of the former in terms of unclear probabilities and high outcome variability [basically following Yates/Stone 1992, Forlani/Mullins 2000]. Within an intrapersonal, subjective process of argumentation, the initiator then has to estimate those variables to decide between the options “start an initiative” or “retain at the status quo”. That is why it is a decision under risk and not one under uncertainty.

The *research question* covering these four aspects is: What is the relative importance and influence of individual and organizational characteristics within the process of information processing, risk perception and decision making of the initiator in the context of radical innovation project during the early FEIP? This work aspires to a deeper understanding of the subjective, cognitive risk construct. The focused process and its context are illustrated in figure 1. For research this paper seems to be

relevant because of (i) analyzing a not observable information processing stage, (ii) individual decision behaviour in the FEIP (iii) of the in research often neglected middle management (iv) about radical innovations (v) within organisational context and (vi) through cognitive perspective. The expected insights are relevant for management regarding OHRM, Strategy and Leadership when configuring the innovation system.

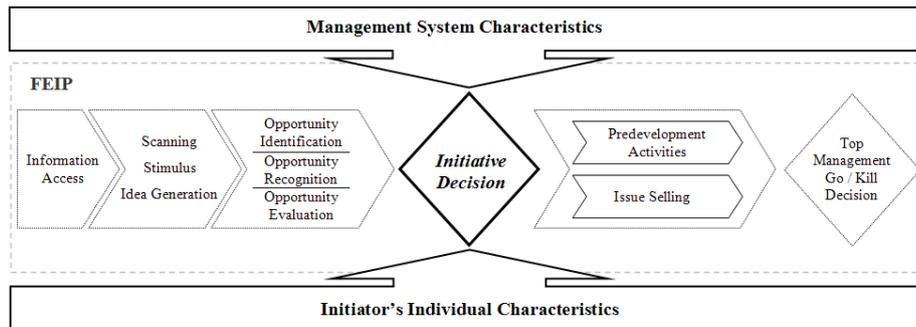


Figure 1: Illustration of the focus of interest and its procedural context

3 State of the Art

In this chapter we briefly describe the current state of the science concerning the Front End Innovation Process and risky decision making.

3.1 FEIP Research

Following Brown/Eisenhardt's [1995] categorisation of NPD research the FEIP literature can be divided into three streams:

(i) **Rational Plan Research (RPR)**: This atheoretical/realtheoretic stream focuses on the direct relationship between predevelopment activities resp. system variables and project resp. financial success. Cooper et al. identified FEIP factors of success such as preliminary market and technology assessment, early product definition and sufficient funding at the project level and clear strategy, management commitment and formalization at the system level [e.g. Cooper 1988, Cooper/Kleinschmidt 1986, 1988, 1993, 1995]. Other authors confirmed these insights [see review by Ernst 2002, meta-analytic reviews by Montoya-Weiss/Calantone 1994, Henard/Szymanski 2001]. Nevertheless this research stream is full of contradictions because many studies disregard important contingencies, especially the degree of innovativeness. When distinguishing between incremental and radical innovations within the context of the latter, classic instruments of market research and financial planning do not work, but strategic planning and product definition develop over a longer period in an experimental learning process [e.g. Song/Montoya-Weiss 1998, Veryzer 1998]. That is the reason why Khurana/Rosenthal [1997] have proposed more than one universal

formalized model, although they did not explicitly make this differentiation with respect to innovativeness.

(ii) Communication Web Work: This stream expands RPR by introducing the factor communication. The individual then becomes the focus of analysis. Relevant for this paper are the roles of the opportunity recognizer who finds potential markets for an invention [O'Connor/Rice 2001, Rice 1999], and the initiator who in addition, proactively forces the transition on the organizational level [Frohmann 1999]. Their personal characteristics are expertise, influence on management, network competence, creativity and risk seeking [e.g. Vojak et al. 2006, Stevens/Burley 2003, Howell/Higgins 1990].

(iii) Disciplined Problem Solving: This stream focuses on the influence of work organization and management practices on non-monetary success factors and individual behaviour. Especially autonomy, organizational barriers, management support and compensation were found critical for the initiative phase [e.g. Hornsby et al. 2002, Burgelman 1983, Tatikonda/Rosenthal 2000]. Despite the partial results, an integrating view to understand individual decision making and the impact of the organization is still missing.

3.2 Risky Decision Behaviour Research

In innovation research risk behaviour is a result of risk propensity which was mostly formulated as a *stable trait*. One stream of risky behaviour literature shares this opinion and could also find some empirical evidence relating to this. Here, risk propensity depends on position, income, age, experience [MacCrimmon/Wehrung 1990] and target level [March/Shapira 1987]. Additionally the decision itself depends on stable systematic judgement biases [e.g. Palich/Bugby 1995].

In the second stream – with its most prominent representative Prospect Theory – risk behaviour depends on the current *situation*. These situations which are positively (negatively) framed compared to a reference point and lead to risk avers (seeking) behaviour [Kahneman/Tversky 1979]. Despite good empirical evidence in experiments, controversial research outcomes indicate the need for further refinements. For example risk behaviour takes the opposite direction when “outcome history” as additional situative factor is manipulated [Thaler/Johnson 1990]. Imirical results indicate that Prospect Theory loses its prognostic explanation when situations become more complex.

This led to the *reconceptualization by Sitkin/Pablo* who introduced an additional *cognitive* level. In this model, the influence of risk propensity on the decision behaviour is (partially) mediated by risk perception. Risk propensity – depending on personal characteristics – is not a stable trait. The characteristics of the problem and the decision context as for example the organizational system influence the risk perception and do not determine decision behaviour directly [Sitkin/Pablo 1992]. Thus, risky decision behaviour depends on trait and perception of the situative factors [Simon et al. 2000]. Empirical evidence exists for different variants of this model [e.g. Sitkin/Weingart 1995, Weber/Milliman 1997, Williams/Voon 1999, Simon et al. 2000, Keh et al. 2002]. However, especially the organizational context was widely neglected so far. This approach seems to be valuable for the present paper. Even

Khurana/Rosenthal [1997] highlighted the importance of those “foundation elements” in the early FEIP.

4 Theoretical Development of the Front End Decision Making (FEDM) Framework

First it must be mentioned that there are only a few FEIP models at the individual level which describe what cognitive processes this decision involves. By synthesizing two models of stress research [Lazarus 1991, Gebert 2002] and R&D management [Baker et al. 1980] three conditions have to be fulfilled to take charge: (1) a performance gap is recognized, otherwise the aspiration level needs to be reconfigured upside to do so; (2) the initiator must perceive the change as capable, otherwise the aspiration level may be moved down to eliminate the gap or it is simply neglected; (3) the alternative which evolves from the problem solving process must match the recognized gap in terms of solving the organizational problem and expecting individual rewards. Figure 2 shows this descriptive process model.

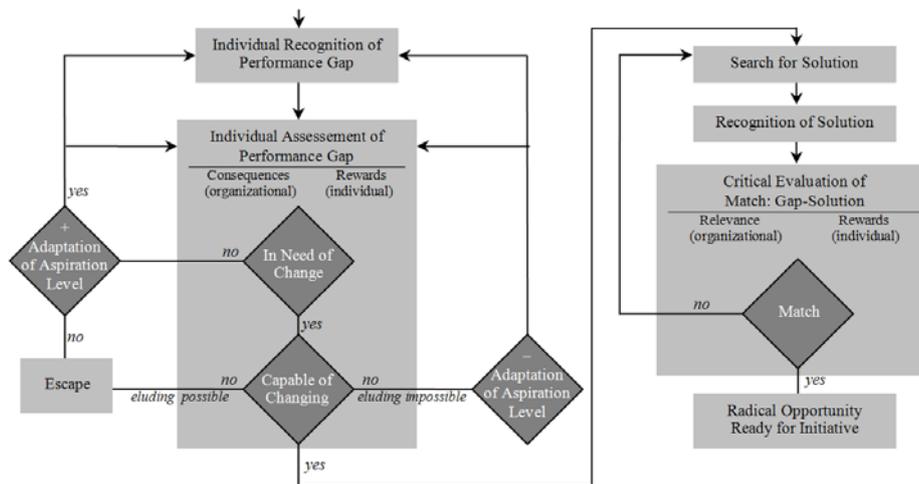


Figure 2: Individual Level Initiative Process Model

Based on the Sitkin/Pablo approach this paper modifies their framework in regard to the described research question and decision setting of radical opportunities. First the elements of the theory are described, followed by the expected interactions.

4.1 Situation of radical Opportunity

As mentioned above the initial situation consists of a product-market combination, which is new to the world. This involves potential multidimensional consequences for the initiator and the organization, which have to be taken into account when the

initiator decides about further investigation, revealing his vision and selling it to the top management.

4.2 Management System

At this level the management practices which are identified as success factors within the atheoretical/realtheoretical stream of the FEIP literature are abstracted on central configuration aspects or attributes, such as (i) Autonomy, in terms of work organization, fit for existing strategies and resource control (ii) Issuing, meaning organization wide sensitization and encouragement for radical innovations through corporate culture, clear mission, goals and innovation charter (iii) Sanctions, in terms of positive and negative consequences in case of success or failure and (iv) Formalization, meaning formal rules of activities, reporting and communication throughout the FEIP. As mentioned above, little is known about the influence of system variables on individual perception and on risk behaviour. Only a few studies empirically investigated the organizational antecedents of knowledge dissemination in the context of R&D, too [e.g. Bij et al. 2003, Song et al. 2006].

4.3 Individual Level of the Initiator

At this level the variables of interest are (i) Personal Characteristics like age, position, target level, domain familiarity and outcome history, whose influence on (ii) Risk Propensity was already shown in different studies [e.g. MacCrimmon/Wehrung 1990, Thaler/Johnson 1990, Mullins et al. 1999, Soane/Chmiel 2005]. Moreover, (iii) motivational characteristics such as self-efficacy [Bandura 1997], need of achievement [McClelland 1961], locus of control [Rotter 1966] and the stable trait proactive personality [Bateman/Crant 1993] are expected to play a key role in initiator's decision making because their high standing in entrepreneurship literature [e.g. Keh et al. 2000, Busenitz/Barney 1997, Palich/Bugby 1995]. The cognitive perspective is represented by (iv) Risk Perception, which is extended by two important refinements: first it would be built formatively, consisting of the possibly independent Initiator's estimations of potential, significance and probability of (positive and negative) consequences. Through this we then respect the multidimensionality of risk in the context of radical innovations which cannot be described through a single number [March/Shapira 1987]. Secondly, the two-level approach requires differing between Perceived Organizational Risk, concerning the risk for the whole firm, and Perceived Individual Risk, concerning the personal risk for the initiator involved [similar Dess/Lumpkin 2005]. The last construct on the individual level is (v) Risky Decision for initiating a radical project or remaining on the status quo.

4.4 Interrelations

We can only give a short overview of the basic supposed interrelations and illustrative examples conceptualized in this paper. First, the *Personal Characteristics* influence Risk Propensity. For example Thaler/Johnson [1990] showed that past outcomes influence the willingness to take risk (so called house-money effect). MacCrimmon/Wehrung [1990] found a positive relationship of Income and Position as well as a negative correlation of Age and Experience with Risk Propensity,

supporting the assumption above. Risk Propensity determines Risk Perception, because rising in risk aversion implies higher perceived risk for the same situation [e.g. Sitkin/Weingart 1995, Cho/Lee 2006]. In addition it is expected, that certain Personal Characteristics such as Decision Domain exert a direct influence on the Risk Perception [March/Shapira 1987] because of correlating with the perceived locus of control.

Within this **Risk Perception** construct, reflection effects are expected which means that higher perceived risk for the organization will automatically lead to the increase of Perceived Individual Risk. This is intuitively traceable: increasing organizational risk implies a higher chance of bankruptcy, which in turn leads to higher personal risk of salary decrease or even dismissal. The Decision Behaviour itself is partly determined by Risk Propensity [Forlani et al. 2002] and Risk Perception [Sitkin/Weingart 1995, Simon et al. 2000]. At constant Risk Propensity, higher Perceived Risk leads to retain the status quo. On the other hand at constant Perceived Risk, an increase in Risk Propensity raises the chance of taking charge of the initiative. That's why both are expected to affect the decision outcome.

Finally the **Organizational Variables** have to be integrated into the model. On the one hand it is expected that they influence Risk Perception in a direct manner. For example higher Formalization provides clear guidelines about "doing the things right" so that in the case of failure blame cannot so easily be assigned to the initiator, which reduces Perceived Individual Risk. The same is expected if the Sanctions are process based, so that the proficiency of the initiator's activities and not the outcome which is highly influenced by external factors as well. However, Issuing provides safety to "do the right things". If experimenting and "thinking out of the box" is of strong cultural value, the initiator will perceive lower individual risk because his engagement is not regarded as abnormal behaviour by the organization. On the other hand these system variables influence the personal characteristics, too. For example through Issuing the organization sets the personal Target resp. the Attention Level which is an aspiration or survival level [March/Shapira 1987 for individuals, Miller/Chen 2004 for organizations] and influences the motives of the decision maker, which consist in achievement or avoidance motives [Higgins 1997, Xie/Wang 2003, Williams/Narendran 1999]. Both influence Risk Propensity and cause risk seeking or risk aversion. Further a higher level of Autonomy may result in a higher perceived locus of control which also leads to higher Risk Propensity [e.g. Simon et al 2000, Keh et al. 2002]. Finally direct effects on the decision outcome are expected. High positive Sanctions in the case of success for example will lead to taking charge because even if the perceived risk is high it is countervailed by the expected returns. Empirical references could not be found for most relations concerning organizational variables so far.

In summary the influences of organizational and individual characteristics on the decision outcome are partially mediated by Risk Perception and Risk Propensity. The main relationships are summarized in figure 3.

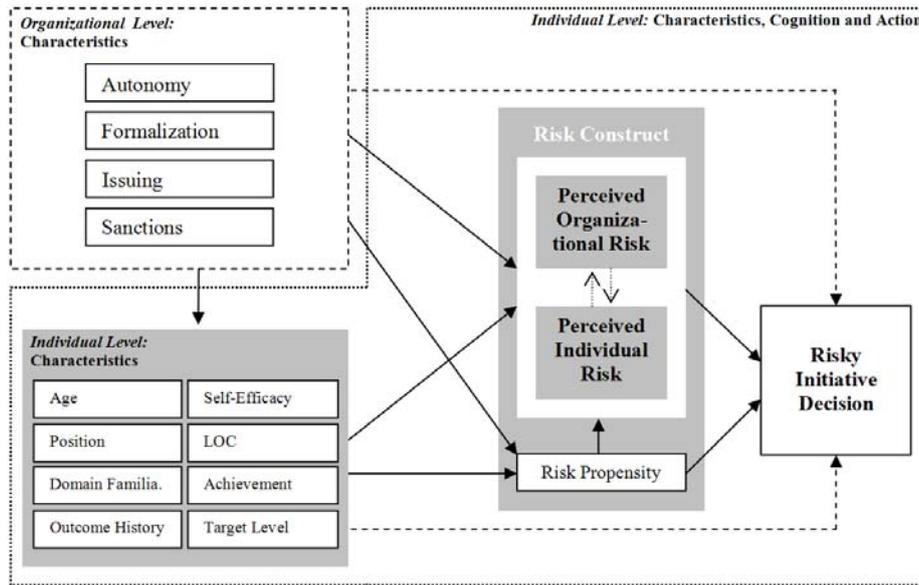


Figure 3: Basic Front End Decision Making Framework

5 Conclusion

A deeper understanding of the impact of the organizational system can assist managers with its configuration to improve bottom up initiatives for radical innovations from below. As opposed to incremental innovations, where suggestion systems or kaizen may institutionalize this, radical innovations are dependent from singular persons – and their knowledge about radical opportunities cannot be brought to the top management through their direct intervention. It needs a systemic approach to motivate the initiator for further information gathering, processing and diffusion inside the company.

The FEDM Framework is the first approach that tries to explain the decision behaviour of the initiator in this early phase of the innovation process. Here, the decision about high risk alternatives is dependent from individual characteristics as well as the organizational context. Both influence the cognition of the situation, which in turn determines the decision outcome. The presented framework helps to understand the determining factors of the initiators behaviour and therefore is of high practical relevance: First of all the FEIP is easily changed at low costs compared to later phases of the innovation process. Secondly, the top management gets to understand an innovation phase resp. decision which is not openly observable. Thirdly, the critical role of the middle management - often cut down in consequence of redesign and lean management - is highlighted. Fourthly, the management can deduce concrete organizational development activities from the model. Finally the decision setting deals with initiatives for radical innovations which are essential for a company's survival.

For the future this framework needs to be expanded and empirically examined to identify the organizational set screws of boosting radical initiative behaviour.

References

- [Antoncic/Hisrich 2001] Antoncic, B. / Hisrich, R.D. (2001): Intrapreneurship: Construct Refinement And Cross-Cultural Validation. In: *Journal of Business Venturing* 16, S.495-527
- [Bandura 1997] Bandura, A. (1997): *Self-efficacy: The Exercise of Control*. Freeman, NY
- [Bateman/Crant 1993] Bateman, T.S. / Crant, J.M. (1993). The Proactive Component of Organizational Behavior: A Measure and Correlates. In: *Journal of Organizational Behavior*, Vol.14 (2), S.103-118
- [Bij et al. 2003] Bij, X. / Song, X.M. / Weggeman, M.: An Empirical Investigation into the Antecedents of knowledge Dissemination at Strategic Business Unit Level. In: *Journal of Product Innovation Management*, Vol.20, S.163-179
- [Brauner/Becker 2006] Brauner, E. / Becker, A. (2006): Beyond Knowledge Sharing: The Management of Transactive Knowledge Systems. In: *Knowledge and Process Management*, Vol. 13, No. 1, S. 62-71
- [Brown/Eisenhardt, 1995] Brown, S.L. / Eisenhardt, K.M. (1995): Product Development: Past Research, Present Findings, And Future Directions. In: *Academy of Management Review*, Vol.20, No.2, S.343-378
- [Burgelman 1983] Burgelman, R. A. (1983): A Process Model of Internal Corporate Venturing in a Diversified Major Firm; In: *Administrative Science Quarterly* 28 (2), S.223-244
- [Cho/Lee 2006] Cho, J. / Lee, J. (2006): An integrating model of risk and risk-reducing strategies. In: *Journal of Business Research*. Vol. 59, S. 112-120
- [Cooper 1988] Cooper, R.G. (1988): Predevelopment Activities Determine New Product Success. In: *Industrial Marketing Management*, Vol.17, S.237-247
- [Cooper/Kleinschmidt 1995] Cooper, R. G. / Kleinschmidt, E. J. (1995): Benchmarking the Firm's Critical Success Factors in New Product Development; In: *Journal of Product Innovation Management*, Vol.12, S.374-391
- [Daneels/Kleinschmidt 2001] Daneels, E. / Kleinschmidt, E.J. (2001): Product innovativeness from the firm's perspective: Its dimensions and their relation with project selection and performance. In: *Journal of Product Innovation Management*, Vol.18, S.357-373
- [Dess/Lumpkin 2005] Dess, G.G. / Lumpkin, G.T. (2005): The Role of Entrepreneurship Orientation in Stimulating Effective Corporate Entrepreneurship. In: *Academy of Management Executives*, Vol.19 (1), S.147-156
- [Dutton/Ashford 1993] Dutton, J.E. / Ashford, S.J. (1993): Selling Issues to Top Management. In: *Academy of Management Review*, Vol.18 (3) S.397-428
- [Ernst 2002] Ernst, H. (2002): Success factors of new product development: a review of empirical literature. In: *International Journal of Management Reviews*, Vol.4 (1) S.1-40
- [Floyd/Wooldridge 1997] Floyd, S.W. / Wooldridge, B. (1997): Middle Management's Strategic Influence and Organizational Performance. In: *Journal of Management Studies* Vol.34 (3), S.465-485

- [Forlani et al. 2002] Folani, D. / Mullins, J.W. / Walker, O.C. (2002): New Product Decision Making: How Chance and Size of Loss Influence What Marketing Managers See and Do. In: *Psychology & Marketing*, Vol.19 (11), S.957-981
- [Frese et al. 1997] Frese, M. / Fay, D. / Hildeburger, T. / Leng, K. / Tag, A. (1997): The concept of personal initiative: Operationalization, reliability and validity in two German samples. In: *Journal of Occupational and Organizational Psychology*, Vol.70, S.139-161
- [Frohmann 1999] Frohmann, A.L. (1999): Personal Initiative Sparks Innovation. In: *Research Technology Management*, May-June, S.32-38
- [Gebert 2002] Gebert, D. (2002): *Führung und Innovation*. Kohlhammer, Stuttgart
- [Hauschildt 2004] Hauschildt, J. (2004): *Innovationsmanagement*; 3. Aufl., München
- [Henard/Szymanski 2001] Henard, D.H. / Szymanski, D.M. (2001): Why Some New Products Are More Successful Than Others. In: *Journal of Marketing Research*, Vol.38, S.362-375
- [Hornsby et al. 2002] Hornsby, J.S./ Kuratko, D.F. / Zahra, S.A. (2002): Middle managers' perception of the internal environment for corporate entrepreneurship: assessing a measurement scale. In: *Journal of Business Venturing*, Vol.17, S.253-273
- [Howell/Higgins 1990] Howell, J.M. / Higgins, C.A. (1990): Champions of Technological Innovation. In: *Administrative Science Quarterly*, Vol.35, S.317-341
- [Kahneman/Tversky 1979] Kahneman, D. / Tversky, A. (1979): Prospect theory: An analysis of decision under risk. In: *Econometrica* 47(4), S.263-291
- [Khurana/Rosenthal 1998] Khurana, A. / Rosenthal, S.R. (1998): Towards Holistic Front Ends in New Product Development. In: *Journal of Product Innovation Management*, Vol.15, S.57-74
- [Kim/Wilemon 2002] Kim, J./Wilemon, D. (2002): Focusing the fuzzy front-end in New Product Development. In: *R&D Management*, Vol.32 (4), S.269-279
- [Koen et al. 2001] Koen, P. / Ajamian, G. / Burkart, R. / Clamen, A. / Davidson, J. / Dámore, R. / Elkins, C. / Herald, K. / Incorvia, M. / Johnson, A. / Karol, R. / Seibert, R. / Slavejkov, A. / Wagner, K. (2001): Providing Clarity and a Common Language to the "Fuzzy Front End". In: *Research Technology Management*, March/April S.46-54
- [Leifer et al. 2000] Leifer, R. / McDermott, C.M. / O'Connor, G.C. / Peters, L.S. / Rice, M.P. / Veryzer, R.W. (2000): *Radical Innovation – How Mature Companies Can Outsmart Upstarts*. HBS Press, Boston/Massachusetts
- [MacCrimmon/Wehrung 1990] MacCrimmon, K.R. / Wehrung, D.A. (1990): Characteristics of Risk Taking Executives. In: *Management Science*, Vol.36 (4), S. S.422-435
- [March/Shapira 1987] March, J. / Shapira, Z. (1987): Managerial Perspectives on Risk and Risk Taking. In: *Management Science*, Vol.33 (11), S.1404-1418
- [March/Simon 1958]: March, J.G. / Simon, H.A. (1958): *Organizations*. John Wiley, NY/London/Sidney
- [McClelland 1961] McClelland, D.C. (1961): *The Achieving Society*. Van Nostrand, NY
- [Mensel 2004] Mensel, N. (2004): *Organisierte Initiativen für Innovationen*. DUV, Kiel 2004
- [Montoya-Weiss/Calantone 1994] Montoya-Weiss, M.M. / Calantone, R. (1994): Determinants of New Product Performance: A Review and Meta-Analysis. In: *Journal of Product Innovation Management*, Vol.11, S.397-417

- [O'Connor/Rice 2001] O'Connor, G.C. / Rice, M.P. (2001): Opportunity Recognition and Breakthrough Innovation in Large Established Firms. In: *California Management Review*, vol.43, No.2, S.95-116
- [Palich/Bagby 1995] Palich, L.E. / Bagby, D.R. (1995): Using Cognitive Theory to Explain Entrepreneurial Risk-taking: Challenging Conventional Wisdom. In: *Journal of Business Venturing*, Vol.10, S.425-438
- [Pralhad 1998] Prahalad, C.K. (1998): Managing discontinuities: The emerging challenges. In: *Research Technology Management*, Vol.41 (3), S.14-22
- [Quinn 1985] Quinn, J.B. (1985): Managing Innovation: Controlled Chaos. In *Harvard Business Review*, May-June S.73-84
- [Rice 1999] Rice, M.P. (1999): Starting the Process – Managing Breakthrough Innovation. In: *Chemtech*, Feb., S.8-12
- [Rotter 1966] Rotter, J.B. (1966): Generalized Expectancies for Internal and External Control of Reinforcement. In: *Psychological Monographs: General and Applied*, Vol.80 (1), S.1-28
- [Simon et al. 2000] Simon, M. / Houghton, S.M. / Aquino, K. (2000): Cognitive Biases, Risk Perception, and Venture Formation: How Individuals Decide to Start a Company. In: *Journal of Business Venturing*, Vol.15, S.113-134
- [Sitkin/Pablo 1992] Sitkin, S.B. / Pablo, A.L. (1992): Reconceptualizing the Determinants of Risk Behavior. In: *Academy of Management Review*, Vol.17, No.1, S.9-38
- [Song/Montoya-Weiss 1998] Song, X.M. / Montoya-Weiss, M.M. (1998): Critical Development Activities for Really New vs. Incremental Products. In: *Journal of Product Innovation Management*, Vol.15, S.124-135
- [Thaler/Johnson 1990] Thaler, R.H. / Johnson, E.J. (1990): Gambling With the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice. In: *Management Science*, Vol.36 (6), S.643-660
- [Veryzer 1998] Veryzer, R.W. (1998): Discontinuous Innovation and the New Product Development Process. In: *Journal of Product Innovation Management*, Vol.15, S.304-321
- [Weber/Milliman 1997] Weber, E.U. / Milliman, R.A. (1997): Perceived Risk Attitudes: Relating Risk Perception to Risky Choice. In: *Management Science*, Vol.23 (2), S.123-144
- [Yates/Stone 1992] Yates, J.F. / Stone, E.R. (1992): The risk construct. In: Yates, J.F. (ed.) (1992): *Risk Taking Behavior*. West Sussex, England, Wiley&Son
- [Zahra et al. 2004] Zahra, S.A. / Hayton, J.C. / Salvato, C. (2004): Entrepreneurship in Family vs. Non-Family Firms: A Resource Based Analysis of the Effect of Organizational Culture. In: *Entrepreneurship Theory and Practice*, Summer, S.363-381
- [Zhang/Doll 2001] Zhang, Q. / Doll, W.J. (2001): The fuzzy front end and success of new product development- a causal model. In: *European Journal of Innovation Management*, Vol.4 (2), S.95-112

A Dynamic Factorial Model of Knowledge Exchange and Sharing – An Exploratory Perspective

Hans Lehmann

(School of Information Management, Victoria University of Wellington
P.O.Box 600, Wellington, New Zealand
hans.lehmann@vuw.ac.nz)

Franz Lehner

(Lehrstuhl für Wirtschaftsinformatik I, Universität Passau
D-93053 Passau, Germany
franz.lehner@wiwi.uni-passau.de)

Abstract: There is surprisingly little literature specifically concerned with theorising and conceptualising of the transfer and sharing of complex information and/or knowledge, despite the fact that its significance is widely and without restriction acknowledged throughout the [mostly Anglo-American] literature on knowledge management and organisational learning. It is the aim of this paper to provide a brief review that allows an eclectic snapshot of the state of the - predominantly continental European - literature on this subject. After an introductory definition and limitation of the concepts involved the theoretical constructs are illustrated with the use of a set of models – selected predominantly for their link to empirical research and the capability to delimit the field. The empirical grounding of the models makes it possible to view them as partial investigations contributing individual elements of a more overarching research framework into which future studies may be integrated. In conclusion, a systemic approach of knowledge exchange in the form of a dynamic factorial model is proposed, the contributing frameworks are further categorised as to the type of knowledge for which they would be of maximum utility and the influence of external issue and problem spaces is shown.

Keywords: Knowledge Transfer, Knowledge Sharing, Information Sharing, Knowledge Exchange Models.

Categories: I.2.4, H.4.m

1 Introduction

“Progress is achieved by the exchange of knowledge” [Becker, 95]. The changes of the last decade or so in business as well as in society have brought with them a much accelerated availability [and volume] of information. To turn this into increased knowledge requires both effective communication processes and a continuing will to learn on behalf of individuals as well organisations. The transformation of the industrial society to one of information and knowledge work points to far reaching socio-cultural change. Concepts such as Total Quality Management and Business Process Re-engineering during the decade of the 90s were mainly concerned to optimise ‘hard facts’. The observed lack of unqualified success and often distinctly sub-optimal results [Gertz, 98] highlighted the fact that ‘soft facts’, i.e. factors dependent on the culture of the enterprise such as Change Management, Team

Building and methods of Organisational Learning [Neumann, 98] also possessed some hidden potential. From the end of the 90s this was accepted wisdom and from then on the hunt for diversification factors that were at once hard to imitate and competitively effective concentrated on the ideas, knowledge and creativity inside the head of an enterprise's people.

Most applications of knowledge management seem to deal with explicit knowledge, such as expert have, which is defined by its functional use, i.e. by what tasks it can help complete or which problem it can assist in solving. The accent is firmly on an instrumental, in the wider sense 'technical', level. This implies that knowledge can be made to exist independent of persons and contexts. In this way – and only in this way – can it be stored, used, extended, multiplied and distributed by automated technology. Implied, or tacit knowledge, however, is equally important. First pointed out by [Polanyi, 67], this includes understanding of interactions and relationships hidden between other elements of knowledge. This is the area where technology for communicating networked knowledge management is increasing in importance.

Thus the knowledge in an organisation exists in a large variety of manifestations, from 'hard facts' to 'softer' representations such as general *Weltanschauung*, specific visions, myths, symbols and behaviour patterns. All can be either explicitly documented or exist 'tacitly' in individual's hunches, thoughts and memories. Applying a more narrow field of definition – in line with the fundamental tenor of this review – then *information sharing* is the distribution of information about (a) specific object set[s], usually from (b) common information source(s), for (c) the common use of individuals. The inclusion of multiple sources of information transcends information sharing to *information exchange*.. Information is furthermore the *sine qua non* element of communication. This 'Siamese-Twin' relationship leads to a logical equivalence of the form: *Communication = Information Exchange = Information Sharing*. Knowledge Exchange is an even more complex process than Information Sharing and usually includes the latter as subordinated processes. [Brödner, 99] introduce the concept of *division of knowledge* "a social interaction process that effectively generates, and at the same time utilises, new knowledge. {Knowledge Division} encompasses the specialisation and fragmentation of knowledge at its genesis as well as the diffusion through collaboration in its {new} use". Drawing parallels with the industrial society's key principle of the division of labour, Brödner sees the division of knowledge playing a similar role in the productivity of knowledge industries in the information age. Although this review will not discuss the concept of knowledge division any further, it is a vivid illustration of the close connection between knowledge management, knowledge identification and classification and the generation, use and sharing of knowledge.

2 Selected Knowledge Exchange Models

Models of the process by which exchanges of knowledge take place have played an important role in the building of explanatory propositions in this field. The models selected for discussion were all required to have a sound empirical base. The selection does not aim for completeness, but rather at demonstrating a delimiting spectrum of the phenomenon. They are:

- Boeglin's Know-How transfer model;
- Szulanski's stepwise model of Best Practices Transfer;
- Richter's Transfer Potential Absorption model.
- Zander & Kogut's Transfer and Imitation model;

In this section they are discussed and related in more detail.

2.1 Boeglin's Model of Know-How Transfer

For [Boeglin, 92] transfer of know-how is a major competitive potential and an important synergy source for internal re-structuring, rapid growth and increased performance of decentralised organisations. Know-how transfer, however, always depends on the compatibility of the sender's and receiver's attitude and capabilities and incompatibilities result in diverse problem/action types.

The Leadership Problems result from the lack of willingness to participate in a knowledge transfer transaction, whereas Communications Problems on both the senders' and receivers' sides can be the result of communication channel problems, but more often have their causes in language difficulties, culture or context considerations and divergent experiences. Management of the problem/action process complex requires a balance between guidance and control as well as acknowledging and influencing the context for the process.

2.2 Szulanski's Step-Model of 'Best-Practice Transfer'

Transferring successful ways of operating to other parts of the enterprise is the focus of 'Best-Practice Transfer'. This type of knowledge exchange often also includes strong elements of *tacit* knowledge, which are embedded in personal skills and previous experiences of collaborations and their social context. The model was developed by assimilating a number of research results from studies in the fields of innovation diffusion, social change, systems implementations and technology transfer [Szulanski, 96]. In the *Initiation* phase the decision is taken whether or not to transfer. *Installation* is the first actual transfer step, where resources between the first sender/receiver pair[s] are established and actual knowledge is transferred in the planned manner. The outcomes of these first attempts are then used to update the plan and to fine-tune the procedure accordingly. The focus during the *Ramp-up* phase is the facilitation of smooth provision of the right knowledge, the willing acceptance of the transfer by the receiver[s] and reaching [or otherwise] the required level of performance improvement. This will necessitate further refinements in the process, mostly to cope with unexpected occurrences or with under- or overestimated organisational or cultural factors. Once an acceptable level of improvement can be assured, the *Integration* of the transferred knowledge and any process modifications/enhancements resulting from it need to be undertaken. This aims to institutionalise the new procedural context as 'routine' work, i.e. making it stable and predictable.

2.3 Richter's Absorption Potential Model

The model was built following a study investigating the learning behaviour of German heavy engineering firms in Japan. It focused mainly on the transfer of cultural knowledge with a view to decide to what extent Japanese management

philosophies and organisational and enterprise culture such as decision behaviour and customer orientation could be integrated into their German head office. The main finding was that the key factor for learning behaviour and thus for successful knowledge transfer was the nature and quality of the personal relationships between the centre and the subsidiary's offices [Richter, 95]. Analysing such personal relationships by themselves, however, is not enough to establish conclusions about the effect this learning behaviour may have on the quality of the resulting transfer of cultural knowledge. There are further influences that shape such a transfer – and also shape each other in a cyclical process [Krogh, 98].

The transfer of knowledge between the centre and the subsidiaries is classified by [Richter, 95] as inter-enterprise organisational learning. Organisational learning can be separated into two groups, namely single-loop and double-loop learning {[Richter, 95], [Lehner, 00], [Pawlowsky, 98]}. Single-loop [or adaptive] learning is essentially a comparison of actual behaviour with a pre-determined norm and behaviour is adapted ['regulatory reflexes'] to counteract any deviations from the norm [Pawlowsky, 98]. For an organisation, this means the sum of all reactive measures in response to perceived shortcomings against the pre-established norms and expectations.

These reactive measures, however, require additional operational and cultural knowledge to create an adaptive potential. If this potential not only leads to behavioural adaptation, but also affects and modifies the target norm and expectations – which the learning process is based on – then double-loop [context sensitive] learning is occurring. Single-loop learning is an imperative precursor to double loop learning.

The study further showed a multifaceted picture for the transfer as well as implementation power factors. This was overshadowed by a limited intention of the centre to learn from its Japanese subsidiary. Similarly, the investment into mediation and absorption resources turned out to be very uneven. With a lack of management effort and little, if any time set aside for the transfer, the subsidiary did not succeed in transferring as much as they expected and soon lost impetus as the centre seemed to ignore their contributions rather than encourage them further – mainly for a similar lack of resources there. Communications and interpretation capabilities, on the other hand, were of a high standard – but on their own they could not achieve a successful transfer of cultural knowledge [Richter, 95].

2.4 The Transfer and Imitation Model of Zander and Kogut

The basic hypothesis underpinning Zander and Kogut's model [Zander, 95] is that internal transfer [defined especially for manufacturing technology across business units of the same enterprise] and imitation [the capability to copy competitors' products] are identical phenomena of organisational capabilities and diffusion processes, determined by common – and not specific, i.e. different – factors. Following the work of [Winter, 87] and [Rogers, 80], Zander and Kogut recognise five main influence elements, namely Codifiability, Complexity, Teachability, Systems Dependence, Parallel Development. The process of transfer is furthermore expected to be determined by the nature of the underlying base knowledge and the ease with which this knowledge can be copied or re-constructed. A further influence is the degree to which the participating firms use common or distinctly different manufacturing facilities. This is especially of importance when the transfer is not

internal, but an imitation between competitors. To analyse this aspect further, they break this commonality of facilities down into three more criteria:

- The ratio between Proprietary vs. Outsourced Equipment is an indication of the degree of original development and internal know-how;
- Key Employee Turnover assesses how many key executives and managers in the area affected by/targeted for an imitation have left their positions – the loss of their knowledge may thus have left the incumbent firm more vulnerable;
- Continuous Development indicates an implicit ability to carry out product changes/improvements at short notice, e.g. to fight off a product imitation.

The hypotheses about the effect the factors had on the speed of transfer and the ease of imitation were tested in a study of 35 innovations by firms in Sweden. It showed that the degree to which knowledge is already gathered and in explicit and documented form or can easily be ‘codified’ is as important for the speed of internal transfer as is the capability to ‘teach’ effectively any new knowledge required. Competitive pressure through ‘parallel development’ by other firms is another strong indicator of transfer speed. Systems dependence and the underlying complexity of the knowledge area were, however, not of any significant influence on internal transfer. Neither were they of import for imitation of another product. In fact, none of the factors influencing enterprise internal knowledge transfer affected external imitation at all. It was only the three specific factors that influenced this form of involuntary knowledge transfer. Loosing key employees depresses the capacity to imitate, whereas continuous development protects the incumbent firm from external imitation attempts. Proprietary equipment, i.e. internally developed technology is more resilient to imitation, however, the opposite policy of outsourcing does not significantly weaken a firm’s position.

On the strength of these findings, the original hypothesis that the internal transfer of technology capability and external imitation both work along the same mechanism did not hold up. A possible reason is that successful imitation of innovations relies on a number of other factors, outside manufacturing, such as marketing and other value chain management. Indirect factors, such as the imitating firm’s reputation and the legal environment it operates in are still further factors.

3 Factors Supporting or Inhibiting Information Sharing and Knowledge Exchange

There is widespread consensus about the importance of knowledge and about the fact that information sharing and knowledge exchange are central processes for the optimal use and distribution of the knowledge resource. The theory in this field goes to some depth about the different types and categories of knowledge and what distinct distribution corollaries result from them. Applied research literature, aimed mostly at the practitioner, focuses more on the barriers to successful knowledge transfer and the dependency of these processes on their specific, often individual context.

Another focal point for the literature is the degree of difficulty with the implementation of information sharing and knowledge transfer processes in different fields, often coupled with analyses of the causes for the failure to actualise the theoretical insights when it comes to practical application. Part of this difficulty is

that the frameworks and models discussed all come from different perspectives of knowledge and from different environments in which the transfer and sharing of such knowledge would be anchored. It may therefore be assumed that not every such model and framework is equally well suited to all situations of knowledge transfer. In the following, the models are categorised as to their usefulness for transferring and sharing the different types of knowledge encountered in the practical applications and contexts of knowledge management.

To find a categorisation scheme for different knowledge types, the [Zander, 95] framework of ‘influence elements’ seems the most useful. However, for the purposes of this discussion, only the factors affecting “Internal Transfer” are of relevance. Of these, furthermore, *Parallel Development*, as an external influence, should be excluded from a model of the transfer process *per se*. The remaining four factors are closely related to each other – in fact only two of the ‘elements’ are not directly related to each other, as Figure 1. shows.

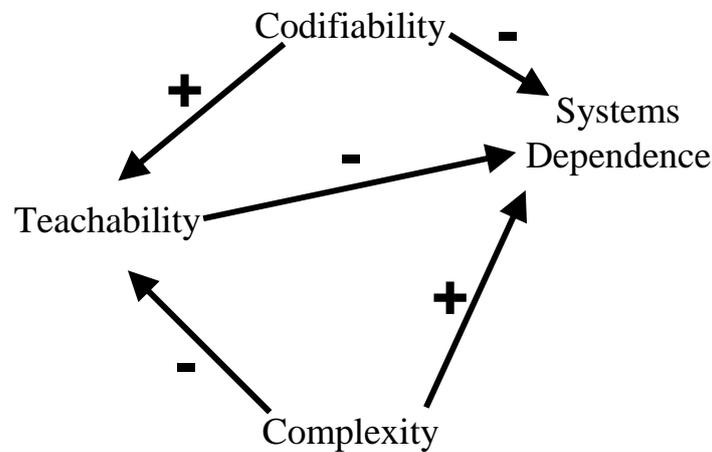


Figure 1: [Co]relationships among Zander & Kogut’s [1995] ‘elements’ that ‘influence’ knowledge transfer and imitation.

Notation: $A \rightarrow + \rightarrow B$: the more of A, the more of B; $A \rightarrow - \rightarrow B$: the more of A the less of B

The relationships between the factors involved in the dynamics are here represented as “cause-effect-loops” introduced by [Weick, 79] Two types of effect are recognised: a change in the *same* direction, i.e. “the more of A, the **more** of B” or “the less of A, the **less** of B”; this is labelled with a ‘plus’ [+] sign. A minus [-] sign signals that the change will move into the *opposite* direction, so that “the more of A, the **less** of B” or the “less of A, the **more** of B” [*ibid.* p71]. Such “causal networks” are considered particularly useful because they “respect [the] complexity” of cyclical interactions where causation is not unilateral and where cause and effect can be interchangeable.

Teachability and Systems Dependence both are correlated to Codifiability: the higher the Codifiability, the higher the Teachability and the less Systems

Dependence. Conversely, the higher the Complexity of the underlying knowledge, the more Systems Dependence, the less Teachability there is. Complexity and Codifiability, however, are not correlated as the effects of any other factor – on the contrary, they themselves are multiple causations for the other factors. Complexity and ‘Codifiability’ [here interpreted as the extent to which the knowledge in question is tacit or explicit] can then be regarded as the bounding dimensions for classifying the type of knowledge to be transferred or shared.

[Boeglin’s, 92] model, with its somewhat mechanical dynamics, is thus particularly useful for explicit knowledge of low complexity, where more sophisticated methods would most likely constitute and overkill. At the other extreme, for achieving the efficient transfer and effective sharing of more – or highly - complex knowledge, especially in explicit manifestation, Richter’s, 95] model seems more appropriate. This two-station model is characterised by a broad and deep structure of determinants which are engaged in an iterative cycle to enable ‘double-loop’ learning.

Where knowledge is mostly tacit, i.e. difficult to codify, but of limited complexity, Szulanski’s [Szulanski, 96] stepwise progression model suffices and is very useful. In contrast, Zander & Kogut’s [Zander, 95] model can cope with high levels of complexity, and its processual nature and pre-defined factor structure means that the knowledge to be transferred may be at different levels of codification. Furthermore, the inclusion of the ‘teachability’ factor means that if a suitable communications structure and process configuration can be established than this model is of specific use for the exchange of tacit knowledge.

4 Conclusion

In conclusion, information sharing and knowledge transfer cannot be considered in isolation from the context that individuals, groups and organisational units find themselves when they participate in the transfer. The behaviour of the actors in this process depends on their individual and collective experiences, entry level of knowledge, their will to learn and the emotions towards other participants, determined by such factors as power seeking/preserving and (dis)trust. Groups are influenced by their collective behavioural characteristics, which themselves are dependent on culture and interaction potential, in turn governed by motivational factors and an open or closed ‘climate’. These factors are reflected in all aspects and phases of information sharing and knowledge transfer. They are all in correlation to one another, although the nature and depth of their relationship often differs considerably.

This underlying complexity is in stark contrast with the relative simplicity of the models introduced in this review and paucity of their predictive power. It is therefore difficult not to agree with [Krogh, 998] that, given the current level of our understanding of knowledge exchange at the phenomenological level, we are some way away from being able to govern or manage information sharing and knowledge transfer processes to a level approaching effective control. A much deeper degree of *insight* in the epistemological sense – to which the models, however, can and do contribute – seems to be necessary before we can begin to think of ours as a knowledge management culture.

References

- [Becker, 95] Becker, J.. Den Austausch von Wissen fördern, in: Gabler Magazin 3/1995, 16-19
- Boeglin, 92] Boeglin, P. Innerbetrieblicher Know-how-Transfer, in: iO Management Zeitschrift, Nr. 9, 86-91, 1992
- [Brödner, 99] Brödner, P., Helmstädter, E., Widmaier, B. [Eds.]. Wissensteilung – Zur Dynamik von Innovation und kollektivem Lernen, München, Schriftenreihe des Instituts Arbeit und Technik im Wissenschaftszentrum Nordrhein-Westfalen, Band 13, 1999
- [Gertz, 98] Gertz, W. Fadenkreuz: Das Wissen der Mitarbeiter, in: CW Nr. 9 vom 27.02.1998, S. 87-88,
<http://www.computerwoche.de/infopoint/heftarchiv/drucken.cfm?id=39265&cfid=735249&cftoken=17814334>
- [Krogh, 98] Krogh von, G., Köhne, M.. Der Wissenstransfer in Unternehmen: Phasen des Wissenstransfers und wichtige Einflussfaktoren, in: Die Unternehmung, 52 Jg., Heft 5/6, 1998, 235-252
- [Lehner, 00] Lehner, F.. Organisational Memory - Konzepte und Systeme für das organisatorische Lernen und das Wissensmanagement, München/Wien 2000
- [Neumann, 98] Neumann, S., Flüge, B., Finerty, T. The Art of Knowledge - Potential aus dem Wissen schöpfen, in: Information Management 1/1998, 66-74
- [Pawlowsky, 98] Pawlowsky, P.. Wissensmanagement, Erfahrungen und Perspektiven, Wiesbaden 1998
- [Polanyi, 67] Polanyi, M.. The Tacit Dimension. New York, 1967
- [Richter, 95] Richter, F. J. Transfer von Kenntnissen und Erfahrungen zwischen Zentrale und Auslandsniederlassung. In: Zeitschrift für Planung [zfP], 6/1995, 227-240
- [Rogers, 80] Rogers, E.. Diffusion of Innovations, New York: Free Press 1980
- [Szulanski, 96] Szulanski, G.. Exploring Internal Stickiness: Impediments to the Transfer of Best Practice within the Firm, in: Strategic Management Journal, Vol. 17 [Winter Special Issue] 1996, 27-43
- [Weick, 79] Weick, K.E.,. The social psychology of organising. Reading, Management, Addison Wesley, 1979
- [Winter, 87] Winter, S. [Winter, 87]. Knowledge and Competence as Strategic Assets, in: Teece, D. [Ed.]: The Competitive Challenge – Strategies for Industrial Innovation and Renewal, Cambridge, 1987
- [Zander, 95] Zander, U., Kogut, B. Knowledge and the Speed of the Transfer and Imitation of Organizational Capabilities: An Empirical Test, in: Organization Science, Vol. 6, No. 1, January-February 1995, 76-92

Practices of Knowledge Management in Companies: A Turkey Survey

Ferit Ölçer

(Mustafa Kemal University, Turkey
folcer@hotmail.com)

Abstract: In today's digital economy, knowledge management is critical to the success of many organizations. This paper is based on a research which examined the practices of knowledge management in Turkish large companies. The research sample was 214 Turkish organizations. Data were obtained by questionnaires and analyzed by SPSS. The main findings revealed that the majority of respondents believed that their knowledge management system was very important to the development of their organizations, and that the most important sources of ideas come from customers. The most important method used by Turkish organizations to facilitate the sharing of knowledge between employees was internet. This research produces some useful insights and leaving a number of issues for future research.

Keywords: knowledge, knowledge management, information

Category: Category A - General Literature

1 Introduction

Interest in knowledge management (KM) has grown with the sophistication of technology for sharing knowledge [Clarke and Rollo 2001; Lim and Klobas 2000; Balla et al. 1999; Demerest 1997]. Extensive literature and case studies support the belief that knowledge management plays a key role in managing businesses successfully. Thus, the ability to learn, acquire, foster and integrate relevant knowledge within an organization and its value chain is recognized as one of the most important competences for a firm to be successful [Liao 2003]. There has been a great deal of debate in the literature about the meaning of the term "knowledge management". While KM has been widely studied by academics, it has no universally accepted definition.

KM is called many different things, and every company has its own distinctive approach, but connecting all of these initiatives is a commitment by companies to developing the production and flow of knowledge, the transmission and use of knowledge to create economic value. KM is the study of strategy, process and technology to acquire, select, organize, share and leverage business-critical information [Zhang and Zhao 2006]. Scarbrough et al. [1999] define knowledge management as follows: "KM covers any intentional and systematic process or practice of acquiring, capturing, sharing and using productive knowledge, wherever it resides, to enhance learning and performance in organizations." KM needs to be regarded differently from simple information gathering in order to take advantage of its competitive potential [Takahashi and Vandenbrink 2004; Zhang and Zhao 2006].

According to Huosong et al. [2003], KM consists of management activities which develop and utilize an organization's knowledge resources efficiently and improve a firm's creative ability. In addition, despite the academic research and organizational practices developed around this concept, according to several authors [Huysman and de Wit 2002; Bhatt 2001; Neef 1999], a considerable gap in the literature remains. Huysman and deWit [2002] suggested that the majority of the empirical studies do not analyze the complete knowledge cycle and that KM has been treated in different views. Davenport and Prusak [1998] suggest that KM can be achieved by effectively utilising the existing resources within an organization. A unique characteristic of their work is acknowledgement that much knowledge is driven by recognition of a need or demand for knowledge. They note that successful knowledge transfer relies on a culture which supports knowledge sharing and includes trust among members of the organization. Nonaka and Takeuchi [1995] established a dynamic model of knowledge creation, with the key assumption that knowledge is created and expanded through social interaction or knowledge conversion between tacit knowledge and explicit knowledge.

Sanchez [2004] identifies the differences between two fundamental approaches to KM: the tacit knowledge approach and the explicit knowledge approach. Sanchez suggests that both approaches have important strengths and weaknesses, and that it is important to find the right creative combination of the two approaches for KM. Marshall [2000] advises organizations that their most important KM task is to take care to retain existing employees. Awad and Ghaziri [2004] after reviewing the existent literature in KM and considering that KM goes through an ongoing life cycle proposed a four step model: capturing, organizing, refining, and transferring knowledge.

Schwartz et al. [2000] presented the acquisition, organization, and distribution model, in which the knowledge cycle is described in three phases: acquisition, organization and distribution of information. Tiwana [2001] suggested a similar model that considers, instead of the phase of organization, the internal and external application of information. According to Stojanovic et al. [2002] the main function of a KM system is to capture and disseminate new sources of information. On the other hand, researchers seek to develop a strategy for capturing, creating, using and transferring knowledge effectively across the organization in order to improve efficiency and provide sustainable competitive advantage [Geisler 2007; Moffett 2003; Muthusamy and Palanisamy 2004].

The organizational and managerial implementation and applications of knowledge in organizations have also received attention from researchers [Gupta et al. 2005]. According to the research of Beccerra-Fernandez et al. [2004], around 81 percent of the large companies in the USA and Europe are using some form of KM. Effective KM pays off in fewer mistakes, fewer redundancies, quicker problem-solving, better decision-making, reduced research development costs, increased worker independence, enhanced customer relations, and improved service. KM has helped organizations identify, select, organize, disseminate, reuse and transfer important information and expertise. These are necessary for problem-solving, dynamic learning, strategic planning, and decision-making [Gupta et al. 2005; Singh et al. 2006].

2 The Purpose of the Research and the Methodology

The purpose of this research is to investigate KM practices and their impacts on effectiveness and performance of the employees and the organizations in large scale manufacturing companies in Turkey. The selected populations for this research are The 500 Biggest Industrial Companies of Turkey, which is determined by the İstanbul Chamber of Industry. A sample size of 214 large scale companies in Turkey was randomly drawn from this list. Some of the unexplored sectors are important from the KM point of view but are beyond the scope of the present study. This is mainly due to time and cost constraints involved in undertaking the whole set of sectors operating in Turkey. In this research, data were collected through a survey questionnaire. The research instrument is based on the work of Singh et al. [2006], Al-Athari and Zairi [2001] and McAdam and McCreedy [1999]. The questionnaire formed two parts. The first part referred to the collection of the information of the respondents and the companies. The second part was designed to collect information about the practices of KM in surveyed companies. A comprehensive critical review of the literature enabled the key areas of interest in KM to be identified. These areas, namely knowledge construction, knowledge embodiment, knowledge dissemination and use, were then used to construct the second part of the questionnaire. The key areas of interest in KM investigated in this research are elements of KM; importance of knowledge; how employees view knowledge; knowledge capture; sources of knowledge; knowledge sharing; employees' input to KM system; effects and benefits of KM; difficulties in the implementation of KM. Initially, the study managers were presented with several statements to assess the practices of KM in Turkish companies. Managers were requested to show how strongly they agreed with these statements on a five-point Likert scale. A total of 214 questionnaires were mailed to different companies and 145 returned, which gives an effective response rate of 68 percent. The survey data was analysed by using SPSS (version 12.0).

3 The Findings

The sample was made up of 110 males, (75.9 percent), and 35 females (24.1 percent). The mean age of sample was 38.2 years and 84.5 per cent of the sample surveyed completed their education at the university. The major portion of the respondents is from the textile sector (26.5 percent), followed by food (23.4 percent), energy (13.8 percent), automobile (11.7 percent), electronics, telecommunication and IT hardware (9.6 percent), chemicals (7.5 percent), and iron and steel industry (7.5 percent).

3.1 Elements of Knowledge Management

Overall, the results show that most of the organizations surveyed consider knowledge to have both factual and social dimensions. For example, Table 1 shows that information technology, knowledge transfer, tools and methodologies, capture and dissemination, core competencies/capabilities and organizational knowledge are considered as key elements of knowledge and KM.

Options	%
Information technology	83
Knowledge transfer	81
Tools and methodologies	72
Knowledge capture and dissemination	69
Core competencies and capabilities	56
Organizational knowledge	52
Organizational learning	45
Organizational memory	34
Intellectual capital	29

Table 1: Terms denote knowledge in organizations

According to these findings, the respondents consider knowledge to be multi-faceted rather than being restricted to any given dimension and they are willing to accept the wider concept of knowledge. As McAdam and McCreedy [1999] stated, this is an essential starting point for the embodiment of such knowledge within organizations and the overall effective use of KM within organizations. Knowledge construction is considered a central feature to all aspects of KM. If managers adopt a narrow concept of knowledge at the knowledge construction stage, resultant KM systems are more likely to be mechanistic and ineffective. So, these findings represent considerable potential for translation into effective KM systems within organizations.

3.2 Importance of Knowledge to Organizations

The respondents were asked about how they view knowledge, or how important it is to them. Their answers reveal that the majority of organizations (64 percent) believe that knowledge is most important, while 36 percent of the organizations believe it is somewhat important. Accepting the importance of the knowledge, is not a significant finding, but collecting, processing, storing and taking action based on that knowledge, that is what certainly matters. The importance of knowledge can be evidenced by its association with actions, and its source can be found in a combination of information, social interaction and contextual situations, which affect the knowledge accumulation process at an individual level [Al-Athari and Zairi 2001].

3.3 How Employees View Knowledge

KM is about finding a way to capture the wisdom of the workers within an organization. That means sharing knowledge is an important factor in creating a KM system. In this context, the respondents were asked to clarify how their employees view knowledge. Their answers reveal that 72 percent of the employees see their knowledge coming from working with organizations and it is not person/private and 69 percent of the employees believe that their knowledge is as an organization asset (Tab. 2). But, 71 percent of the employees see their knowledge as power, while 62 percent of the employees see their knowledge as something private. In most environments, information is power. Holding vital information ensures that the employee will be seen as indispensable, and free from termination. In this type of environment, there is very little incentive to share information.

Options	Responses*
Employees knowledge comes from working with the organizations and it is not personel and/or private	72
Employees see their knowledge as power	71
Employees' knowledge is an organization asset	69
Employees see their knowledge as something private	62
Employees guard their knowledge to protect their positions	57
Employees think no one has the right to share their knowledge	52

* The percentage of respondents who mark "strongly agree" and "somewhat agree".

Table 2: How employees view knowledge

On the other hand, Zack [1999] argued that the most critical problem encountering KM is seeing knowledge as power. Schwarzwalder [1999] argued that the core to the idea of KM is establishing an environment where information is shared and openly accepted. The findings show that it may be difficult to have a KM system in Turkish companies with the above attitudes (Tab. 2).

3.4 Sources of Knowledge

The respondents were asked about the sources of their KM system. The majority companies use customer knowledge and others mistakes as a sources for their knowledge system (Tab. 3). These responses are seen in the light of the obvious trends in the market, wherein the customer is considered as the king and his feedback is valued the most. In addition, 85 and 83 percent of the companies, respectively, use their employees' performance and employees' knowledge, while 84 percent use organizational knowledge as a source for their knowledge system.

Knowledge Sources	%	Knowledge Sources	%
Customers	90	Team leaders	80
Others mistakes	88	Appraisals	79
Senior/middle management	87	Benchmarked companies	78
Lessons learned from projects	86	Daily experience	77
Competitors	86	Conferences/seminars/panel	74
Employee performance	85	Informal discussions	75
Knowledge about organizations	84	Journals/books	73
Employees	83	Inductions	68
Interviews	82	Formal discussions	67
Supplier	82		

Table 3: Sources of knowledge

These findings show that knowledge is considered to be systematically captured in Turkish companies and the captured knowledge comes from both internal and external sources, and KM must not be limited to one source, either internal or external. This finding is consistent with Coates [1999] who argued that KM should not be limited to internal sources.

3.5 Knowledge Sharing

The findings of the research show that a majority of respondents (92 percent) agree that their organizations facilitate knowledge sharing and dialogue within the organization. Furthermore, 78 percent of the respondents agreeing that employees take responsibility for their own learning and that dialogue and discussion are encouraged throughout the organization. The respondents were asked about the KM system which they use to capture, process, create, measure, and transfer their knowledge. Table 4 shows the results relating to how knowledge is shared or disseminated across the organization.

Options	%	Options	%
Internet	93	Multimedia presentations	75
Electronic mail	90	Workshops/forums/seminar	73
Face-to-face meetings for dialogue sharing	82	Creativity techniques	72
Mentoring and coaching	79	Discussion forums	67
Databases	77	Intranet	54
Cross-functional learning	76	Groupware	53
Training needs analysis	75		

Table 4: Methods to assist knowledge sharing

As can be seen from this table, the majority of the companies use internet and e-mail as a source for sharing their knowledge. Likewise, Singh et al. [2006] found that e-mail and internet have emerged as the extensively used enablers of KM in organizations. Al-Athari and Zahiri [2001] and Schwarzwalder [1999] found that the majority of the organizations use an internal journal as a source for sharing their knowledge. As Al-Athari and Zahiri [2001] stated, the important point in developing a KM system is the issue of process. To be effective, a knowledge system needs to operate in a manner that fits the work process flow in the organization. Information relevant to the onset of a project must be made available to staff in time and in an appropriate manner. According to the findings, technology, education and reward system are seen as the key important factors to encourage knowledge sharing.

3.6 Employees' Input to Knowledge System

The respondents were asked about the important knowledge input in the system. The major input in the companies' KM systems is employees' job experience. The findings shows that 81 percent of the companies use their knowledge system both to record employees' new ideas and information about customers and while 77 percent of them use it to record information about their suppliers. But, 72 percent of the companies use their knowledge system to record organization performance and 73 percent use it to record organization activity. Al-Athari and Zairi [2001] found that the major input in the government and private organizations' knowledge systems is to record organization performance and organization activity to a great extent.

3.7 Impacts of Knowledge Management

The perceived numerous impacts and benefits coming from the use of KM were also investigated. These are shown in Table 5 and are arranged in order of preference. As can be seen from this table, there are many benefits of KM.

Business Impacts/Benefits	%	Employee Impacts/Benefits	%
Improved quality	83	Employee productivity	83
Improved productivity/efficiency	81	Employee creativity	80
Being competitive	80	Teamwork	79
Improved management learning	80	Employee using abilities/experience	79
Launch of new product or services	79	Employee satisfaction	78
Improved effectiveness	79	Employee motivation	78
Customer satisfaction	77	Employee belonging	77
Time-to-markets improvement	75	Employee learning	76
Using knowledge for organization development	74	Changing employee behaviour	76
Profit margin	74	Increased power	75
Reduced operating costs	74	New career choices	72
Creating training system based on new knowledge	73	New job choices	70
Cycle-time reduction	73		
Increased sales	71		

Table 5: Impacts of using knowledge management in organizations

The most important benefits of KM are improving quality and employee productivity (consistent with the findings of McAdam and McCreedy 1999 and Singh et al. 2006), followed by, increased employee creativity, building an environment for teamwork, and increased employee satisfaction. On the other hand, the top-scoring items are seen to relate to improving both internal and external efficiency within the organizations as Table 5 shows. In summary, the findings showed that the respondents were able to grasp the breadth and depth of benefits, both business and employee, to be achieved by effective KM.

3.8 Knowledge Management Difficulties

Table 6 shows the various difficulties and obstacles the companies face with in the implementation of KM. The two main difficulties are identified as inability to utilize information and high staff defection. The other reported important obstacles in the success of KM in organization include the lack of contact among employees to share their knowledge, not knowing how to interpret information, insufficient employee performance and determining what knowledge should be managed.

Difficulties/Obstacles	%	Difficulties/Obstacles	%
Inability to utilize information	87	People do not disclose knowledge	67
High staff defection	85	Organization unaware of available information	66
Lack of contact among employees to share their knowledge	80	Lack of methodology and resources	65
Not knowing how to interpret information	79	Difficulties with sharing	64
Insufficient employee performance	76	Less use of information technology	64
Knowledge management concept not well understood	76	Insufficient knowledge about costumers	63
Determining what knowledge should be managed	73	Insufficient knowledge about competitors	59
Useful knowledge is ignored	72	Non-standardization of key organizational processes	56
Unable to access knowledge	71	Insufficient knowledge about suppliers	55
Lack of top management's commitment	71	Lack of organizational culture	54
Justifying the use of scarce resources for knowledge initiatives	70	Lack of time	52
Knowledge not available	69	Inter-department interactions are insufficient	49
Methods of dissemination	69	Lack of documentation	40
Poor incentive system	67		

Table 6: Obstacles to the success of knowledge management

4 Conclusions

The majority of the companies surveyed believed that KM was very important to the development of their organizations, and that the most important sources for their KM systems would be customer knowledge. It can be concluded from the research that understanding knowledge construction is fundamental to the overall concept of the KM process, in regard to theory and practice. The majority of employees see their knowledge coming from working with organizations and it is not personel/private and as an organization asset. However, some employees see their knowledge as power, and something private. Furthermore, they guard their knowledge to protect their positions. The most important method which is used by Turkish companies to facilitate sharing of knowledge between employees is internet and electronic mail. The most important benefits of KM are that the use of the KM may be deliver strong benefits by the way of improved quality and employee productivity. On the other hand, the most important impact of KM system was improving quality and increasing employee productivity, followed by efficiency, then employee creativity. Based on the findings of this survey, it appears that KM initiatives are getting well grounded in Turkish manufacturing firms. Top management is aware about the need for a cultural

transformation needed for the purpose. Consequently, the management in Turkish companies should encourage staff members to start thinking within their organizations who could benefit from their own knowledge, and determining which organization departments and colleagues have information that, in turn, would be useful to them. The possibility of exchanges of information with departments that share similar goals and needs could then be investigated. Staff members need to accept the concept that shared knowledge increases the total power of the group or organization. Management should also consider how knowledge is transferred when a staff member resigns. A system that not only encourages knowledge transfers among staff members, but also prevents exiting staff members from taking everything they know with them when they leave, could prove twice as profitable. This research produces some useful insights and leaving a number of issues for future research. The study could be extended to consider the impact of other elements of the KM cycle or other managerial themes, such as technological readiness and management support.

References

- [Al-Athari and Zairi 2001] Al-Athari, A., Zairi, M.: “Building Benchmarking Competence through Knowledge Management Capabilities –An Empirical Study of the Kuwaiti Context”, *Benchmarking: An International Journal*, 8, 1, (2001), 70-80.
- [Awad and Ghaziri 2004] Awad, E., Ghaziri, H.: *Knowledge Management*, Pearson Education, New Jersey, NJ (2004)
- [Balla et al. 1999] Balla, H.: “Knowledge Management Comes of Age”, 13, 7, (1999), 22-29.
- [Beccerra-Fernandez et al. 2004] Beccerra-Fernandez, I., Gonzalez, A., Sabherwal, R.: *Knowledge Management: Challenges, Solutions and Technologies*, Pearson- Prentice Hall, Upper Saddle River, NJ (2004)
- [Bhatt 2001] Bhatt, G.: “Knowledge Management in Organizations: Examining the Interaction between Technologies, Techniques and People”, *Journal of Know. Man.*, 5, 1, (2001) 68-75.
- [Clarke and Rollo 2001] Clarke, T. & Rollo, C.: “Corporate Initiatives in Knowledge Management, Education + Training, 43, 4/5, (2001), 206-214.
- [Coates 1999] Coates, F.: “The Inevitability of Knowledge Management”, *Research Technology Management*, 42, 4, (1999), 6-7.
- [Davenport and Prusak 1998] Davenport, T.H., Marchand, D.A.: “Is KM Just Good Information Management?”, *Financial Times*, March 3, (1998), 2-3.
- [Demerest 1997] Demerest, M.: “Understanding Knowledge Management”, *Journal of Long Range Planning*, 30, 3, (1997), 374–384.
- [Geisler 2007] Geisler, E.: “A typology of Knowledge Management: Strategic Groups and Role Behavior in Organizations”, *Journal of Knowledge Mangement*, 11,1, (2007), 84-96.
- [Gupta et al. 2005] Gupta, S., Bhatanagar, V., Wasan, S.: “Architecture for Knowledge Discovery and Knowledge Management”, *Knowledge and Infor. Systems*, 7, 3, (2005), 310-19.
- [Huosong et al. 2003] Huosong, X., Kuanqi, D., Shuqin, C.: “Enterprise Knowledge Tree Model and Factors of KMS based on E-C”, *Journal of Knowledge Management*, 7,1, (2003) 96.

- [Huysman and de Wit 2002] Huysman, M., deWit, D.: *Knowledge Sharing in Practice*, Kluwer, Dordrecht (2002)
- [Liao 2003] Liao, S. H.: "Knowledge Management Technologies and Applications – Literature Review from 1995 to 2002", *Expert Systems with Applications*, 25, 2, (2003), 155-64.
- [Lim and Klobas 2000] Lim, D., Klobas, J.: "Knowledge Management in Small Enterprises", *The Electronic Library*, 18, 6, (2000), 420-432.
- [Marshall 2000] Marshall, M.P.: *Managing Your Company's Knowledge*, Agency Sales, Irvine.
- [McAdam and McCreedy 1999] McAdam, R., McCreedy S.: "The Process of Knowledge Management within Organizations: A Critical Assessment of Both Theory and Practice", *Knowledge and Process Management*, 6, 2, (1999), 101-113.
- [Moffett 2003] Moffett, S. McAdam, R., Parkinson, S.: "An Empirical Analysis of Knowledge Management Applications", *Journal of Knowledge Management*, 7, 3, (2003), 6-26.
- [Muthusamy and Palanisamy 2004] Muthusamy, S., Palanisamy, R.: "Leveraging Cognition for Competitive Advantage: A Knowledge-Based Strategy Process", *Journal of Information and Knowledge Management*, 3, 3, (2004), 259-72.
- [Neef 1999] Neef, D.: "Making the Case for Knowledge Management: The Bigger Picture", *Management Decision*, 37, 1, (1999), 72.
- [Nonaka and Takeuchi 1995] Nonaka, I., Takeuchi, H.: *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*, Oxford Uni. P., Oxford (1995).
- [Sanchez 2004] Sanchez, R.: "Tacit Knowledge versus Explicit Knowledge: Approaches to Knowledge Management Practice", No. 04-1, Working Paper Series, Department of Industrial Economics and Strategy, Copenhagen Business School (2004)
- [Scarborough et al. 1999] Scarborough, H.; Swann, J., Preston, J.: "Knowledge Management: A Literature Review", Institute of Personnel Development Report, London (1999).
- [Schwartz et al. 2000] Schwartz, D.G., Divitini, M., Brasethvik, T.: *Internet-based Organizational Memory and Knowledge Management*, Idea Group Publishing, Hershey, (2000)
- [Schwarzwalder 1999] Schwarzwalder, L.: "Librarians as Knowledge Management Agents", *Econtent*, 22, 4, (1999), 63-5.
- [Singh et al. 2006] Singh, M.D., Shankar R., Narain, R., Kumar, A.: "Survey of Knowledge Management Practices in Indian Manufacturing Industries", *Journal of Knowledge Management*, 10, 6, (2006), 110-128.
- [Stojanovic et al. 2002] Stojanovic, L., Stojanovic, N., Handschuh, S.: "Evolution of the Metadata in the Ontology-Based Knowledge Management Systems", German Workshop on Experience Management, Berlin (2002)
- [Takahashi and Vandenbrink 2004] Takahashi, T., Vandenbrink, D.: "Formative Knowledge", *Journal of Knowledge Management*, 8, 1, (2004), 64.
- [Tiwana 2001] Tiwana, A.: *The Essential Guide to Knowledge Management*, Prentice Hall, Upper Saddle River, NJ (2001)
- [Zack 1999] Zack, M.: "Developing a Knowledge Strategy", *California Management Review*, 41, 3, (1999), 125-45.
- [Zhang and Zhao 2006] Zhang, D., Zhao, L.: "Knowledge Management in Organizations", *Journal of Database Management*, 17, 1, (2006), 1-8.

Becoming Knowledge Focused: Developing a Practice for Managing Knowledge in International Organizations

Ugo N. Ugbor

(Vienna University of Technology, Austria
uugbor@opec.org)

Carlos Chanduvi-Suarez

(Vienna University of Technology, Austria
C.Chanduvi-Suarez@unido.org)

Abstract: All organizations need knowledge, and international organizations are no different. Intellectual capital in international organizations needs to be nurtured to optimize their performance; and understanding the challenges of managing knowledge, and aligning technology solutions are crucial to achieve that end.

This paper explores the knowledge in international organizations and the dynamic interactions between stakeholders. It also proposes tools to classify the various knowledge products and to contextualize knowledge initiatives in international organizations and other types of non-profit organizations interested in knowledge management.

Organizational leaders, knowledge managers and consultants involved in creating and implementing knowledge management initiatives aimed at process improvements in international organizations context could benefit from the reflections and learning insights from this paper.

This paper does not address efficiency matters from the political or governing sphere, nor does it propose specific programmatic changes to work of international organizations; rather it presents a pragmatic “bottom-up” approach, based on case studies and interviews conducted at eight different international organizations, including the European Union (EU), the International Atomic Energy Agency (IAEA), the International Maritime Organization (IMO), the OPEC Fund for International Development (OFID), the Organization of the Petroleum Exporting Countries (OPEC), the Organization for Security and Co-operation in Europe (OSCE), the United Nations Industrial Development Organization (UNIDO) and the United Nations Office on Drugs and Crime (UNODC) and were classified according to their functional orientation. Over 30 interviews were conducted, with organizational leaders, knowledge managers, IT managers and human resources managers. Emphasis was on UNIDO and IAEA, as these two had developed independent knowledge management programs.

The grounded theory methodology as described by Strauss and Corbin [Strauss and Corbin, 1990, Strauss and Corbin, 1994], using interviews and document studies for data collection, combined with findings from literature, helped identify processes involved in managing knowledge in international organizations. The KM processes discovered suggests that managing knowledge in international organizations is a complex cyclical activity that leads an international organization to become more knowledge focused.

Keywords: Knowledge management, international organizations

Categories: A.0, SD I.2.6, SD I.2.4

1 Background to research

This review is limited to intergovernmental organizations. Within this category, it is important to understand how their functions have evolved in relation to the changing global landscape, particularly as regards the debate around international organizations' operational and normative functions. Evidence on the classification is drawn from various public documents, including the organizations' websites.

1.1 Normative functions

Normative functions are intended to influence the perceptions of actors in the international community within a specified set of issues [Magarinos, 2005], while implementation is left to an organization's stakeholders [Potter, 1935]—specifically its member countries. Such international organizations function mainly through forums for debate on issues of economic and political significance, setting targets and monitoring the performance of member countries vis-à-vis their regulatory mandate. For example, the International Maritime Organization (IMO) is responsible for over 40 international conventions and agreements related to maritime safety and environmental standards. The treaty on non-proliferation of nuclear weapons (NPT) established a safeguard system under the responsibility of the International Atomic Energy Agency (IAEA), which also plays a central role in nuclear technology transfer for peaceful purposes. The United Nations Office on Drugs and Crime (UNODC) is responsible for developing and implementing treaties and conventions related to drugs crime and terrorism. Similarly, the Organization of the Petroleum Exporting Countries (OPEC) was established to help co-ordinate and unify the petroleum policies of its members [OPEC, 2005].

1.2 Operational functions

Operational functions involve the implementation in recipient countries of decisions made by participating governments [Magarinos, 2005]. Whilst many normative international organizations have an operational component, some international organizations with operational functions have no formal regulatory mandate, and as such they provide consultancy services and implementation assistance for their members. For example, the work of the United Nations Industrial Development Organization (UNIDO) is focused on three broad thematic priorities [UNIDO, 2005]: poverty reduction, building trade capacity and energy and environmental concerns. The Organization for Security and Co-operation in Europe (OSCE) is another example of a operational-oriented organization, concerned with security issues, ranging from the politico-military to the socio-economic dimension [OSCE, 2006].

International organizations with a normative orientation tend to be more focused than international organizations with a more operational orientation, as the former generally create and maintain specific international agreements, while the latter provide consultancy services and implementation assistance on a broad range of issues. The dynamics of the two types of organization have implications for the types of knowledge that they are dependent on. The following [Section 1.3] examines the knowledge types against organizational orientation.

1.3 Organizational orientation and knowledge

Blackler [Blackler, 1995] proposed that different types of organizations depend on five different types of knowledge, including embrained, embodied, encultured, embedded and encoded knowledge. Encoded knowledge affects the other types of knowledge, depending on those that have adequate knowledge to decode meaning from the externalized symbols.

In the context of international organizations' functions, all knowledge types are present, although in varying degrees, based on the orientation of the organization (cf. Figure 1). The operational functions of international organizations require them to share the problem of implementing policy decisions. The knowledge processing load [Zack, 2000] therefore focuses on novel problem solving, requiring a larger degree of embodied, encultured and embrained knowledge. Problem solving in technical co-operation projects carried out by international organizations is not standardized, requiring direct contributions of key technical specialists (embodied knowledge), communications and collaboration with stakeholders (encultured knowledge), and the creative ability of experts (embrained knowledge).

Normative functions depend more on established research knowledge (embedded knowledge), intended to influence the perceptions of international actors within a specified set of issues [Magarinos, 2005], while implementation is left to member countries [Potter, 1935]. In the case of OPEC, research into oil supply and demand is informed by current oil market fundamentals [Reynolds, 2000]. Trends in the oil market inform the petroleum policies of the member countries, but the implementation of oil production levels vis-à-vis quotas is left to the members.

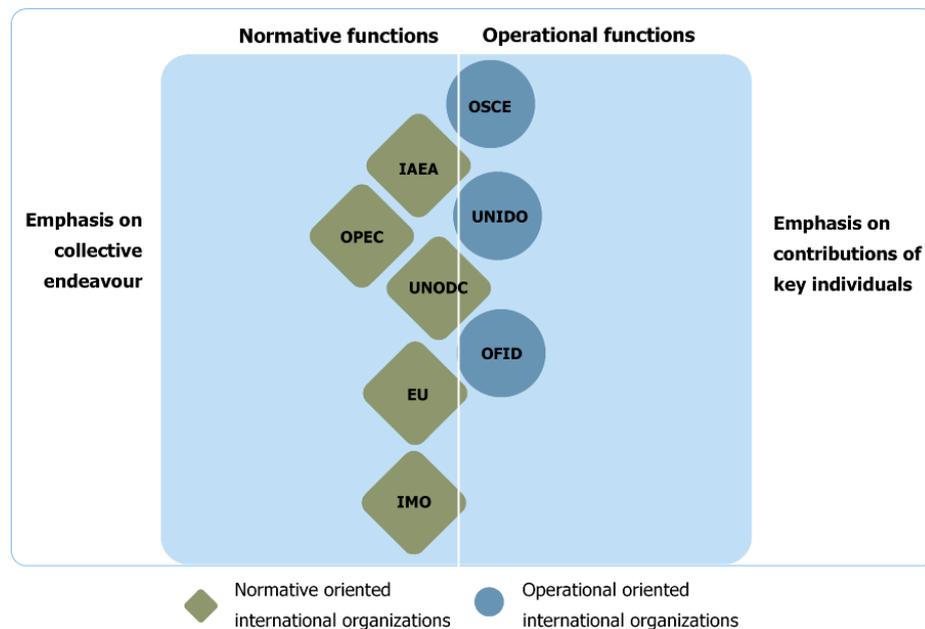


Figure 1: Organizational orientation

2 Knowledge interaction between stakeholders

The knowledge flows between organizations (normative or operational oriented) stakeholders is shaped by the type of knowledge interaction with stakeholders. The primary stakeholders of international organizations are its member countries and the organizations staff. The knowledge relations can be categorized into three interrelated services including: project management, expertise, and advice. A hieratical depiction of the service requests are shown below (Figure 2), juxtaposed against KM strategy [Hansen, et al., 1999], and organizational functions.

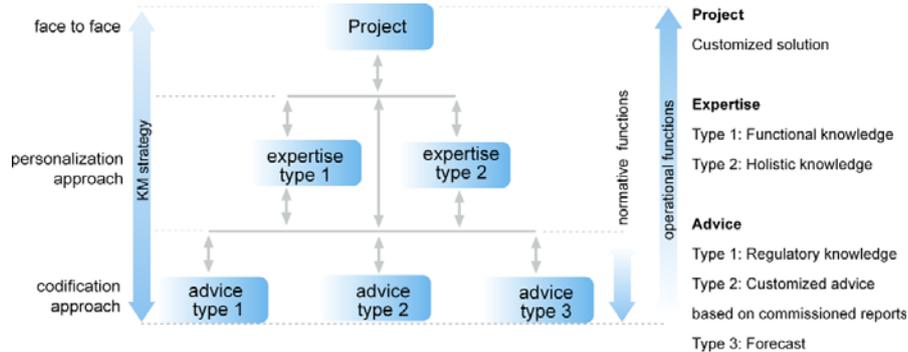


Figure 2: *Technical cooperation KM service request hierarchy*

Project requests, at the top of the hierarchy (cf. Figure 2) imply strategy formulation: assessing a country's strengths, weaknesses and opportunities and threats related to the request. A request for expertise (cf. type 1 or 2) usually requires transferring specialists across national boundaries.

These specialists (cf. expertise type 1 or 2) are managed by the organization in a database, alongside their competencies. A request for advice (cf. type 2) can be a result of a commissioned project request, where as an outcome of feasibility studies, other related sectors are identified. Advice (cf. type 1 and 3) can also be automated, in the sense that member countries have access to the organizations website and knowledge portal, where they can search for needed information.

Stakeholder requests (cf. advice type 1, 2 and 3) of a normative oriented organization are a result of geopolitical concerns related to the organizations mandate. These can be categorized into general public concerns (media, academics, industry etc) and member country concerns. To address such concerns, use of explicit knowledge (cf. advice type 1, 2 and 3) (treaties, research, commissioned reports, agreements, and so on), are needed to build consensus, influencing stakeholder behaviour and reaching new agreements, taking into consideration implicit factors such as calculative trust [Child, 1998].

3 Knowledge processing activities

Using the knowledge chain model [Holsapple and Singh, 2000], managing knowledge in international organizations can be better understood in terms of knowledge processing activities. The outcome of knowledge management (cf. Figure 3) is used as key performance indicators, and their applicability at each stage in the knowledge chain is depicted graphically. Examples of activities categorized alongside the functions of international organizations can be seen in the figure below (cf. Figure 3).

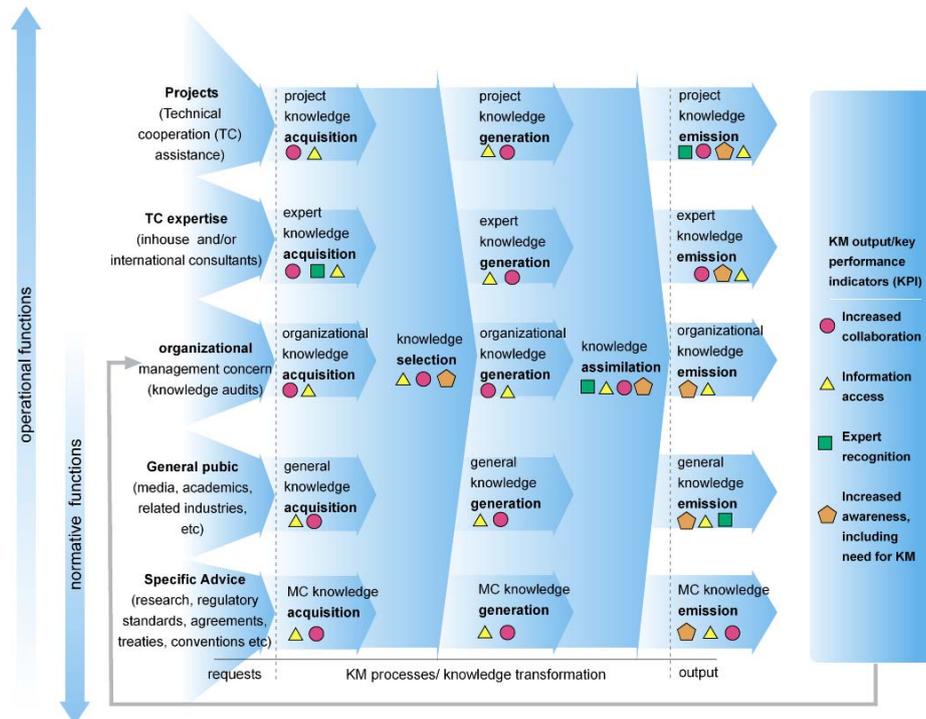


Figure 3: Knowledge chain for international organizations

The knowledge chain above (Figure 3) depicts the knowledge processing activities involved in both roles (normative and operational), embedded with the key performance indicators (KPIs) at each stage of the knowledge chain. The KPIs are sorted from left to right in the order of importance below the knowledge chain activity (Figure 3). For example “acquiring” the relevant expertise to implement parts of projects, hinges on successful collaboration among stakeholders involved, followed by a recognition of the most qualified person to do the job (if the expert is a national of the country), as well as access (information access) to an international expert database—in case no national experts are identified—that the project manager can pool from to help in the project implementation. A general listing of types of activities involved in the knowledge chain is shown below in Figure 4.

Building a knowledge dimension into the management practices in international organizations is about making knowledge flow and transfer within and beyond the organization and its member countries as fluid and effective as possible. This is in contrast to competitive environments where “knowledge leaks” or “knowledge loss” to the competition is a challenge.

Knowledge managers tend to focus their efforts differently based on functional orientation of the organization as shown below (Figure 4). For example in IAEA, a normative oriented organization, KM efforts were focused on organizing and preserving nuclear knowledge. In UNIDO, an operational oriented organization, the KM efforts were focused on knowledge sharing, and how to ensure that there is adequate feedback from the projects implemented in the field (member countries) into future projects.

A two dimensional matrix that describes the knowledge chain activities against four primary knowledge management challenges [Jordan, et al., 1998] is shown below (Figure 4). Normative oriented organizations (cf. Section 1.1) tend to focus their knowledge processing activities on problems (Figure 4, quadrant 1) within a specific domain. The newly developed integrated energy information system in OPEC (also referred to as the “chart bank”), is used to reduce complexity, by presenting complex oil market information in graphical form, in the organizations portal, enabling staff and member country officials to “know what the organizations knows” (Figure 4, quadrant 1), thereby making better use of the knowledge assets stored in databases. In operational organizations a focus is on solving novel problems. For example the OSCE, an operational oriented organization, is mandated with a wide range of tasks in Kosovo [OSCE, 2007]: from institution building and democracy-building to promoting human rights and rule of law. Knowledge generation (Figure 4, quadrant 2) and innovation are keys to remaining relevant to the stakeholders of operational oriented organizations.

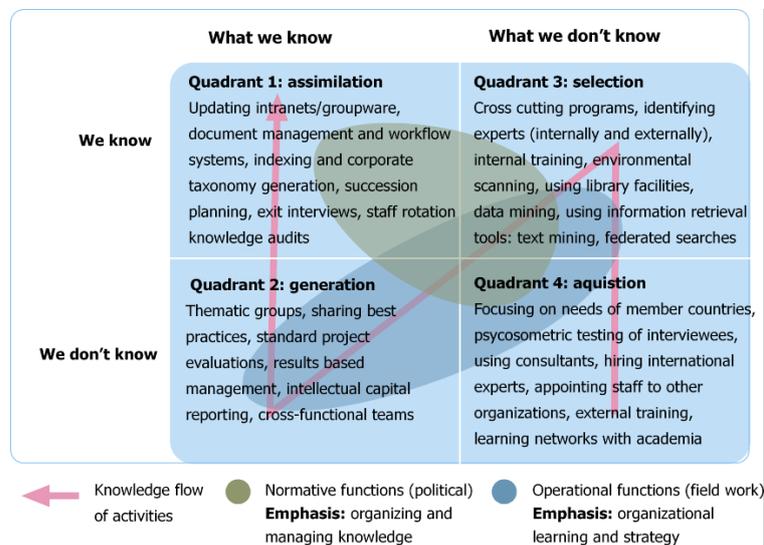


Figure 4: KM focus in international organizations

4 Conclusions

KM research has rarely studied the strategy for becoming knowledge focused in international organizations, especially from the view of the organizations. Such research has confined itself to identification and measurement of adaptation [CCPOQ, 2000] or “maladaptation” of KM theory [Zack, 2000], to international organizations, with evidently poor results [CCPOQ, 2000]—only 2 of the 17 international organizations surveyed reported “successful” results (*ibid.* p.10).

The key areas of success to date reported by knowledge managers at international organizations are increased collaboration, information access, recognition for expertise and finally an awareness of the need for KM within and beyond the organization and its member countries. Many old rules of competitive advantage of KM breakdown in an international organization—they operate not for profit, but to increase international cooperation and collaboration. The calculative interests of stakeholders [Child, 1998] to implement KM requires time and conscious effort, the differentiating factor between business as usual and managing knowledge as a needed “meta activity” to meeting the organizations goals.

According to the knowledge-based view of the firm from literature such as Rumelt [1984], Wernerfelt [1984], Dierickx [1989], Barney [1991], Hall [1993], the main emphasis is on building and managing reflective human capacity to know and derive value from such knowledge, and to offer a process perspective on how to improve work. Therefore the knowledge-chain activities in Figure 3 provides a process perspective by complementing conventional approaches to an international organization’s functions, fostering best practices in managing knowledge in this context.

Acknowledgements

This research has been supported by the Institute of Design and Assessment of Technology, Vienna University of Technology, Austria.

References

- [Barney, J. B. 1991], Barney, J. B.: "Firm resources and sustained competitive advantage"; *Journal of Management*, Issue 17 (1991), pp. 99-120
- [Blackler, F. 1995], Blackler, F.: "Knowledge, knowledge work and organizations: an overview and interpretation"; *Organizational Studies*, Vol.16, No. 6 (1995), pp. 1021-46
- [CCPOQ 2000], CCPOQ: "Knowledge Management and Information Technology"; [online], 17th Session of the Consultative Committee on Program and Operational Questions (CCPOQ), Available from: <http://www.unssc.org/web1/programmes/km/documents/crp-14.pdf> [Accessed 22 July 2006], (2000)
- [Child, J. 1998], Child, J.: "trust and International Strategic Alliances". IN Lane, C. and Bachmann, R. (Eds.) *Trust within and between organizations*. New York, Oxford University Press.(1998)
- [Dierickx, I. and Cool, K. 1989], Dierickx, I. and Cool, K.: "Asset stock accumulation and sustainability of competitive advantage"; *Management Science*, Issue 35 (1989), 1504-1511

- [Hall, R. 1993], Hall, R.: "A framework linking intangible resources and capabilities to sustainable competitive advantage"; *Strategic Management Journal*, Issue 14 (1993), pp. 607-618
- [Hansen, M. T., Nohria, N. and Tierney, T. 1999], Hansen, M. T., Nohria, N. and Tierney, T.: "What's your strategy for managing knowledge"; *Harvard Business Review*, Vol. 77, No 1, (1999), 106-16
- [Holsapple, C. W. and Singh, M. 2000], Holsapple, C. W. and Singh, M.: "The knowledge chain model"; Third Annual Conference of the Southern Association for Information Systems (SAIS), (2000),
- [Jordan, B., Goldman, R. and Eichler, A. 1998], Jordan, B., Goldman, R. and Eichler, A.: "A technology for supporting knowledge work: the reptool". IN Borghoff, U. M. and Pareschi, R. (Eds.) *Information Technology for Knowledge Management*. Berlin, Springer-Verlag.(1998)
- [Magarinos, C. 2005], Magarinos, C.: "Economic development and UN reform"; United Nations Industrial Development Organization (UNIDO), Vienna, Austria (2005)
- [OPEC 2005], OPEC: "OPEC Long-Term Strategy"; [online], OPEC Secretariat, Available from: <http://www.opec.org/library/Special%20Publications/pdf/OPECLTS.pdf> [Accessed 16 March 2006], (2005)
- [OSCE 2006], OSCE: "The Organization for Security and Co-operation in Europe"; [online], The Organization for Security and Co-operation in Europe, Available from: <http://www.osce.org/> [Accessed 29 August 2006], (2006)
- [OSCE 2007], OSCE: "OSCE Mission in Kosovo"; [online], The Organization for Security and Co-operation in Europe, Available from: <http://www.osce.org/kosovo/> [Accessed 25 March 2007], (2007)
- [Potter, P. B. 1935], Potter, P. B.: "Classification of international organizations"; the *American Political Science Review*, Vol. 29, No. 2 (1935), pp. 212-224
- [Reynolds, D. B. 2000], Reynolds, D. B.: "The case for conserving oil resources: the fundamentals of supply and demand"; *OPEC Review*, Vol. 24 No. 2 (2000), pp 71-86
- [Rumelt, R. P. 1984], Rumelt, R. P.: "Towards a strategic theory of the firm". IN Lamb, R. B. (Ed.) *Competitive Strategic Management*. New Jersey, Prentice Hall.(1984)
- [Strauss, A. and Corbin, J. 1990], Strauss, A. and Corbin, J.: "Basics of Qualitative Research"; Sage Publications, Newbury Park, California (1990)
- [Strauss, A. and Corbin, J. 1994], Strauss, A. and Corbin, J.: "Grounded theory methodology: An overview. " IN Denzin, N. and Lincoln, Y. (Eds.) *Handbook of Qualitative Research*. Thousand Oaks CA, Sage.(1994)
- [UNIDO 2005], UNIDO: "UNIDO's Thematic Priorities"; [online], United Nations Industrial Development Organization, Available from: <http://www.unido.org/doc/51918> [Accessed 22 August 2005], (2005)
- [Wernerfelt, B. 1984], Wernerfelt, B.: "A resource-based view of the firm"; *Strategic Management Journal*, Issue 5 (1984), 171-180
- [Zack, M. H. 2000], Zack, M. H.: "If Managing Knowledge is the solution, then what is the problem?" IN Malhotra, Y. (Ed.) *Knowledge management and business model innovation* Hershey, PA, Idea Group Publishing.(2000)

Special Track Integrating Working and Learning in Business

Introduction

Stefanie N. Lindstaedt

(Know-Center, Austria
slind@know-center.at)

Tobias Ley

(Know-Center Graz, Austria
tley@know-center.at)

Armin Ulbrich

(Know-Center Graz, Austria
aulbrich@know-center.at)

Since the start of the IWL special track in 2005, work-integrated learning has grown into a broad and active field of research. Several research initiatives on the European level, including the APOSDLE project¹, give evidence of this growing importance. The Special Track on Integrating Working and Learning IWL'07 addresses this field of research by focussing on technological support for work-integrated learning. IWL'07 focuses on methods, tools and technologies for supporting and enabling informal learning episodes motivated by and embedded within the daily work processes.

The first four contributions to this special track address the field of work-integrated learning with a strong process focus. In the first contribution, Stoll seeks to make documentation from a standard based process oriented management system (like ISO 9000) usable for workplace learning. She also reports about first experiences obtained in a case study. Godehardt, Doebling, Faatz and Goertz present their experiences of using a workflow modelling language (YAWL) for supporting work-integrated learning and suggest some extensions. Thirdly, Thalmann and Enoksson suggest using application profiles for arranging knowledge elements for on-demand eLearning in knowledge work. Finally, Wolpers, Martin and Duval present a way to utilize contextualized attention metadata (CAM) to deal with agility in business processes.

In the next two contributions, the authors present ways to represent skills for learning purposes. Olivella proposes a task-assignment model for industry workers which predicts task performance for an employee based on his/her experiences in similar tasks. Results of an experiment are presented which support the model.

¹ APOSDLE (<http://www.aposdle.org>) is an integrated project partially funded under FP6 by the European Community within the IST work programme (project number 27023).

Colucci, Di Noia, Di Sciascio, Donini and Ragone present a system which seeks to satisfy a user's knowledge need with a search for professional profiles or for personalized training programs. The model is based on Description Logics.

Finally, the last paper of the special track by Braun and Schmidt deals with the concept of how knowledge matures over time. The concept is evaluated by performing a qualitative and quantitative analysis of a sample of Wikipedia content.

This year's special track is the third in a series of special tracks at the I-Know conference. In 2005, IWL'05 has started to investigate into more effective knowledge management methods and technologies that integrate into existing work and learning practices. With this year's special track we proceed with these efforts by augmenting our conceptualisations in the research area of Work Integrated Learning.

Organizers

- Stefanie N. Lindstaedt, Know-Center, Graz, Austria, slind@know-center.at
- Tobias Ley, Graz Know-Center, Graz, Austria, tley@know-center.at
- Armin Ulbrich, Know-Center, Graz, Austria, aulbrich@know-center.at

Program Committee

- Manuel Görtz, SAP Research, Germany
- Ton de Jong, University of Twente, The Netherlands
- Dominik Heckmann, DFKI, Germany
- Robert de Hoog, University of Twente, The Netherlands
- Jan M. Pawlowski, University of Duisburg-Essen, Germany
- Eric Ras, Fraunhofer IESE, Germany
- Andreas Schmidt, FZI, Germany
- Luk Vervenne, Synergetics, Belgium

The special Track is endorsed by



<http://www.aposdle.org>

Graz, September 2007

Stefanie N. Lindstaedt, Tobias Ley, Armin Ulbrich
Know-Center Graz, Austria

Workplace Process Integrated Learning and Knowledge Organization

Margareth Stoll

(O&ICT, EURAC Research Bozen, South Tyrol, Italy
Margareth.stoll@dnet.it)

Abstract: Standard based management systems (as ISO9001 quality, ISO14001 environmental, ISO27001 information security, ISO 22000 hygiene management systems and others) are widely implemented and based on common principles: objectives and strategies, business processes, resource management and continuously optimization. These systems must also be documented, communicated, implemented and continuously improved. Therefore they represent an optimal basis for an integrated learning organization, if the requirements of management systems are completely implemented. This documentation contains the explicit organizational knowledge, but it is almost felt as additional workload with a little or no advantage and it is not totally corresponding with the lived processes. Therefore it is hardly used as reference book for workplace need-oriented process integrated learning to improve the organization. Based on this situation we prepared the system documentation according to media-pedagogical and didactical principles and published it on organizational learning and knowledge system based on constructivist theory. In the case study the documentation support a confidence-based, open and fault-tolerant corporate and learning culture by promoting a need-oriented workplace process integrated learning, shorter initial training periods for new collaborators and a continuous optimization of the organization for securing a sustainable success for the organization.

Keywords: standard based management systems, e-learning, workplace learning, knowledge management, optimization of organization

Categories: K.6.4, SD G.1.6, SD K.3.2, SD I.2.6, K.4.3, I.2.4

1 Introduction

Many organizations of different sizes and scopes are implementing, already since several years, management systems, such as ISO9001, ISO14001, ISO27001, ISO22000 or others, which are based on international standards with common principles: organization objectives and strategies, business processes, resource management and continuously optimization of the organization. The established management system must be documented, communicated, implemented and continuously improved.

The need of developing an integrated solution for organizations and the shift to holistic management systems, which fulfill optimally all requirements of organizations, influenced the introduction of knowledge management in such standard-based management systems. Thereby the system documentation contains the entire actual explicit knowledge and supports the knowledge distribution for promoting the knowledge implementation by establishing a basis for the knowledge generation.

The documentation of standard based management systems (system documentation) was distributed for long time as books, whereby the collaborators received one-time the information and used them afterwards for possible questions scarcely ever. In the last ten years the system documentation was distributed more electronically

through web-based intranets, document management platforms or as pdf. The delivered system documentation is in these way used a little more as reference book for solving ad hoc learning needs. But changes proposal are not related to the documentation and the collaborators are not able to discuss new ideas or questions in a context-sensitive way. The change and improvement processes of the management system are isolated from the documentation.

Therefore the primary use of the system documentation is as a tool for the process optimization during the system introduction and afterwards it is primarily for the auditor to fulfill the standard requirements. It is usually changed only once a year, if at all and it frequently does not correspond to lived processes. Moreover in addition to it instructions and forms are developed as hidden system.

Sometimes also external sample manuals are used, but they do not correspond with the lived processes and are providing no contribution to the organizational knowledge base.

The requests for always shorter change cycles, the ever stronger IT involvement and the great importance of the system documentation as knowledge base for organizations led us to the extension of a constructivist learning platform as an organizational learning and knowledge system [Back and Bendel, 2002], [Bick and Pawlowsky, 2007] for distributing the system documentation of a standard based and knowledge integrated management in accordance to media-pedagogical and didactical principles.

Firstly we present the common requirements of international standards for management systems. Based on these requirements we describe the project objectives and our approach. Finally we document the project experiences and results of the implementation in different organizations with distinct management systems.

2 Main requirements of standard based management systems

The standards for management systems have different specialized focuses, but are all based on common principles [Osanna et al. 2002], [EN/ISO 9001:2000]:

- organizations objectives and strategies must be established regarding customer requirements;
- all business processes including management process, support processes, resource processes and optimization processes must be defined and promote the optimized fulfillment of the organization objectives under the focus of the respective standard;
- process oriented resource management must be promoted including personnel development, IT – management and other infrastructures, tools and instruments;
- the organization and their objectives and strategies must be continually optimized according to established processes in sense of a PDCA cycle (plan, do, check, act);
- the established management system must be structured and systematically documented and communicated within the organization and the collaborators must be continually motivated for implementing the system and for recommending improvements.

These standard based management systems are implemented more frequently in a holistic way. In accordance with the organizational purposes and objectives are integrated in a management system different aspects like quality, environment, hygiene, data security, occupational health and safety, as well as personnel development, re-

source management, IT - management, communication management, controlling and also knowledge management.



Figure 1: Main requirements of standard based management systems

3 Project objectives

How can the system documentation be modified and communicated in order to contribute to the organization success and to be considered, used and improved as organizational knowledge base?

By preparing the system documentation in accordance to media-pedagogic and didactical principles and managing it on an organizational learning and knowledge system by extending a constructivist learning system we expect to foster:

- workplace need-oriented process integrated learning,
- organizational development and knowledge generation,
- employee involvement and knowledge exchange,
- practice-oriented system documentation
- more quickly initial trainings for new collaborators,
- optimization of the organization and improved achievement of its objectives.

Thus the learning, knowledge generation, knowledge representation, knowledge communication and knowledge application should be optimized and the organization will be promoted in achieving its objectives including knowledge objectives.

4 Approach

Firstly the system documentation must be established [see 2 Main requirements of standard based management systems]. Considering the needs and expectations of all interested parties we establish the organization policy with consistent strategies and goals integrating knowledge objectives and plans. Afterwards the applied processes are analyzed bottom up by interviewing the collaborators involved and by integrating knowledge management and optimized in accordance with the objectives of the organization. The necessary resources, tools, instruments, trainings required for achieving the objectives and for improving the management system must be identified, made available and continually optimized. Also the training and human resource development processes must be optimized and established. The organization must also plan and implement the monitoring, measurement, analysis and optimization processes

needed to continually improve the effectiveness of the management system. Thus also knowledge management and learning organization must be integrated and their improvement must be structured and systematically planned. The entire system documentation must correspond with the lived processes. All improvements or changes must be approved by the involved collaborators, documented, communicated and implemented.

After the development of the management system documentation, it must be prepared regarding media-pedagogical, motivation-psychological and didactical principles [Reinmann-Rothmeier, 2002]. According to these principles it must be divided into small modules, which must be typecasted and functionally well structured. Thus, we use a constructivist method for organizing the content. In order to support the collaborators need-oriented workplace process, the way for accessing the single modules should be as short as possible, optimal structured and practice oriented.

All needs and expectations of the collaborators must be analyzed and considered as early as possible. They demand particularly an efficient and effective search function and a clear structured system. Therefore we need effective indexing and different start assistances (for new collaborators, department oriented, management oriented, topics referred, based on the standard and others). Apart of the self-driven learning approach, the system must also offer guided learning for new collaborators or for collaborators with little IT competences.

After the appropriate editing the system documentation can be published on the organizational knowledge and learning system. The upload of the content should be user-friendly and simply, because it is normally an activity done by the system manager, which has organizational, managerial and controlling competences and competences in developing management systems and must not be an IT expert. The constructivist approach in learning requires the possibility of linking concepts and objects following one's own needs. This feature must be available also for graphical objects and different document and media formats, which are widely used in hypertext contents.

To promote the acceptance of the system the collaborators must be trained on the handling the system and acquire also the necessary media competence.

Subsequently, questions will be answered by means of the discussion forum, the help desk and personal partners, e.g. the knowledge and/or system manager. Using the constructivist organizational learning and knowledge system every collaborator can introduce his/her suggestions, ideas or questions directly in a context-sensitive way by referencing all the possible objects of the system documentation. Thus, ambiguous formulations in the documentation are showed up and the system documentation - organizational knowledge base - becomes the basis for the continuously optimization of the organization. Thereby we integrate also optimal process modeling, process standardization and transparency with need and objective oriented flexible process implementation. The collaborators have sometimes problems to distinguish between problem and optimization, or they do not know to which section of the system documentation they should relate their suggestion or question. Therefore we have integrated the organizational knowledge and learning system with the existing workflow driven problem reporting system (help desk system).

Following the management system requirements also clearly structured processes for the introduction of new collaborators, for the maintenance of the system documen-

tation and for the management of questions, optimizations and problems must to be established, documented, implemented and maintained.

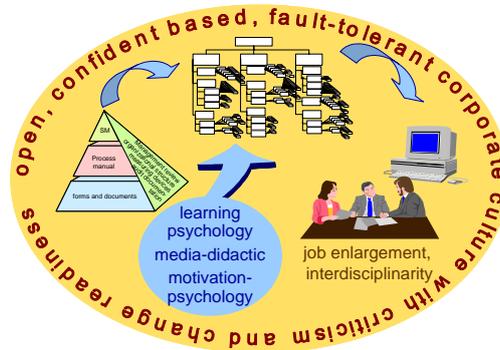


Figure 2: Integrating need oriented workplace process integrated learning and standard based management systems

4.1 Requirements for an organizational learning and knowledge system

Based on our project experiences an organizational learning and knowledge system for standard based management systems demands, in addition to the general requirements of learning or knowledge management systems, the following particular characteristics:

- It must be simple and intuitive to handle, it must provide the possibility to use different views of (or leading to) the same object, different start facilities, comfortable search functions, filtering of content using object types, simple uploads of content and links to external literature and it must promote individual learning and organization by allowing personal bookmarks, annotations [Auinger and Stary, 2005], summaries and notes, as well as glossary, FAQs, etc..
- It must offer secure context-sensitive communication to all elements (discussion forum, chat), as well as newsgroups, wiki, newsletters and whiteboards, support the knowledge distribution, the collective knowledge generation and collective learning.
- Uploading and the administration of new content must be simple, with as large didactical and media pedagogic support as possible. There must be the possibility to insert, annotate and discuss context sensitive the content of most different document and media formats (as text, graphic, table, picture, sound records, video). Also creativity tools should be integrated.
- Due to the requirements of the standards for management systems, we need the administration of access rights, the support of change processes, versioning with change history and the efficiently and traceable distribution / communication of new or changed content. Depending on the organization culture also testing and examination tools for the traceable learning must be integrated.
- The ideas of the collaborators, their discussion contributions and problem reporting must be implemented in accordance with the established systematic and structured

process following the standard requirements and their effectiveness must be evaluated.

- Open interfaces must be available for fostering optimal connectivity with other systems in order to support process integration and simplify the administration. Within the controlling process all measured values should be evaluated by the responsible departments and consequently changes and optimizations should be implemented using the organizational learning and knowledge system.
- The web-based accessibility of the organizational learning and knowledge system supports mobile and working hour's independent learning and teleworking, making also easier the integration of absent collaborators.

4.2 Success factors for an organizational learning and knowledge system

Corporate culture processes and IT technologies must be integrated optimally according to the objectives of the organization, to the collaborators needs and requirements and considering media-pedagogical and didactical principles, in order to successfully introduce an organizational learning and knowledge system into management systems. The system is thereby only a tool, which supports the optimization of the organization so far as this is admitted by the culture. Therefore we need an open, confident based, fault-tolerant corporate and learning culture with criticism and change readiness. The collaborators should be interested in new knowledge, able for self-driven learning, have personal employment, team ability and change willingness apart from necessary IT-competences. All managers must use constantly and actively the system and motivate their collaborators in following these principles, promoting in this way learning and new knowledge generation. A strengthening point for achieving this is maintaining an optimal communication internally as well as with external partners.

By introducing this system the system manager and/or knowledge manager extend their own job, needing also the acquisition of the necessary media-pedagogical and didactical knowledge for preparing the content and the necessary skills for supporting e-learning and knowledge management.

Sufficient IT-infrastructure and IT-support are also very important for the project success. Only by using an organizational learning and knowledge system, which meets as far as possible the stated requirements [see 4.1 requirements to an organizational learning and knowledge system] and by promoting a workplace need-oriented process integrated learning, a continuously optimization of the organization, in accordance with its objectives, and a sustainable organization success are secured.

5 Project experience and results

The concepts expressed so far have been implemented successfully in several organizations with different management systems.

The interdisciplinarity was a great challenge. Knowledge about organization theory, management methods and standards were required for the preparing and structuring the content, information-technical knowledge were needed for extending the learning platform to meet the requirements of an organizational learning and knowledge sys-

tem; didactical, media-pedagogical and communication science knowledge were required for editing the contents.

The implemented learning system was improved and meets now some of the requirements stated for an organizational learning and knowledge system.

5.1 Achieving the project objectives

The described concept has been implemented in medium-size service organizations by integrating knowledge management into management systems. Most of the organization's collaborators own good media competences and use frequently e-tools.

Nevertheless the platform used in the case studies fulfills only some of the system requirements [see 4.1 requirements to an organizational learning and knowledge system], the adoption of the organization adapted system documentation according to media-pedagogical and didactical principles and taking into consideration a confident based fault-tolerant corporate culture, leads the following case study results collected by measuring the system accesses / user contributions and interviewing the leadership and collaborators:

- Workplace need-oriented process integrated learning: the accesses to the system documentation are increased monthly at averaged two accesses per collaborator.
- Integrated learning organization: we receive five times more suggestions and ideas, which improves the organization.
- Employee involvement and knowledge sharing: the communicated ideas, problems and suggestions are discussed and read on the average by a quarter of the collaborators. Thus, the advantages and disadvantages of ideas are discussed and examined by all departments before their possible implementation and they are substantially more balanced and more considered for implementation.
- Practice oriented documentation for workplace requirements: while conventional system documentation is usually adapted only once a year, now unclear formulations or missing regulations are soon analyzed and immediately changed according to established processes and automatically communicated to all. Thus, the documentation is adapted optimally to the changing requirements of the organization.
- Shorter initial training periods for new collaborators: new collaborators are quickly introduced into the system at their first working day and afterwards they learn in a self-driven need oriented way, all relevant contents for their specific work. The lead time could be abbreviated around a quarter. New collaborators identify themselves substantially earlier with the corporate culture, they feel themselves also more integrated into the organization, can execute their tasks faster well, whereby the productivity increase and possible errors are avoided.
- Knowledge optimization by learning, knowledge generation, knowledge representation, knowledge communication and knowledge use: the collaborators bring in their ideas and questions and improve thereby the knowledge base (system documentation) and the organization itself. Thus, also the achievement of the organization objectives including the knowledge objectives is optimized.
- Promoting an open, confident based, fault-tolerant corporate and learning culture with criticism and change readiness: the manager of an organization with an organizational knowledge and learning system is particularly enthusiastic for the culture promotion.

6 Outlook

Learning or knowledge management systems must be extended [see 4.1 requirements to an organizational learning and knowledge system] to organizational knowledge and learning systems for need-oriented, workplace process integrated learning.

By introducing an organizational knowledge and learning systems for standard based management systems with integrated knowledge management all success factors should be considered [see 4.2 Success factors for an organizational learning and knowledge system].

The training of system managers should consider more the principles for didactical and media-pedagogical restructure of system documentation.

Also e-learning experts should analyse more the requirements of standard based management systems and knowledge management systems.

The standards for management systems should underline more and promote the importance of an open, confident based, fault-tolerant corporate and learning culture, knowledge management and integrated learning.

7 Conclusion

Integrating knowledge management and standard based management systems, restructuring the system documentation in accordance to media-pedagogical and didactical principles and implementing it on a constructivist organizational learning and knowledge system with an open, confident based, fault-tolerant corporate and learning culture promotes workplace need-oriented process integrated learning and a continuous optimization of the organization by means of collaborators ideas and discussion contributions.

Thus by fostering workplace need-oriented process integrated learning and improved knowledge management, the sustainable success becomes guaranteed.

References

- [Auinger and Sary, 2005] Auinger A. and Sary C.: „Didaktikgeleiteter Wissenstransfer“, Deutscher Universitäts-Verlag, Wiesbaden (2005).
- [Back and Bendel, 2002] Back A. and Bendel O.: „Lern- und Wissensportale“; in Glotz P. and Seufert S.: „Corporate University“, Institut für Medien- und Kommunikationsmanagement, Universität St. Gallen (2002).
- [EN/ISO 9001:2000] „EN/ISO 9001:2000, “Quality management systems - requirements“; ISO, (17.12.2000).
- [Osanna et al. 2002] Osanna, P., Durakbasa, M. and Afjehi-Sadat, A.: „Qualitätssicherung und Qualitätsmanagement“; Vienna University of Technology, Vienna (2002).
- [Bick and Pawlowsky, 2007] Bick M. and Pawlowsky J.M.: “A Framework for Integrated Ambient Learning and Knowledge Environments” Proc. 4th Conference on Professional Knowledge Management – Experiences and Visions, Gito Verlag, Berlin (2007), volume 2 335 - 342.
- [Reinmann-Rothmeier, 2002] Reinmann-Rothmeier G.: „Mediendidaktik und Wissensmanagement“; MedienPädagogik 10/2002, www.medienpaed.com/02-2/reinmann1.pdf.

Deploying YAWL for Workflows in Workplace-embedded Learning

Eicke Godehardt, Markus Doehring, Andreas Faatz, Manuel Goertz

Fraunhofer IGD, Germany

eicke.godehardt@igd.fraunhofer.de;

SAP AG, Germany

{markus.doehring, andreas.faatz, manuel.goertz}@sap.com

Abstract: In today's information society the border between the roles of learners, workers and teachers becomes more and more fuzzy. To support this, we are using the workflow language YAWL. It provides guidance in work processes and workplace embedded learning. Nevertheless, using workflows in user driven (e-learning) environments, triggers challenges. We discuss the ex-post OR-split, a novel recommendation of how to cope with OR-splits. Furthermore, the visualisation strategy on how to even simplify the graphical view of YAWL on the workflow during the workplace-embedded learning application is discussed.

Key Words: workflow, YAWL, workplace-embedded learning

Category:

1 Introduction

We are faced with a shift to the knowledge society. A great challenge for the next generation of e-learning systems is workplace-embedded learning, which brings electronically available learning material and knowledge items of any kind directly to the working user at her workplace. In such a scenario, the learning material will not be organized in curricula or courses any longer, but in relevance to the working situation. In particular this is challenging for knowledge intensive work. APOSDLE (Advanced Process-Oriented Self-Directed Learning Environment) [1, 2, 3] is a research project partially funded by the EU, which investigates and prototypes seamlessly integrating roles of teaching, (knowledge-intensive) working and learning in next-generation e-learning systems.

Our paper concentrates on aspects of melting working and learning. Today large organizations—but increasingly small and medium sized enterprises too—formalize their work by processes/workflows. Workflows are instances of business processes, i.e., of “a set of one or more linked procedures or activities, which collectively realize a business objective or policy goal, normally within the context of an organizational structure, defining functional roles and relationships” [4]. We restrict ourselves to those parts of the workflow, which encapsulate and order the working tasks, i.e., the order of steps a worker has to follow. These steps are expressed as the control flow of a workflow [5]. The more an e-learning system

would be able to reflect these real-world workflows, the more likely e-learning becomes workplace-embedded in those cases, where workflows determine a worker's day. Thus, the core questions of our paper are:

- How can we design and deploy the e-learning functionality in an arrangement, which is as similar as possible to real-world workflows?
- What is needed in terms of a workflow language to close the gap between workflows in e-learning and working?
- Which additional interpretations and language constructs are needed to apply a concrete workflow language in workplace-embedded learning?

The rest of the paper is organised as follows. Section 2 gives a brief overview on related work, which already used workflow systems as a part of e-learning systems. Section 3 describes our approach, in particular how we selected a suitable workflow language, its characteristics and which adjustments as well as additional implementations were necessary for modelling and graphical presentation at the workplace. We summarise our findings in section 4 and give an outlook on what to solve in future increments of the APOSDLE system. We will use the term APOSDLE for the project as well as for the overall technical system developed in the project.

2 Related Work

There are some related research projects in using workflows for learning. But none of them is workplace embedded or focusing on informal learning based on workflows. Cesarini et. al. [6] uses workflows to model and structure e-learning content. In a similar manner [7] tries to use the greater flexibility for structuring e-learning content to enhance SCORM based content. Another project [8] is also not workplace-embedded, but uses workflows to model traditional learning content to make it more adaptive.

Quite similar approaches are using business-process oriented knowledge management (BPOKM) methods or are process oriented in general [9]. For example the project LIP (Learning In Process) [10], which focuses more on e-Learning supported by processes rather than informal learning and do not use a dedicated process definition language but ontologies. Another project facing an analog direction is APO-Pilot [11]. They use processes mostly to model the learning activities. On the other hand APOSDLE focuses more on helping the knowledge worker to get her work done by providing resources and collaboration to experts matching the context of the user in one selected (pre)defined process.

To summarize all of these approaches do not have the goal to integrate informal learning into the workplace and thus still stick to fixed roles for learners and teachers in a dedicated e-learning environment.

3 Approach

This section describes our approach in detail. We explain, why we chose YAWL as a modelling language, introduce necessary constructs and present our visualization principles.

3.1 Choice of the workflow language and environment

The process of choosing a suitable workflow language was driven by the following core requirements: intuitive graphical notation of the language, full formal specification, expressiveness, availability, costs and extendability. An easy and intuitive graphical notation is necessary in the APOSDLE context to enable users with a non-IT background to model the workflow in a reasonable time. On the other hand, the workflow language should be formally specified, such that a workflow expressed in this language has a unique non-ambiguous interpretation. Along with the last two requirements, the high expressiveness of the language should minimize the “effort needed to construct models that reflect the process logic in a direct manner” [12] while manipulating/editing via a graphical user interface. This goes beyond usability—the concern is more on the logical expressiveness and the encapsulating constructs of the language.

The remaining requirements are non-functional. Availability is concerned about the existence of editing tools and a workflow management system, which should be deployable with reasonable costs, as the workflow modelling tool and the workflow engine are part of a larger environment, which will produce further costs itself, and allow extensions. We compared five approaches or languages (Petri Nets [13], OMG’s Business Process Modeling Notation (BPMN) [14], OMG’s Unified Modeling Language (UML) [14], IDS Scheer’s Event Driven Process Chains (EPC) [15], ConcurTaskTrees (CTT) [16] and YAWL [17]) by hands-on evaluation and by reflecting on/studying theoretical literature [17].

Criterion	Petri Nets	BPMN	UML	EPC	CTT	YAWL
(1) Easy / Intuitive Graphical Notation	O	O	+	O	O	+
(2) Full Formal Specification	+	O	–	O	O	+
(3) High Degree of Expressiveness	–	+	O	O	O	+
(4) Available Tools and Environment	+	+	+	+	+	+
(5) Low cost	+	–	+	–	+	+
(6) Flexibility	O	O	+	–	O	+
+ accomplished O partly accomplished – not accomplished						

Table 1: Decision Matrix for the Selection of a Workflow Management System

The YAWL language and environment fits best the requirements. It combines a clear formal specification with an easy-to-learn and (in comparison) simple-to-use language, which can express complex workflows. Moreover, YAWL is not only a theoretical approach, but comes along with free open-source applications available at [18].

The description of all language constructs and the YAWL system is beyond the scope of this paper. The concerned reader is referred to [17] for a deeper insight. In the following section we will explain our changes and additions to YAWL.

3.2 Application of YAWL constructs

After finally choosing YAWL as our underlying workflow system, we figured out, that there is still some missing functionality according to the environment in APOSDLE or workplace-embedded e-learning in general. Especially from the fact, that human beings are using and driving the system, additional needs arose. First the user (APOSDLE knowledge worker) does not have to follow the proposed path through the workflow, she can choose every task at any time. This will lead to another obstacle—the handling of OR-splits. At start time of an OR-split the system cannot determine which tasks/paths inside the OR-split will be executed by the user. Thus this decision have to be postponed, while the original behavior of YAWL as a workflow engine was to decide at the OR-split start, which paths are chosen.

We will now show, how we applied and extended the basic YAWL constructs to fulfill the requirement defined in APOSDLE.

3.2.1 Treatment of Ex-Post OR-Splits

A significant obstacle, which has to be overcome regarding the control-flow perspective is the retroactive selection of tasks after an OR-split. As shown in figure 1, the workflow engine decides about the further routing (i.e., which tasks are subsequently enabled—tasks *A* and *B*) right at the point of time when the OR-split takes place. This is exactly after the corresponding OR-split task has finished firing. The further routing depends on circumstances like user selections or internal data variables.

This kind of process structure occurs every time the users has the ability to choose n of m parallel task until she decides to finish and go to the following task. To give an example for this kind of OR-split the overall task is to book a business trip. Usually the user has to book a flight, a car at the destination and a hotel. Nonetheless all of them are optional, she might go by train, do not need a car or stay with a friend at the business location, but the user knows which of the paths she has to take to finish the booking.

The problem is, that once a decision about the further routing has been made, it is not possible to alter it any more. Nevertheless, it is sometimes desirable, especially in user-driven environments, to allow for a more **“dynamic” decision** without requiring a complex underlying model. This is an eminent issue in the APOSDLE learning environment, because a worker rarely knows in advance what she will do next, if she has the choice. For example, if he has the options to complete his travel expense report and to work on a power-point presentation, the processing of the expense report may depend on the progress she makes with the presentation.

We conclude that the exact composition of chosen tasks should not be fixed at the beginning. It should rather be permitted to enable tasks even after the processing of the actual OR-split. We call this an “ex post or-split” or “lazy or-split”. Related to our example, even if only task *A* and *B* had been selected (marked black within figure 1), task *C* could be retroactively selected, but only before the firing of the corresponding OR-join.

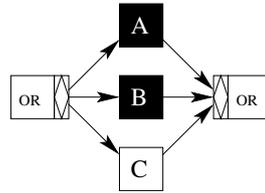


Figure 1: Problems Resulting from YAWL’s OR-Split Semantics

We elaborated two different workarounds to circumvent this problem. The first one makes use of the YAWL multiple instances (MI) process element. Instead of statically defining the control-flow using an OR-split element as shown above, we replace the whole construct with multiple instances of a composite task.

The decomposition of the MI-task is made up of a XOR-split and a XOR-join, which embrace the tasks of our former OR-split-construct (tasks *A*, *B* and *C* in figure 2). The effect of the resulting complex is, that at the starting of the MI-task, a number of instances are created (equal to the number of branches that would have been chosen by the OR-split). Until all instances have terminated, it is possible to add new instances and therefore to enable formerly inactive tasks. Within the decomposition, the XOR-split decides which particular task has to be activated. To prevent a double activation of a task, the XOR-split stores and evaluates data within the workflow engine. If an instance would cause a double activation of a task, the XOR-Split chooses the empty task *E* and the instance is consequently immediately terminated.

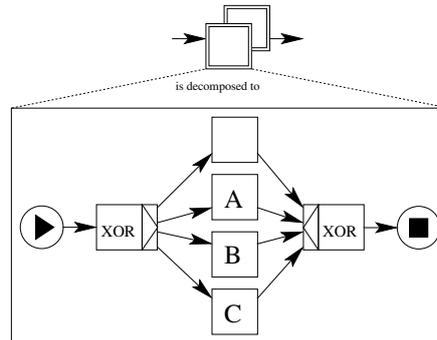


Figure 2: Solution Approach to Ex-Post OR-Split with Multiple Instances

Especially because of the double activation of tasks, we decide to use another solution approach instead. This time we realize the ex-post OR-split with the help of an initial AND-split and a cancellation region. Cancellation region is a construct from YAWL. If a task is connected to a cancellation region and the task is fired, all tasks in this region are canceled and all tokens from conditions inside the region are removed. This solution can be gathered from figure 3, where the AND-split prophylactically enables all succeeding tasks.

It is then up to the user to process and execute a selection or all of the tasks. The task complex can be finished, as soon as the explicit condition P contains a token (which again realizes the OR-join's preliminary of at least choosing and completing one branch). A characteristic of this approach is that the processing of the task complex has to be explicitly finished by the execution of the task F , which cancels all remaining active tasks and remove all tokens from conditions in the cancellation region. The cancellation of those enabled (or even running) tasks is smoothly possible. Its explicit finishing also does not constitute a critical problem within the application domain of APOSDLE. Generally, no counter-arguments against the use of this construct could be spotted. We will consequently use this **second approach** instead of the classical OR-construct in APOSDLE.

3.3 Visualizing

The intension of visualization the worker's process in APOSDLE is to give her an idea of the current task as well as the context of this task, e.g. which tasks are previous/next. On the other hand, this process visualization is also used as a user interface to select a different task, i.e., to change the worker's context.

In order to provide an intuitive user interface, we hide all special/artificial constructs from the user. Especially the joining condition P and the direct arc

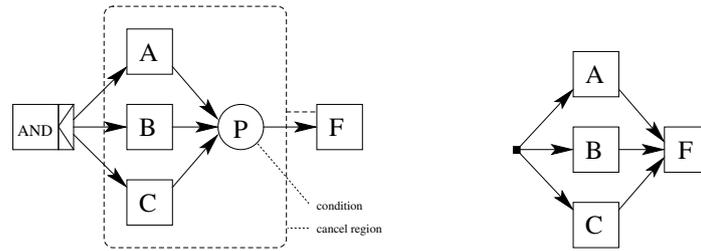


Figure 3: Ex-Post OR-Split with AND-Split and Cancellation (left side) and its simplified visualization (right side)

to this condition would confuse regular users. Every branch just ends in the task following this condition. This is the same approach YAWL uses for implicit conditions between two tasks, which are hidden as well. The Ex-Post OR-Split solution together with the visualization concept can be seen in figure 3. As we have implemented all splits in the workflow models with our ex-post or-split, we do not have to distinguish AND- from OR-splits in the user interface. Please note that this kind of guidance is specific for the less restrictive role a workflow plays when serving as a navigation in e-learning instead of data processing.

4 Conclusion and future work

In this paper we have shown how we choose YAWL as our workflow management system, based on the requirements in the APOSDLE project. We also explained in depth the obstacles we faced and our approach by using YAWL's means to fulfill our needs and to visualize this in a way, that is easy understandable even by non-IT users.

Further research topics are the integration of desktop monitoring [19] allowing us to automatically recognize task/context changes of the user. In addition the arbitrary "jumping" in a predefined process is not completely handled yet.

Other open topics are using a different way to visualize the process, e.g., using Topic Maps [20] or using the desktop monitoring to build ad-hoc processes instead of predefine them by knowledge designer.

After showing the feasibility of our approach, it would also be interesting to transform or convert other existing workflows to YAWL and analyse their behavior applying our results.

References

1. H. Mayer, W. Haas, G. Thallinger, S. Lindstaedt, and K. Tochtermann. Aposdle - advanced process-oriented self-directed learning environment. Poster Presented

- on the 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies, 30 November – 01 December 2005.
2. S. N. Lindstaedt, T. Ley, and H. Mayer. Integrating working and learning with aposdle. In *Proceedings of the 11th Business Meeting of Forum Neue Medien, 10–11 November 2005, Vienna*. Verlag Forum Neue Medien, 2005.
 3. APOSDLE. Aposdle (advanced process-oriented self-directed learning environment) - website. <http://www.aposdle.org>.
 4. Workflow Management Coalition. Terminology and glossary (wfmctc-1011). http://www.wfmc.org/standards/docs/TC-1011_term_glossary_v3.pdf, 1999.
 5. W.M.P van der Aalst, A.H.M. ter Hofstede, B. Kiepuszewski, and A.P. Barros. Workflow patterns. In *Distributed and Parallel Databases*, volume 14, pages 5–51, Norwell, 2003. Kluwer Academic Publishers.
 6. Mirko Cesarini, Mattia Monga, and Roberto Tedesco. Carrying on the e-learning process with a workflow management engine.
 7. Kwang-Hoon Kim, Hyuk-Jae Yoo, and Hak-Sung Kim. A process-driven e-learning content organization model. In *ICIS '05: Proceedings of the Fourth Annual ACIS International Conference on Computer and Information Science (ICIS'05)*, pages 328–333, Washington, DC, USA, 2005. IEEE Computer Society.
 8. Peter Brusilovsky, Elmar W. Schwarz, and Gerhard Weber. A tool for developing adaptive electronic textbooks on www. In *WebNet*, 1996.
 9. A. Abecker, G. Mentzas, M. Legal, S. Ntioudis, and G. Papvassiliou. Business-process oriented delivery of knowledge through domain ontologies. In *Proceedings of DEXA conference, TAKMA-2001, Second International Workshop on Theory and Applications of Knowledge Management*, 2001.
 10. Andreas Schmidt. Bridging the gap between knowledge management and e-learning with context-aware corporate learning. In *Professional Knowledge Management*, pages 203–213, 2005.
 11. Frank Fuchs-Kittowski, Katja Manski, Daniel Faust, Marko Prehn, and Ingo Schwenzen. Arbeitsprozessorientiertes e-learning mit methoden und werkzeugen des prozessorientierten wissensmanagement. In Arndt Bode, Jörg Desel, Sabine Rathmeyer, and Martin Wessner, editors, *DeLFI*, volume 37 of *LNI*, pages 392–401. GI, 2003.
 12. W.M.P. van der Aalst. Making work flow: On the application of petri nets to business process management. In *Lecture Notes in Computer Science*, volume 2360, pages 1–22. Springer-Verlag, 2002.
 13. Carl-Adam Petri. *Kommunikation mit Automaten*. PhD thesis, Institut für instrumentelle Mathematik, Bonn, 1962.
 14. OMG (Object Management Group). Website. <http://www.omg.org>, 2006.
 15. G. Keller, M. Nüttgens, and A.-W. Scheer. Semantische Prozeßmodellierung auf der Grundlage Ereignisgesteuerter “Prozeßketten (EPK)”. In *Veröffentlichungen des Instituts für Wirtschaftsinformatik (IWi)*, Universität des Saarlandes, volume 89, Saarbücken, 1992. Universität des Saarlandes.
 16. F. Paterno, C. Mancin, and S. Meniconi. Concurtasktrees: A diagrammatic notation for specifying task models. In *Proceedings Interact '97*, Sydney, 1997. Chapman&Hall.
 17. W.M.P van der Aalst and A.H.M. ter Hofstede. Yawl - yet another workflow language. In *Information Systems*, volume 30, pages 245–275, Orlando/New York, 2005. Elsevier Ltd.
 18. YAWL-Foundation. Yawl (yet another workflow language) - website. <http://www.yawlfoundation.org>.
 19. Robert Lokaiczkyk, Andreas Faatz, Arne Beckhaus, and Manuel Goertz. Enhancing just-in-time e-learning through machine learning on desktop context sensors, 2007. SUBMITTED.
 20. TopicMaps. ISO/IEC 13250.

*Proceedings of I-KNOW '07
Graz, Austria, September 5-7, 2007*

An Approach for On-Demand E-Learning to Support Knowledge Work

Stefan Thalmann

(University of Innsbruck, Austria
stefan.thalmann@uibk.ac.at)

Fredrik Enoksson

(Royal Institute of Technology, Sweden
fen@csc.kth.se)

Abstract: The requirements on learning support from knowledge work differ compared to traditional work. Based on those observations an approach for supporting learning in knowledge work is proposed considering requirements from e-learning as well as from knowledge management. In addition to traditional e-learning, on-demand e-learning takes the current situation of the knowledge worker into consideration to ensure learning support of knowledge work is proposed. For using a broad variety of resources in on demand e-learning a single metadata schema for describing seems not sufficient for every organisation. Therefore, application profiles appear adequate for describing resources used in the proposed approach for arranging knowledge elements. Identifying the knowledge workers current situation a learning need should be derived and to use it afterwards for selecting and delivering knowledge elements.

Keywords: on-demand e-learning, workplace learning, metadata, learning need, work-based learning, knowledge work

Categories: H.1.2, H.5.4, H.3.3, H.3.5

Introduction

Looking upon e-learning (EL) and knowledge management (KM) there are fundamental differences influencing the design of EL solutions. Currently, EL see learning as a continuous, lifelong process transferring knowledge for predefined tasks. Primary the transfer of more general knowledge usable in different situations to build competencies and more general job requirements are well supported. Therefore, predefined, didactically well prepared courses that end with an examination, to verify the learning success, are used.

In KM the demand for knowledge is triggered directly from business process, an unforeseeable knowledge gap initiates the learning process. Learning is primary seen as process to acquire knowledge required to solve current problems and after learning the knowledge worker returns into the business process. In that perception learning processes are directly linked to business processes.

Consequently, for knowledge work with its high variety often unforeseeable tasks and thus unforeseeable required knowledge [Maier, 05] learning support by current EL approaches are not ensured. Therefore, in the following the requirements from KM at learning support as well as the current situation in EL are regarded. Building on that a model for supporting learning in knowledge work is introduced. In

the following chapters the proposed approach is to use application profiles for describing the resources and arranging learning and knowledge resources to learning arrangements are described.

Learning support of knowledge work

From a KM perspective competencies and general knowledge applicable in numerous situations are highly relevant. But because of the scarce time in the daily work life, long-lasting learning processes brokering general knowledge appear unworkable. The prior transfer of the required knowledge seems not practicable, because on the one hand the demand of knowledge work is mostly unforeseeable and with a high variance [Maier, 05] and on the other hand the transfer of all potentially valuable knowledge is very time-consuming and expensive.

EL is understood as learning, a continuous, life-long process resulting from acting in situations [Brown, 89], that is supported by information and communication technology. Typically, learning material is offered as web-based training (WBT), which can be characterized either as monolithic, static courses or as more or less pre-defined compositions of learning objects (LO). The result is in both cases a static WBT that should fit all knowledge workers in all situations, irrespective of their personal attributes and that of their current context. If the knowledge worker's context and needs are not considered, motivation, success and acceptance may decline [Dagger, 05]. In order to solve this problem, it is necessary to consider the concrete situation of the knowledge worker and to compose resources during run-time into learning arrangements. Therefore, a detailed and processable description of the situation must be available. Building on the theory of task-technology fit [Goodhue, 95], a better fit between the knowledge worker's situation (task) and the proposed WBT (technology) should improve knowledge worker's performance.

Generally, it seems improbable that the predefined, static and bulky courses of current EL can ensure the fit to the knowledge worker's current task. However, none of the current approaches clearly focuses on considering typical situations in which employees encounter an opportunity to learn which is at the centre of just-in-time KM [Davenport, 02], workplace learning [Illeris, 03] or on-demand KM [Sampson, 02]. Here, knowledge services are triggered by a situation in which the user switches to a learning-oriented action, in this case knowledge elements are selected, composed and delivered considering as much context as possible [Maier, 07b]. Educational Adaptive Hypermedia Systems (EAHS) are an established approach to adapt learning material to the learners' situation. But in contrast to the proposed approach it strongly focuses on learners characteristics, like previous knowledge or learning styles, and not at work situations, primary focused here.

In the rest of this paper we propose an approach for on-demand EL supporting knowledge work considering requirements from KM and EL depicted in figure 1. Despite the fact that knowledge work and KM requires more flexible approaches for learning, traditional solutions developed by course authors [1] are still necessary. Didactically refined courses can be used to transfer more general knowledge and competencies identified as generally applicable in a multiplicity of business processes by learning processes before [2]. In the case of insufficient knowledge in the business process [3] situation-oriented arrangements are assembled

by resources from the organizational knowledge base in order to compensate the knowledge gap. Therefore, the organizations whole knowledge base (including external material) should be used. Thus, knowledge elements in different stages of maturity, like lessons learned, best practice or patents, [Maier, 07a], as well as contact information, like blue or yellow pages for proposing people helpful for the problem can be considered. By system-led adaptivity [4] a set of knowledge elements appropriate for the knowledge workers specific situation will be arranged [5]. In contrast to the traditional EL the composed course is not didactically designed or prepared. Finally, the knowledge worker can adopt the codified knowledge into the current work situation and apply in the business process [6].

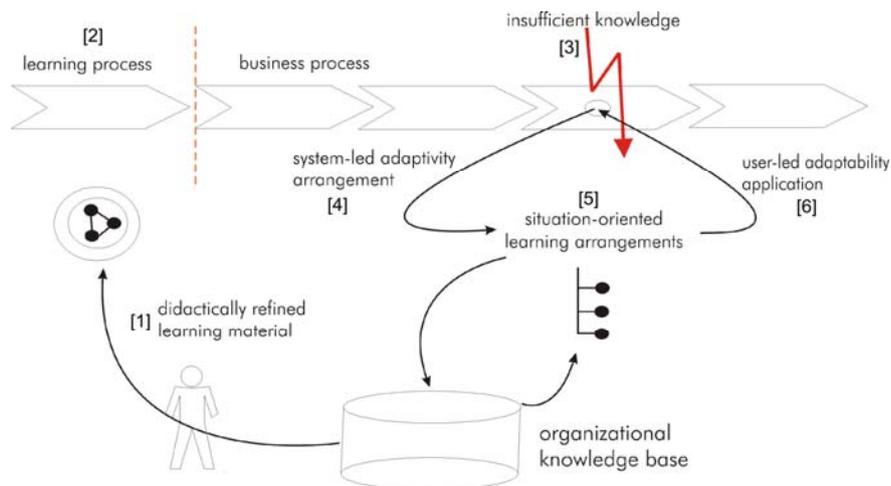


Figure 1: learning support for knowledge work

Application profiles to support on-demand e-Learning

In the last couple years, many organizations have applied concepts of business process re-engineering and numerous methods and techniques to support business process modeling have been proposed in the literature. Recently, a number of authors have proposed extensions to business process modeling techniques that model (some of the) specifics of KM (for a comparison see [Maier, 2007c]). Main extensions are on one hand additional object types, e.g., knowledge object, i.e. topics of interest, documented knowledge, employees, and skills, and on the other additional model types, e.g., knowledge structure- or communication diagram and knowledge map.

Even though the added concepts describe a portion of the context of knowledge work, they are not suited to model the often unstructured and creative learning practices in knowledge work and particularly their link to business processes (for a detailed comparison see [Maier, 05]). To overcome this obstacle [Maier, 07b] proposed an approach for on-demand EL based on the concept of activity theory. The activity-theoretic perspective [Engström, 87], [Blackler, 95] focuses creative, dynamic, and communication-intensive tasks, unstructured problems, membership in communities, self-organizing teams and demand for learning. In contrast to the clearly

defined sequence of events and functions, there is no predetermined flow of knowledge-oriented actions. Nevertheless, deriving a learning need from the mentioned extended modeling approaches seems also realizable.

In order to reuse knowledge elements from the organisational knowledge base (see figure 1) contextual information, i.e. metadata, represented in a semantically described specification is needed [Pawlowski, 06]. For LO the IEEE LOM [IEEE, 02] is regarded as the dominant standard in the field [Duval, 04], the standard is rather extensive standard that covers many of relevant aspects in order to mark up any LO. Depending on the intention of the use of the LOs and a cost/benefit analysis the set of metadata chosen can either be local to the organisation or from an established standard, in order to prepare for interoperability with other organisations. Some of the common problems with using a standard in practice are that they are sometimes very extensive and a lot of the elements are never used, even if the standard chosen is extensive it might not have enough elements to cover the need to describe a LO needed within an organisation [Godby, 04]. One solution to both apply to standards and describing the resources the way needed is to use application profiles [Heery, 00]. They are defined as: "schemas which consists of data elements drawn from one or more name-spaces, combined together by implementers, and optimised for a particular local application" [Heery, 00]. The application profile can furthermore restrict each of it elements further than it was originally intended, however the opposite is however not possible in order to be consistent with the standard. This means that an organisation can choose which metadata elements to use from one or more standards and also from locally developed used elements.

Some guidelines in order to create an application profile are given in [Duval, 06] where the starting point is to analyze organizational requirements. With the requirements it is also recommended to start from a certain standard that covers most of the needs and start from those elements of the standard. From there it is possible to further make the selection of appropriate metadata elements from other standards.

In [Maier, 07b] the requirements for a knowledge worker in a business process has been analysed and the paper further discusses what metadata elements from that standard can be used for different contexts aspects for a knowledge worker in a business process. The context dimensions taken into consideration are: process, person, group, product, location, time, technology and they are mapped against the elements of IEEE LOM.

In [Maier, 07b] it is concluded that IEEE LOM can generally be mapped to these context dimensions. However, In organization specific parts and in the description of media types, usage descriptions and educational descriptions is not totally covered. The solution is to extend these description by using application profiles with IEEE LOM as base schema. An example of an application profile for Learning Objects in form of video or audio content would use the IEEE LOM in combination with elements from the MPEG-7 standard, that is used to describe the technical aspects of video and audio content.

Arranging Knowledge Elements for on-demand e-learning

In order to create situation-oriented learning arrangements for realizing on-demand EL arranging knowledge elements, is proposed in the following. Starting point for on-

demand learning systems is a learning need arise by situations in business processes. Learning activities are triggered by an opportunity in a business process in which an employee with a certain learner profile can, should or must learn in order to be able to complete the present task. A learning system then delivers knowledge elements according to the worker's situation [Maier, 07b]. The situation can be described by the seven context dimensions mentioned in previous chapter, and the description is derived from basic systems, like ERP Systems or learning management systems. Table 1 presents the seven context dimensions with a short description of what they are and also how they can be found and gathered from the basic systems.

context dimensions	description	direct takeover of elements	applying rules or ontologies to gain usable metadata
process	sequence of tasks	- parts of the process description like keywords	- attributes from modelling language → required knowledge - granularity level from roles
person	individual engagement in the learning process	- entered keywords - input from defined input fields	- personal attributes from learner profiles - derive metadata from user behaviour, e.g. click stream data - user feedback to calibrate the deriving process
group	teams or communities the knowledge worker is engaged in	- elements from group description	- evaluate system usage of the other group members in similar situations
product	electronic resources which the knowledge worker use	-non-varying attributes from the metadata of content in use	- analyzing relationships defined in the metadata to identify typical attributes that fits - analyze the LO history to find typical successors
location	geographic location, points of interest	- predefined points of interests directly entered by the user, e.g. by using select lists	- derive metadata from points of interest (assign points to geographic coordinates) , e.g. keywords, degree of interruption
time	available time, current time	- information directly entered by the user	- deriving from schedules or process descriptions
technology	technical attributes of devices and applications	- technical specifications from the devices stored in device profiles, e.g. (CC/PP) model	- derive technical requirements from the device information

Table 1: deriving metadata from context dimensions

Providing situation-oriented learning arrangements can be realized in the following steps. Firstly, context data needs to be gathered which can be implemented by sensor services that provide context from basic systems such as repositories, ERP or HR systems. Those systems containing a huge amount of data about resources and their usage, the knowledge worker itself and his engagement with the systems.

Because of the heterogeneous data formats used in the multifaceted amount of basic systems and this raw data represents isolated views on the situation they must be aggregated, filtered and related in a separate consolidation process for describing the entire situation and to ensure processing in the next levels. For example, raw data

in form of a geographic coordinate must be linked to a point of interest, the description of which in form of metadata can be processed in the next levels. The methods depicted in Table 1 can be used in order gain valuable metadata stored in application profiles used in the organisation. After preparing the input data, the context is available as processable metadata stored with an application profile schema. This data is the basis for automatic query generation. The generated query is used for querying one or more repositories, in which knowledge resources and their metadata are stored. [Maier, 07b] showed that descriptions of the proposed context dimensions (see Table 1) match to metadata elements from existing metadata standards, e.g. IEEE LOM. Thus selection of suitable resources by using context information and practices described in table 1 is enabled.

A way to structure and support the search of knowledge elements would be to attach each of them to a learning process, this process would in turn correspond to a "learning need" that the knowledge worker has in the business process. In order to find the "learning need" for a knowledge worker, i.e. what a knowledge worker needs to learn in order to fulfil the part of the business process that are currently lacking. To derive this in an automatic way the description of the business process had to include roles in the process with some kind of competency description. It would also require a competency description of the knowledge worker, from those descriptions a learning need will be derived by comparing the two descriptions, the competency gap is then the learning need. To automatically find the learning need is rather complex unless all information already exist in advance. To manually find out a learning need it would first need to be recognized by the knowledge worker that such a need exist. This learning process needs to be found by the knowledge worker and from a selected learning process the attached LOs can be selected with the context dimensions taken into consideration in order to further personalise the selection.

In order to be interoperable over several repositories some kind of common infrastructure between the systems are needed. A way of searching for resources that is quite common in the library world is the Z39.50-protocol standard [Hammer, 96]. The protocol is client/server based and both sides have to implement a Z39.50 interface. Furthermore it needs to be decided what kind of data that can be searched for. Z39.50 can be used for non-semantic described resources, traditionally stored in databases or document management systems [Hammer, 96]. In the case of semantically described resources, e.g. described in RDF, searching can be made quite easy by using a query language like SPARQL specially designed for RDF.

The outcome of the query is a set of metadata describing knowledge elements that fulfil the criteria specified in the query. Therefore, the corresponding knowledge elements fit in the knowledge worker's current situation to a certain degree. Knowledge elements are selected from the return set and arranged into learning chunks. The size of the learning arrangements can vary from atomic knowledge elements to complex sets.

The decision is placed by a personal software agent or by a rule-based approach based on a fitness function as benchmark for evaluation [Goodhue, 95]. Finally, the resources identifiers are used to retrieve them from the repository and to present them to the knowledge worker.

The endpoint of a so-defined learning activity would not be passing a test, but completing the original task in the business process. By adopting the absorbed

knowledge into the current work situation in a business process and applying the knowledge in order to solve a certain task the on-demand learning process is finished and the knowledge worker turns back into the business process.

Conclusion and outlook

The paper identifies the requirements of learning support for knowledge work and presented an approach to consider these requirements. The need for acquiring new knowledge arises directly from acting in business process. Because of the high variance and unstructured processes of knowledge work [Maier, 05] the demand for knowledge cannot be identified before running the business process. Therefore, a flexible, situation-oriented reaction of a demand for knowledge arising from acting in a business process is necessary (see figure 1).

In order to ensure this support an approach for on-demand EL was proposed. What needs to be taken into consideration when searching for and arranging knowledge elements are the context-dimension as described in the paper. This information needs to be gathered and consolidated, perhaps from different systems in order to detect the learning need and afterwards query repositories in order to receive a personalised set of knowledge elements.

Resource descriptions with the LOM standard fits with the context dimensions [Maier, 07b], but could be extended with the help of application profiles in order to make a better description for a particular organisation.

Problematically are interoperability aspects between different metadata descriptions. The idea behind application profiles is rather simple, but when implementing such a solution problems can occur that is caused by differences in the abstract model behind different metadata standards. It is usually possible to be interoperable with one standard when using several in an application profile, but that is not really what is wanted. This problem is described in more detail in [Nilsson, 05].

In further research empirical studies to evaluate proper application profiles and a prototypical implementation of the presented approach are planned. Interesting questions here would be how to derive the necessary information needed to describe the seven contexts described here and then also how to keep it in a format that would be interoperable with each other.

References

- [Bach, 00] V. Bach, H. Österle (Eds.), Customer Relationship Management in der Praxis. Erfolgreiche Wege zu kundenzentrierten Lösungen. Springer, Berlin 2000.
- [Blackler, 95] F. Blackler, Knowledge, Knowledge Work and Organizations: An Overview and Interpretation. Organization Studies 16 (1995) 6, p. 1021-1046.
- [Brown, 89] J. S. Brown, A., Collins, P. Duguid, Situated Cognition and the Culture of Learning. Educational Researcher 18 (1989) 1, 32-42.
- [Dagger, 05] D. Dagger, V. Wade, O. Conlan, Personalisation for All: Making Adaptive Course Composition Easy. In: Journal of Educational Technology&Society, Vol. 8, Issue 3, pp. 9-25.

- [Davenport, 02] T. Davenport, J. Glaser, Just-in-Time Delivery Comes to Knowledge Management. *Harvard Business Review*, 80(7), p. 5-9.
- [Duval, 04] E. Duval, Learning Technology Standardization: Making Sense of it All. *Computer Science and Information Systems*, 1 (2004) 1, 33-43.
- [Duval, 06] E. Duval, N. Smith, M. Van Coillie, Application Profiles for Learning, *icalt*, pp. 242-246, Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06), 2006.
- [Engström, 87] Y. Engeström, Learning by Expanding: An Activity-theoretical Approach to Developmental Research. *Orianta-Konsultit Oy*, Helsinki 1987.
- [Godby, 04] C. J. Goodby, What Do Application Profiles Reveal about the Learning Object Metadata Standard?, *Ariadne* 41, UKOLN, UK, October 2004, <http://www.ariadne.ac.uk/issue41/godby/>
- [Goodhue, 95] D. L. Goodhue, R. L. Thompson, Task-technology fit and individual performance. *MIS Quarterly* 19(1995) 2, 213 - 237.
- [Hammer, 96] S. Hammer, J. Fayaro, Z39.50 and the World Wide Web, *D-Lib Magazine*, March 1996.
- [Heery, 00] R. Heery, M. Patel (2000). Application profiles: mixing and matching metadata schemas, *Ariadne*, Issue 25, 24-Sep-2000. <http://www.ariadne.ac.uk/issue25/app-profiles/intro.html>.
- [IEEE, 02] IEEE, Draft Standard for Learning Object Metadata (LOM). New York. URL: http://tsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf.
- [Illeris, 03] K. Illeris, Workplace Learning and Learning Theory. *Journal of Workplace Learning*. Vol.15, No 4.
- [Maier, 05] R. Maier, Modeling Knowledge Work for the Design of Knowledge Infrastructures, in: *Journal of Universal Computer Science*, Vol. 11, No. 4, 2005, 429-451.
- [Maier, 07a] R. Maier, A. Schmidt, Characterizing Knowledge Maturing: A Conceptual Model Integrating E-Learning and Knowledge Management. 4th Conference of Professional Knowledge Management (WM07), Potsdam, 325-333.
- [Maier, 07b] R. Maier, S. Thalmann, Describing Learning Objects for Situation- Oriented Knowledge Management Applications, 4th Conference of Professional Knowledge Management (WM07), Potsdam, 343-351.
- [Maier, 07c] R. Maier, *Knowledge Management Systems*, 3th Edition, Springer, Berlin, 2007.
- [Nilsson, 05] M. Nilsson, P. Johnston, A. Naeve, A. Powell, The Future of Learning Object Metadata Interoperability. In A. Koohang (Ed.), *Principles and Practices of the Effective Use of Learning Objects*, Informing Science Press.
- [Pawlowski, 06] J. M., Pawlowski, M. Bick, Managing & re-using didactical expertise: The Didactical Object Model. *Educational Technology & Society*, 9 (1), 2006, 84-96.
- [Sampson, 02] D. Sampson, C. Karagiannidis, A. Schenone, F. Cardinali, Knowledge-on-Demand in e-Learning and e-Working Settings, *Educational Technology & Society Journal*, Special Issue on Integrating Technology into Learning and Working, vol. 5(2), pp. 107-112, April 2002.

CAM in Process Execution

Martin Wolpers

(Katholieke Universiteit Leuven, Belgium
martin.wolpers@cs.kuleuven.be)

Gunnar Martin

(Deutsches Forschungszentrum für Künstliche Intelligenz, Germany
gunnar.martin@iwi.dfki.de)

Erik Duval

(Katholieke Universiteit Leuven, Belgium
erik.duval@cs.kuleuven.be)

Abstract: Existing business process, learning and knowledge management infrastructures work quite well in their specific context. Despite this, they are barely able to present any precise information about the usage of knowledge in certain working conditions, occurring learning needs or the identification of competency gaps and the detection of appropriate learning objects to fulfil the needs. Receiving this data from one single source is necessary to improve the individual and the organizational knowledge base continuously. Thus, the employees' qualification as well as the enhancement of their competencies constitutes an important precondition for an effective and efficient business process execution, the accomplishment of change management and in terms of "Time-to-Competency" their ability to anticipate cause-and-effect relations of process and market changes better and faster. To address the identification and provision of adequate (formal and especially informal) learning content in weakly structured or agile workflow environments, contextualized attention metadata can be useful. The attention metadata approach facilitates the improved collection of recent usage data across the various IT-system boundaries and applications used in different working surroundings. Consequently, it represents information about each session a user attends in the IT environment.

Keywords: business process management, workflow management, contextualized attention metadata, process performance measurement

Categories:

1 Introduction

As most organizations using integrated business ICT architectures, the integration of learning and enterprise environments must focus on the interface of business applications such as enterprise resource planning (ERP), supply chain management (SCM), customer relationship management (CRM) and business intelligence (BI) tools with workplace learning infrastructures [Bullinger 2006; Wahlster 2006]. Such integration has not been fully implemented today. Several (model-driven) approaches exist that strive to complement the loosely coupled business process resp. workflow execution with fitting – formal and/or informal – trainings to address the learning at

the workplace (e.g. EU/IST funded projects PROLIX and APOSDLE). Nevertheless, existing ICT systems and their embedded process model logic provide only incomplete information about the usage of knowledge in agile working conditions and its allocation. Furthermore, today's business information systems do not show the correlation of knowledge to the individual user nor a satisfactory definition of requirements for the development of learning and knowledge supporting actions. On the other hand, such information are necessary to ensure an efficient process execution, the implementation of change management strategies and the realization of competitive advantages that base on the individual employee [Levy, 2003]. This paper advocates an innovative approach that is targeted to improve the performance of workflow processes through the education of employees that is aligned to change management activities. Two possible sources of managerial inefficiency are identified in chapter 2. Chapter 3 focuses on capturing contextualized attention metadata that provides the ability to compare real with planned process execution. In chapter 4, we will detail how observations can be used to increase efficiency of the workflow execution by outlining its influence on (key) performance management. Chapter 5 on summary and future work concludes the paper.

2 Agility of business processes

A business process provides the context information necessary to identify learning needs and design matching course material that is "sensemaking" for business and individual learning goals [Weick 1995]. The process lifecycle and its major phases (1.) process design, (2.) process implementation and (3.) process control serves as conceptual baseline to add business weight to the most innovative learning infrastructures and thus make them economically sustainable. A business process mainly focuses on functions which have to be carried out in order to create an output that helps to achieve (strategic) goals or is steered by these goals indirectly. Therefore, the business process functions predefine the relevant information and supporting information systems as well as organizational and individual skill/competency references needed to operate in given business contexts. Based on these explicit pre-conditions set by the process knowledge, learning goals can be derived and trainings provided to ensure the (further) development of the employees' professional competencies and her process executing abilities (see figure 1). Recent observations indicate that most changes in workflows occur gradually. Participating functions in business processes usually are not completely exchanged but, instead, modified to suite the new workflows. Such modifications lead to activities that slightly change and therefore described as derivative functions f' of a function f (function f describe the original activity). A gradual shift in the function f means that some small part of an activity differs from the original task or execution procedure and leads to the definition of the derived function f' . Recently, business analysts from Aberdeen Group stated that the major problems of intrafunctional or procedural changes are the information of the employee involved resp. the provision of re-designed trainings [Aberdeen Group, 2006]. Often, the employees are not informed about the change within the set of activities they are responsible for. As a consequence, the employees still carry out function f instead of fulfilling the "new" or "revised" requirements of function f' correctly. Consequently, the employees have to

correct their work results to suite f' and the workflow in general is executed suboptimal in terms of quality and thus less efficient related to the dimensions of time and costs.

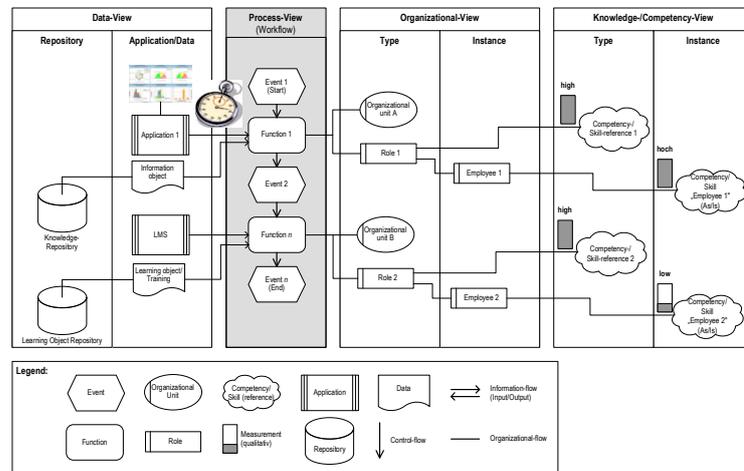


Figure 1: Business processes as learning context [Martin, 2007]

The process control identifies the shortcoming in process execution. The function and the employee can be identified on an instance level. Nevertheless, suitable educational measures cannot be taken because, so far, no measurement exists to pinpoint the exact causes of the malfunctioning workflow. Instead, the employee is given a rather broad education to address the whole of the derived function f' . The training is not efficient and probably not effective because, as f' is a derivate of f , the employee already knows most activities carried out in f' . The employees' real training need is embedded in the activities that change in the execution of f' compared to the execution of f while taking her experience and knowledge into account. So far, no ICT-tool exists to identify or derive training needs in a contextualized manner.

The approach outlined here will provide a suitable measurement to detect learning needs that occur through small functional or procedural changes in workflows. The proposed measurement is derived from the observations about the activities of the employee while carrying out the activities of functions f and f' . The comparison of the observations with the function (type) definition describes the difference between real activities and required activities which we call delta. The larger this gap is, the larger is the training need. Furthermore, the comparison describes which activities the employee carries out faulty. Therefore, the gap defines the learning needs in terms of the individual employee and also taking her knowledge and experience into account.

Additionally, the approach enables the identification of learning needs that arise through procedural changes in the workflow process. The observations also provide data about the communication behavior of the employees involved in a workflow process in the sense of cause-and-effect-relations. Therefore, the combination of the

observations about the employee with tacit knowledge about the workflow and function structure enables the deduction of learning needs.

Last but not least, observing the employees' user data during process execution and coupling the results with social web 2.0 technologies – such as blogs and wikis – will lead to the continuous increase of explicated organizational knowledge and goes beyond traditional workflow management systems. Based on the identification of the communication flow of employees, the employees are asked to explicate their knowledge, store and share it within the organizational knowledge base.

3 Observing activities

Observing the employee habits and data in her daily work enables the building of detailed user profiles. In the context of learning, providing the user with the appropriate material in the appropriate time and format requires collecting and mining usage information about activities in different systems she may use to execute a function in a workflow. This enables building specific profiles that are based on the user behaviour when executing functions, and therefore eliminate the need to rely on predefined stereotypes. As indicated before, such data will be used by process controlling systems to more correctly conclude on the user learning needs and requirements.

Hereby, detailed observations on user activities within functions of workflows and related systems are defined as Attention Metadata, because it is information about the attention a user gives to material she works with, for how long she works with it, when the action took place and what was her opinion about it (extracted from annotations, tags and rating information). The collected observations are generalized into behavioural patterns. Behavioural patterns describe in general how a user handles information, e.g. which activities she carries out with them. The comparison of behavioural patterns from various users allows clustering similar users. Based on such clusters, we expect to be able to precisely identify learning needs. The contextualized attention metadata approach utilizes to capture behavioural information of users in learning contexts that can be used to deal with the information overload in user centric ways. It relies on the schema for collecting contextualized attention metadata (CAMs) [Wolpers, 2007]. CAMs is designed to enable collecting observations about the attention users give to content and their specific contexts. CAMs enables the correlation of the observations, thus reflects the relationships that exists between the user, her context and the content he works with. CAM is captured through applications that interface with existing applications of daily use, e.g. the Microsoft Office Suite, Web Browsers, server logs, ARIS process platform, etc. Each of these tools continuously provides observations, thus generates a stream of information on the interactions of the user with the respective application. For here, the notion of context is broadened by describing the context through all additional information (e.g. information about parallel activities) available at the time the user activity within the business process is taking place. The user observations, captured in a generalized format, will be merged and processed thus providing the ability to contextualize each observation, e.g. by identifying which activities the user carried out simultaneously or within a short time-span. For example, with which keywords the user found relevant documents which he really wanted.

The observations of the employee directly correspond with the carried out activities. Either one activity represents one observation, or, more generally, a subset of observations represents an activity.

Using CAM, the delta (introduced in chapter 2) can be derived. CAM data provides the user's activities while performing function f as well as its derivate f' . Function f yields a set of observations o and respectively function f' yields a set of observations o' . The delta is derived by comparing the observations o' of the user with those that are to be expected based on the definition of function f' . Furthermore, by matching the set of observations o' of the user to those to be expected, the function activities are identified which the user executes erroneously. Consequently, the delta identifies the learning need of the employee and thus supports the improvement of business process performance and management of employee training.

4 Performance Management

Combined business and learning process control allows measuring the impact of learning and training on the execution of business processes within the organization. It is important to get feedback about the educational activities in order to improve adaptive, individualized learning material and learning design. Specifically, controlling learning processes from a business perspective provides insights on the impact of the training on the business process execution and the process performance. This means, employees who are already trained should perform better in their daily work. If not, the training is not sufficiently adjusted to the learning goals which correlate with business goals. Then the training content has to be (re-)engineered in order to optimize its effects on business process execution. In order to reveal the impact of workplace learning on the overall process, training measurements must be added to the already common key performance indicators of process monitoring system. This is where the true Return on Investment (ROI) is quantified and provides feedback for process improvement.

Business process-oriented learning design, learning material production and distribution to the learner create a comprehensive learning experience embedded in learner-oriented business process flows. This supports the understanding of transaction-oriented cause-effect-relations, which aligns individual and organizational learning goals. Flexible knowledge distribution on the basis of an improved technology support provides only relevant information and learning material to the employees. Therefore, it supports the reduction of time lags caused by competency deficiencies and enables faster readiness for business tasks, faster decision making and a shorter response time of involved stakeholders. Consequently, business execution performance and customer satisfaction increases. In order to couple business processes with learning processes, a SOA-based learning management framework offers controlled feedback channels for the automatic tailoring of learning experience to the single learner and the learner's performance controlling [Westerkamp 2006]. For controlling purposes the user performance during the learning process and afterwards is monitored. Based on the collected and analyzed data, respective adjustments to the learning process, the user profile and the learning process selection criteria are carried out. This feedback mechanism enables a flexible and continuous adaptation to evolving and arising learning needs, controlled by user

as well as company requirements [Martin, Leyking, Wolpers 2006]. To identify the users' working context, data from business process automation tools and workflow management systems (WfMS) can be analysed [Workflow Management Coalition 2003]. WfMS provide detailed description of the business process including organizational units, employees, applications etc., and information on the processing status, execution times and the context of the respective employee involved. The data provides a good quantitative basis for the process performance measurement and the cost/time controlling [Scheer 1998; zur Mühlen 2003], but does not contain any information about the usage of knowledge by the individual employee and his competencies.

4.1 Intra-functional changes

Consider a situation, where an employee works with an workflow process. The employee carries out a set of activities to fulfill function f . These activities yield observations o . Now, function f' as a derivate of f replaces f in the workflow chain. Function f' has a different set of activities but that overlaps with the set of activities of function f . As f' is only a derivate of f , the employee is not necessarily informed about the change in the function he has to carry out. He still carries out activities related to function f . In some cases, the employee is not clear which activities need to be carried out and thus mixes activities of both functions f and f' . Consequently, the observations about her activities are still those that result from the activities belonging to function f . The respective function f is identified through the observations that represent the activities.

A workflow management system is thus able to detect that the carried out activities do not match the expected activities. The observations therefore enable the detection of mistakes in carrying out the workflow process and enable the workflow management system to raise respective warnings.

Furthermore, the system is able to detect which activities in particular are carried out in a wrong way. Assuming that suitable learning materials are associated with all activities, and by having identified which activities are performed erroneously, the system is able to suggest the appropriate learning material to the employer. Note that the system will only suggest those where the employee makes mistakes.

In addition to identifying possible learning needs of one employee, the system is able to also deduce where systematic errors are included in the function definition. Comparing the observations of a number of employees carrying out the same function, differences become obvious. Such differences in observations enable a controller to identify the most effective and efficient set of activities. Eventually, this knowledge enables the controller to advise the employees and initiate respective learning processes.

4.2 Procedural changes

So far, we looked at the employee $e1$ that is directly involved in carrying out a function $f1$. Stated earlier, we said that a function $f1$ yields a result that is used as input for the subsequent function $f2$. $f2$ is also carried out by an employee $e2$ whose activities are observed.

If $f1$ is replaced by its derivate function $f1'$ and if the employee $e1$ carries out the respective activities erroneously, the result of $f1'$ is inadequate for function $f2$. In order to correct the result of $f1'$, the employee $e2$ contacts employee $e1$ to request the correction. Subsequently, the observations about the activities of employee $e2$ include representations of the communication activity as well as information of practical experiences of the employee and how they are utilized within the changed process.

The workflow management system is able to detect these variances by comparing the activities of employee $e2$ to those that define function $f2$. Measuring the variance enables the system to trigger corrective actions. For example, the system suggests to the employees $e1$ and $e2$ to include respective information about the erroneous result of $f1'$ and how to correct this into the corporate information system thus triggering the externalization of knowledge as described in the SECI model [Nonaka, Takeushi 1995] and outlined in figure 2. Today, a suitable tool to capture the information is based on Web 2.0 technologies like a corporate wiki or blog.

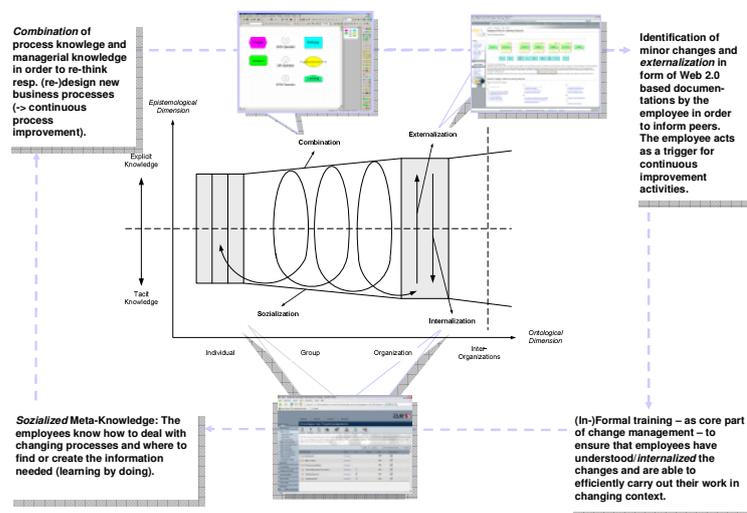


Figure 2: Orchestrating ICT-tools through the SECI-phases

Furthermore, the observations of activities of employee $e2$ in respect to function $f2$ provide evidence of the performance of the employee. For example, a controller detects that the time to carry function $f2$ needs 10min. instead of 2min. Based on the derived evidence, the controller decides to take appropriate corrective actions.

5 Conclusion and future work

We describe in this paper an approach that makes use of contextualized attention metadata for improving the performance of workflow process execution. The

observations derived from employees carrying out functions in a workflow process represent the activities carried out by the employees. Comparing these activities with those that define the respective function establishes a measurement for process execution performance. We argue that performance improvements are basically achieved through a better and more targeted training and education of the employees. In order to target the training better to the individual needs, the chosen approach enables the identification of respective knowledge gaps on the level of the activities carried out.

References

- [Aberdeen Group 2006] Aberdeen Group: Die größten Herausforderungen in Projekten für das Geschäftsprozessmanagement. In: Computer Zeitung 37 (2006), Nr. 38, S. 1
- [Bullinger 2006] Bullinger, Hans-Jörg: E-Learning darf keine Insellösung bleiben. In: Computer Zeitung 8 (2006), S. 6
- [Duval et.al. 2003] Duval, Erik and Hodgins, Wayne. Hencsey, Gusztáv, White, Bebo, Chen, Yih-Farn Robin, Kovacs, László, and Lawrence, Steve, eds. A LOM Research Agenda. In: Proceedings of the 12th international conference on World Wide Web 2003. Budapest, ACM.2003: 1-9.
- [Gartner, 2005] Gartner Inc. Prozesswissen ist gefragt: Welche IT-Dienstleistungen deutsche Unternehmen im kommenden Jahr einkaufen werden. Computer Zeitung 17/2005, 1. 2005.
- [Levy, 2003] Levy, Jonathan D. Creating a Change Management Knowledge Infrastructure. *IM Information Management & Consulting* 18(1), 12-16. 2003.
- [Najjar, 2003] Najjar Jajjar, Duval, Erik, Ternier, Stefaan, Neven, F.: Towards interoperable learning object repositories: the Ariadne experience. In: Proceedings of the IADIS International Conference WWW/Internet 2003. 1. (2003) 219-226
- [Martin, 2007] Martin, Gunnar: Geschäftsprozessorientiertes Learning Management: Konzept und Anwendung. Saarbrücken, Universität. 2007
- [Martin, Leyking, Wolpers, 2006] Martin, Gunnar; Leyking, Katrina; Wolpers, Martin: Business Process-driven Learning. In: Proceedings of the EU IST Africa 2006 Conference, Pretoria/South Africa., 2006
- [Nonaka, Takeushi, 1995] Nonaka, Ikujiro; Takeushi, Hirotaka: The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. Oxford, New York : Oxford University Press, 1995
- [Scheer, 1998] Scheer, August-Wilhelm: ARIS - business process frameworks. 2., completely rev. and enl. ed. Aufl. Berlin [u.a.] : Springer, 1998
- [Wahlster 2006] Wahlster, Wolfgang: Lernen der Zukunft - Möglichkeiten des onlinegestützten und mobilen Lernens. Saarbrücken : Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), 2006
- [Wolpers et.al. 2007] Martin Wolpers, Jehad Najjar, Katrien Verbert and Erik Duval, Tracking Actual Usage: the Attention Metadata Approach. *International Journal Educational Technology and Society* (ISSN: 1436-4522 (online) and 1176-3647 (print)), Special Issue on "Advanced Technologies for Life-Long Learning", 2007

An Experiment on Task Performance Forecasting Based on the Experience of Different Tasks^{*}

Jordi Olivella

(Technical University of Catalonia,
jorge.olivella@upc.edu)

Abstract: Performance in a task is influenced not only by the experience obtained in doing this task, but by how recent it is and by the experience obtained in doing similar tasks. Competence-Performance Approach is used as the theoretical framework. A modified version of Nembhard and Uzumeri learning and forgetting function is proposed to forecast performance by including the experience derived from other similar tasks. An experiment with voluntary students of telecommunication engineering was carried out. The tasks require assembly of electronic circuits. The results fitted well with the proposed model.

Keywords: Learning by doing, Competence-Performance Approach, Work organisation

Categories: J.4

1 Introduction

In production processes today the use of knowledge is increasingly present in all positions. The concept of knowledge worker [Pyöriä, 05] can be extended to almost all jobs. Learning, and particularly learning at work, becomes essential. Needs for current work and needs for learning can be contradictory: performance will be higher when a worker is assigned to a task in which he is experienced, while learning requires him to undertake new tasks. To manage this problem we have to be able to forecast performance for each level of experience. To evaluate the experience we will consider experience doing that task, the time that has passed since it was done and the experience obtained in other similar tasks.

Mechanical and mental tasks have to be distinguished, although both can be considered knowledge tasks. Mechanical tasks are those not involving decisions, while a task becomes more mental as more decisions have to be taken. It has been verified that learning and forgetting processes are simpler for mechanical tasks than for mental tasks [Arazi and Shtub, 97]. The analysis developed here could be applied to any kind of work; however in the experiment developed only mechanical tasks are performed, and thus at the present moment we have results only applicable to this kind of task.

The theoretical basis of our analysis is the Competence-Performance Approach [Korossy, 97]. Forecasting is done by using a modification of the learning and forgetting curve proposed by [Nembhard and Uzumeri, 00]. Previous works applying learning and forgetting models to solve work organisation problems were found in the literature. [McCreery et al., 04] considers a learning and forgetting model in

^{*} Supported by the Spanish MEC Project DPI2004-05797 co-financed by the FEDER

addressing a design problem while [Sayın and Karabatı, 07] includes one in a planning problem. Our contribution is to consider a wider range of factors and to carry out an experiment to test the model that we propose.

2 Theoretical basis

The dependence of performance of a task on experience obtained in performing other tasks is supported by the Competence-Performance Approach [Korossy, 97]. A distinction is made between competences – understood to be unobservable constructs representing “pieces of knowledge”– and performance of a task – a measurable outcome. To define the concept of knowledge structure two more elements have to be considered: relations of precedence and states of knowledge. ‘Precedence’ means that the acquisition of some competences precedes the acquisition of others. A state of knowledge is the situation of a worker in relation to competences, and is formed by the set of pieces of knowledge learned and known by the worker. Precedence rules determine which states will be feasible and which will not. A state of knowledge that includes a piece of knowledge and not all its precedents will be unfeasible; otherwise it will be feasible. A set of feasible states of knowledge is defined as a knowledge structure [Falmagne et al., 06].

Knowledge structure leads naturally to the concept of learning path – a sequence of learning experiences. It has been used to describe a situation where competences can be reached by following alternative sequences, as happens in the example we describe later. It has been used to design a lifetime learning path according to the characteristic of the person and his achievements [Karampiperis and Sampson, 06].

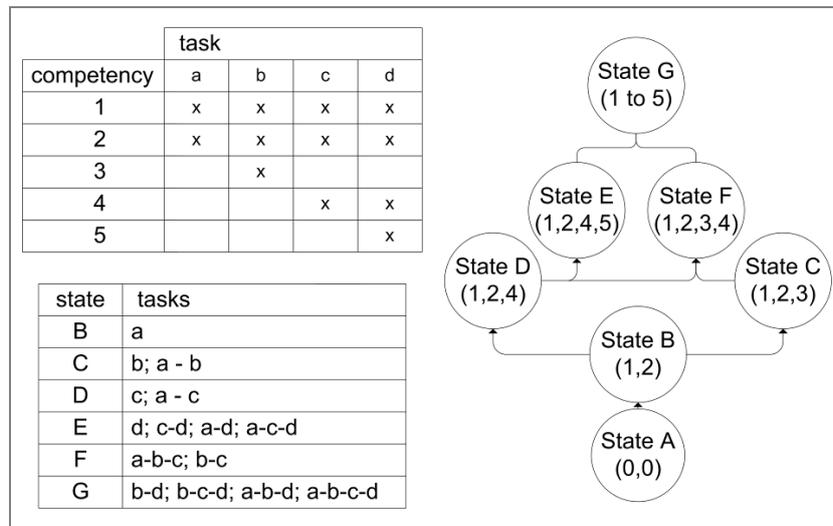


Figure 1.: Example of states of knowledge and precedence.

A simplistic example is illustrated in [Fig. 1] to clarify the model. We consider 4 tasks (a to d) each of which implies the use of several competences (1 to 5). This

supposes the existence of 7 states of knowledge (A-G). Effectively, a state consisting in competences 4 and 5, for example, is not feasible because it is not possible to acquire these competences without acquiring 1 and 2. The graphs show the paths that minimize the number of competences to be acquired at each step, and that are thought to be the best ones from the learning point of view. The other steps will be penalized by longer learning time and lower performance.

This theoretical approach has been applied to a methodology to assess underlying competences [Ley and Albert, 03]. Here we use it to define the states of knowledge and justify the differences between time and performance according to previous experience. In real cases the changes from one state to another may not be complete, but in any case the model is applicable because we can easily use states representing partial acquisition of competences.

To apply the Competence-Performance Approach to the performance at work requires us to consider a set of employees and a set of tasks where all tasks can be assigned to any worker. This situation corresponds to the case of a team or a category of workers, for example. Requirements for the job, such as personal traits and offline training, have to be tested previously. The worker's performance when doing a task is experience-dependent, and then it is a learning-by-doing approach, where competences are acquired by working. As competences are acquired, performance in the task will increase. When the new state of knowledge has only a limited number of new competences, the learning process is expected to be faster [Borthick et al., 03]. Then the time of learning – the time required to acquire the competence – and performance in the transition between states of knowledge will depend on the path followed.

In this situation the acquisition of competences depends on the work to be done and their assignment to the workers. The work to be done is not determined by the needs for learning but by the needs of production. Contradiction between these needs has been analysed in the case of apprentices, where learning is mandatory [Brooker and Butler, 07]. The sequence of work is thus conditioned to the needs of work. This limits the number of possible learning paths to use at each moment. Forecasting of performance and final knowledge goals can be used to support the decision between the remaining options.

3 Learning and forgetting model

We intend to forecast task performance according to experience of the task, the time that has passed since this experience and experience in other similar tasks. A learning and forgetting model modified to include experience in other similar tasks has to be defined.

A learning curve is an equation that shows the relation between experience and performance. Learning curves are based on the premise that people increase their performance in a task as they repeat it. A measure of the experience in doing the task has to be taken into account in any learning curve, while other variables can be included or not. Next the model that we propose is defined in three steps.

3.1 Model depending on repetitions

In a comparison of a total of 11 alternative learning curve models, a three parameter hyperbolic model is found to give the best approximation [Nembhard and Uzumeri, 00]. In [Eq. 2] y_x is a measure of the productivity with x the number of times that the task has been done, p the prior expertise in the task, k the limit of y_x as x approaches infinity – i.e. productivity when learning has been completed – and r a parameter that determines the slope of the curve.

$$y_x = k \cdot \left(\frac{x + p}{x + p + r} \right), \quad \text{with } y, p, k, x \geq 0; p + r > 0 \quad (1)$$

3.2 Model depending on repetitions and forgetting

The model has been extended by introducing a forgetting factor. [Shafer et al., 01] defines a measure of recency [Eq. 2] – i.e. how recent the experience is. Here x is the number of times that the worker did the task and t_i is the moment when the repetition number i of the task is performed – with t_0 the initial moment and t_x the last moment the worker did the task. The more recent the experience is, the closer R_x is to 1; the less recent it is, the closer it is to 0. The resulting learning and forgetting model has proved to be efficient when applying it to real data [Shafer et al., 01].

$$R_x = \frac{\sum_{i=1}^x (t_i - t_0)}{x \cdot (t_x - t_0)} \quad (2)$$

The factor R_x is then introduced to the model [Eq. 3]. The factor R_x makes the performance forecast higher the more recent the experience is. In addition a factor α is also introduced. Factor α depends on the degree to which the individual forgets the task.

$$y_x = k \cdot \left(\frac{x \cdot R_x^\alpha + p}{x \cdot R_x^\alpha + p + r} \right), \quad \text{with } \alpha, y, p, k, x \geq 0; p + r > 0 \quad (3)$$

3.3 Model depending on repetitions, forgetting and experience of other tasks

Next, the model is modified so that when we examine the performance of a task we consider the experience of other tasks. This effect is included by modifying the parameter p by the expression defined in [Eq. 4], where p_j represents the previous experience useful to do task j that has been gained by doing other tasks, j belongs to a group of indices of tasks K , $q_{j,j'}$ are parameters and $x_{j'}$ is the number of times that the task j' has been done. Over a certain threshold more experience of task j' does not provide any advantage for doing task j ; in this case x_j takes the value of this threshold.

$$p_j = q_{j,0} + \sum_{j' \in K, j' \neq j} q_{j,j'} \cdot x_{j'} \quad (4)$$

The parameter p is substituted in [Eq. 3] by the expression that has been defined, resulting in the model defined by [Eq. 5].

$$y_{j,x} = k \cdot \left(\frac{x_j \cdot R_x^\alpha + q_{j,0} + \sum_{j' \in K, j' \neq j} q_{j,j'} \cdot x_{j'}}{x_j \cdot R_x^\alpha + q_{j,0} + \sum_{j' \in K, j' \neq j} q_{j,j'} \cdot x_{j'} + r} \right), \tag{5}$$

with $\alpha, y, p, k, x \geq 0; p + r > 0$

4 Experiment

We developed an experiment to test the validity of the model to explain the performance in a mechanical task. The experiment consists of the assembly of electronic circuits by 12 volunteers, students of telecommunications engineering. The parameters of the model are estimated by using nonlinear regression, and performance forecasted by the model and real data are compared.

The electronic circuits were composed basically of amplifiers, resistors and transistors. Different tasks with different levels of difficulty were prepared. Finally three tasks A, B, C were selected. Task A consists of assembling an inverse amplifier. The material used is an electrical supplier, a functions generator and several electronic components such as resistors, amplifiers and wires. The subjects have to assembly the elements and verify the result with the use of an oscilloscope. Task B is very similar to A but the circuit is more complex. Finally, task C uses the same material as in the other experiments but an intermediate accumulator is used and a transistor is added.

3 - Col·locar les resistències tal i com s'indica en la figura següent

Resistència	Posició
10KΩ	25A - 25C
100KΩ	25C - 26C

4 - Afegir el filferro vermell que surt del jack vermell de +12V de la protoboard cap a la pota 4 del TL074 (sector 23C) i el filferro negre que surt del jack negre de -12V cap a la pota 11 TL074 (sector 23D). Connectar el filferro vermell del jack de +5V de la placa protoboard a la secció 25A (al costat de la resistència de 10kΩ).

Figure 2: Example of task standard.

A standard of work has been defined for the three tasks. All the volunteers must follow the same steps and operations so that their work is homogeneous and they obtain results independent of any individual's methods or ideas. Everybody thus does the same task in the same way. The standard contains all the information about the correct position of each component, the configuration of the appliance and the sequence of operations. Instructions are simple to understand for a

telecommunications engineering student. [Fig. 2] shows two steps of the standard for task A.

Sequence 1:	A A A A A A A A A B B B B B B B B B C C C C C C C C C
" 2:	B B B B B B B B B C C C C C C C C C C A A A A A A A A A
" 3:	C C C C C C C C C A A A A A A A A A B B B B B B B B B
" 4:	A A A B B B A A A B B B A A A B B B A A A B B B A A A B B B
" 5:	C C C A A A C C C A A A C C C A A A C C C A A A C C C A A A
" 6:	B B B C C C B B B C C C B B B C C C B B B C C C B B B C C C

Table 1: Sequences of task in the experiment

We defined six sequences of work, show in [Tab. 1]; these provide a sample of the possible situations. Two volunteers followed each sequence of work. The tasks were done by the subjects in more than one day, and not always on consecutive days. To take the important effect of these interruptions into account the factor of recency (R_x) has been calculated in days from the first day a task was done, whichever this task was.

	Parameter	Estimated value	Standard deviation	95% confidence interval	
				lower limit	upper limit
Task A	k	114.636	24.604	65.896	163.376
	$q_{A,0}$	6.456	2.832	.845	12.067
	$q_{A,B}$.110	.034	.042	.177
	$q_{A,C}$.010	.011	-.011	.032
	r	-6.456	2.653	-11.713	-1.200
	α	.548	.073	.403	.694
Task B	k	167.014	26.518	114.482	219.546
	$q_{B,0}$	4.949	2.079	.832	9.067
	$q_{B,A}$.241	.061	.120	.361
	$q_{B,C}$.000	.083	-.164	.164
	r	-4.949	1.866	-8.645	-1.253
	α	.287	.113	.062	.511
Task C	k	73.341	40.153	-6.202	152.884
	$q_{C,0}$	13.702	11.473	-9.026	36.430
	$q_{C,A}$.000	.054	-.107	.107
	$q_{C,B}$.138	.043	.053	.223
	r	-13.620	11.193	-35.793	8.554
	α	.613	.116	.382	.843

Table 2: Results of the experiment.

Prior to the experiment the volunteers answered a questionnaire about their personal situation and character (Situational Personality Questionnaire). With these test special situations that could affect the results were discarded.

The results were used to estimate, for each task (A, B and C), the parameters of the model defined above in [Eq. 6]. SPSS 14 nonlinear regression module was used. The complete results are show in [Tab. 2]. The parameters representing the influence of the experience on one task to the performance of another were basically those we expected. A and B are similar tasks and thus experience in A influences performance in B and conversely, as values of $q_{A,B}$ (influence of the experience of B when doing A) and $q_{B,A}$ show. The coefficients of determination have also been calculated and are show in [Tab. 3]. The model explains the results of the experiments to a good extend. The difference between the results when including ‘experience of other tasks’ and when not including it indicates the suitability of the inclusion of this factor in the model.

Experience in other tasks	Forecasting of performance of task		
	A	B	C
Included	78.24%	69.26%	62.60%
Non-included	66.72%	29.05%	38.53%

Table 3: Coefficients of determination (R^2).

5 Conclusions and further research

We have addressed the problem of forecasting performance in a task by considering the experience in this task, how recent this experience is, and the experience in similar tasks. An experiment with voluntary students gave us data for testing the efficiency of the model we proposed. The results obtained give us a reasonable confidence in the adequacy of our proposals and thus justify further work to corroborate the model. Data from more experiments and, especially, from real work situations has to be analysed to validate the model.

In addition, developments to take advantage of the model are proposed. The Competence-Performance Approach justifies why the performance of one task is influenced by the experience of other tasks. The consequence of this phenomenon is that the set of prior experiences affects the rate of improvement of the performance – i.e. the rate of learning. Learning at work is thus path-dependent. The order in which the tasks are done influences the performance and the process of learning. This can be used in planning to increase the rate of improvement of the performance, increase the volume of acquired skills at the end of one period, or both. We foresee, then, developing a planning model that takes into account the relations between experience and performance and that includes learning objectives.

References

- [Arazi and Shtub, 97] Arazi, Y. & Shtub, A.: "Learning and forgetting in mental and mechanical tasks: a comparative study "; *IIE Transactions*, 29, 9 (1997), 759-768.
- [Borthick et al., 03] Borthick, A. F., Jones, D. R., & Wakai, S.: "Designing Learning Experiences within Learners' Zones of Proximal Development (ZPDs): Enabling Collaborative Learning On-Site and Online"; *Journal of Information Systems*, 17, 1 (2003), 107-134.
- [Brooker and Butler, 07] Brooker, R. & Butler, J.: "The learning context within the workplace: as perceived by apprentices and their workplace trainers"; *Journal of Vocational Education & Training*, 49, 4 (1997), 487 - 510.
- [Falmagne et al., 06] Falmagne, J.C., Cosyn, E., Doignon, J.P., & Thiéry, N. The Assessment of Knowledge, in Theory and in Practice, Formal Concept Analysis: 61-79.
- [Karampiperis and Sampson, 06] Karampiperis, P. & Sampson, D.: "Adaptive Learning Objects Sequencing for Competence-Based Learning"; *Proc. ICALT'06*, IEEE Computer Society (2006), 136-138.
- [Korossy, 97] Korossy, K.: "Extending the theory of knowledge spaces: A competence-performance approach"; *Zeitschrift für Psychologie*, 205, 1 (1997), 53-82.
- [Ley and Albert, 03] Ley, T. & Albert, D.: "Identifying Employee Competencies in Dynamic Work Domains: Methodological Considerations and a Case Study"; *Journal of Universal Computer Science*, 9, 12 (2003), 1500-1518.
- [McCreery et al., 04] McCreery, J. K., Krajewski, L. J., Leong, G. K., & Ward, P. T.: "Performance implications of assembly work teams"; *Journal of Operations Management*, 22, 4 (2004), 387-412.
- [Nembhard and Uzumeri, 00] Nembhard, D. A. & Uzumeri, M. V.: "An individual-based description of learning within an organization"; *Engineering Management, IEEE Transactions on*, 47, 3 (2000), 370-378.
- [Pyöriä, 05] Pyöriä, P. : "The concept of knowledge work revisited "; *Journal of Knowledge Management* 9, 3 (2005), 116 - 127.
- [Sayın and Karabatı, 07] Sayın, S. & Karabatı, S.: "Assigning cross-trained workers to departments: A two-stage optimization model to maximize utility and skill improvement"; *European Journal of Operational Research*, 176, 3 (2007), 1643-1658.
- [Shafer et al., 01] Shafer, S.M., Nembhard, D.A., & Uzumeri, M.V.: "The Effects of Worker Learning, Forgetting, and Heterogeneity on Assembly Line Productivity"; *Management Science*, 47, 12 (2001), 1639-1653.

A Semantic-based Integrated Solution to Personnel and Learning Needs

Simona Colucci

(Politecnico di Bari, Bari, Italy,
s.colucci@poliba.it)

Tommaso Di Noia

(Politecnico di Bari, Bari, Italy,
t.dinoia@poliba.it)

Eugenio Di Sciascio

(Politecnico di Bari, Bari, Italy,
disciascio@poliba.it)

Francesco M. Donini

(Università della Tuscia, Viterbo, Italy,
donini@unitus.it)

Azzurra Ragone

(Politecnico di Bari, Bari, Italy,
a.ragone@poliba.it)

Abstract: In knowledge intensive companies intellectual capital assumes a crucial role in the organizational strategy and, as any other strategical asset, it needs to be scheduled to achieve sustainable competitive advantage. When the required knowledge is a resource available inside the company, its assignment represents a key success factor, which many research efforts are devoted to. On the other side, when the needed competence is unavailable within the company, training programs may be seen as methods to strengthen such a strategic asset. In this paper we show a semantic-based integrated system aimed at supporting both the assignment of available intellectual resources in three different multiplicity scenarios and the search for training programs ad-hoc composed to fill possible knowledge gaps.

Key Words: Skill management, E-learning, Knowledge Management Systems

Category: I.2.4, K.6.1, I.2.1

1 Introduction

The vision of knowledge intensive companies intuitively focuses on competencies human resources are endowed with. Any organization aiming at sustainable competitive advantage needs then to schedule and acquire competencies as any other asset. In such a manufacturing metaphor, internal personnel may be seen as the knowledge warehouse a company holds; when such a warehouse runs out

of stock, management may revert to outsourcing, which means hiring external personnel, or to a further supplying, which consists in making new competencies available in the knowledge warehouse. In both such solutions, the company needs to search for a knowledge provider able to fulfill the arising competency need. Even though it is straightforward thinking of knowledge providers in terms of workers, also resources like e-learning modules (learning objects) may be considered available to fit the need. In this paper we propose an integrated system implementing the approach designed in [?] for the solution of knowledge needs, either in terms of search for professional profiles or for personalized training programs. In order for the search and the composition process to be automated, the descriptions of the knowledge need, the available professional profiles and the learning objects to be composed are modeled according to the formalism of Description Logics(DLs)[?] in order to take advantage from some provided reasoning services useful in the processes our system supports. In particular, together with standard DLs services, non standard explanation services have been exploited to investigate on missing and conflicting competencies in knowledge provider descriptions w.r.t. formalized knowledge needs. An overview of DLs formalism and needed reasoning services is given in [?]. The use of DLs ensures also the full interoperability with Semantic Web [?] languages, as witnessed by the existence of a subset of OWL language¹, OWL-DL, developed to map DLs constructs in OWL.

The paper is organized according to the following outline: in next section the proposed system is detailed in all its components; in Section ?? system working mode is explained with the aid of an extended example. Related work on the subject and conclusions close the paper.

2 The Integrated System

The integrated solution we propose to solve personnel and learning needs implements the framework designed in [?] and grounds on the architecture shown in Figure ?. In the initial step, the user chooses the matching scenario (Arrow 1) and models according to a reference ontology the tasks to be satisfied by knowledge provider profiles available within the system repository. The following three multiplicity cases are allowed:

1. **Single one to one matching:** one knowledge need at a time is taken into account and assigned to only one knowledge provider through the algorithm *Assign* proposed in [?].
2. **Multiple one to one matching:** more than one knowledge need at a time is considered and each of them is assigned to only one knowledge provider among the set of available ones, according to the algorithm *MultipleAssign* proposed

¹ OWL. www.w3.org/TR/owl-features

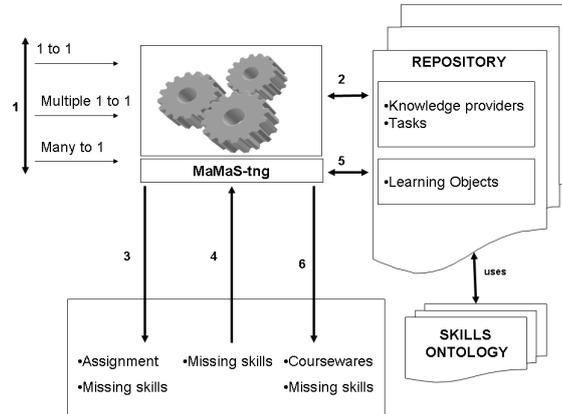


Figure 1: System architecture

in [?].

3. Many to one matching: one knowledge need at a time is taken into account and assigned to an ad-hoc created team of knowledge providers by applying the algorithm *TeamComposer* proposed in [?].

In all of the three multiplicity cases, the assignment is performed on the basis of the suitability of the knowledge provider descriptions to the task descriptions. Such a suitability is evaluated through a function U outlined in [?]. In a nutshell, U evaluates a numerical score for the similarity of knowledge provider to task descriptions by taking into account:

- the part H of knowledge need not explicitly provided by the analyzed knowledge provider. Such an explanation is provided by solving a Concept Abduction Problem [?], a non standard DL reasoning service
- the part G of the knowledge need to be given for the selection of the analyzed knowledge provider. Such an explanation is the result of a non standard problem by DL: Concept Contraction [?].

All the required tasks, $T_i \in D$ with $i = 1 \dots n$ and the knowledge provider profiles $P_j \in S$ with $j = 1 \dots m$ have to be formalized in a subset language of DLs, \mathcal{ALN} [?], by using the terminology of the skills ontology, independently on the chosen matching scenario. This requirement, related only to the current system implementation, is determined by the use of MaMaS-tng² as reasoning engine. To the best of our knowledge, MaMaS-tng is the only reasoning engine able to solve Concept Abduction and Concept Contraction problems. The system queries MaMaS-tng in order to search in the repository for knowledge

² **MA**tch**MA**ker**S**ervice is available at <http://dee227.poliba.it:8080/MAMAS-tng>.

providers able to satisfy the formalized tasks (Arrow 2) and returns an assignment proposal, together with the explanation about skills still missing in the assignee profile (Arrow 3). Such lacking knowledge represents the learning need to be covered (Arrow 4) by a learning path automatically returned (Arrow 6) by composing the learning objects advertised in the repository as \mathcal{ALN} descriptions (Arrow 5). An explanation about skills still missing after learning path fruition is also returned to the system user.

A framework for the automatic composition of learning objects into learning paths, denoted as coursewares, was proposed in [?]. Such a framework implements an algorithm, *teacher*, automatically computing personalized coursewares and the part of learning need still uncovered after courseware fruition. Also in the composition process, Concept Abduction is exploited to get the explanation about lacking knowledge. On the contrary, we do not need here to solve any Concept Contraction Problem, because the situation in which the learning need is incompatible with a learning object description and needs to be contracted may not ensue because of the nature of knowledge itself: knowledge about a given skill is always compatible with any other sort of knowledge. Both the learning objects and the learning need are formalized in a *description* (pedex D) and a *background knowledge* (pedex BK) component: each learning object λ teaches the knowledge described by λ_D and requires the knowledge described by λ_{BK} ; the learning need ρ is a request for the knowledge described by ρ_D , starting from a knowledge background expressed by ρ_{BK} . Such a formalization causes the proposed courseware to be personalized: only the providers holding the required background knowledge may be asked to learn the competences detailed in the learning object description.

Our system supports user's decision also in choosing the learning process which requires the least effort for covering the learning need, given that different personalized processes are possible. In our opinion a completely automatic selection is not the most suitable solution in this phase, because several highly subjective choice factors have to be taken into account. Courseware complexity, evaluation of missing skills and of knowledge bonus learned after courseware fruition are possible choice factors. In particular the knowledge bonus is computed by solving a Concept Abduction Problem returning the part of the courseware description not explicitly formalized in the knowledge need. Instead of proposing an automatic learner selection our system presents an explanation facility for these three factors to the system user, so making available the whole information relevant for her decision. The final selection of the candidate is then up to the system user.

3 Working example

Imagine the scenario in which an user needs to solve a set of tasks, by assigning each of them to only one knowledge provider. She can ask the system for a multiple one to one assignment process. Imagine skill domain is described in the ontology shown in the following in its DL formalization:

ComputerScience \sqsubseteq Skill	Engineer \equiv \exists hasMasterDegree \sqcap	\forall hasMasterDegree.Engineering \sqcap
InternetUse \sqsubseteq ComputerScience	\exists advancedKnowledge \sqcap	\forall advancedKnowledge.Design \sqcap \exists basicKnowledge
InternetTechnologies \sqsubseteq ComputerScience	\forall basicKnowledge.(Mathematics \sqcap Physics)	
CGI \sqsubseteq ComputerScience	ManagerialEngineer \equiv Engineer \sqcap	\exists advancedKnowledge \sqcap
TCP/IP \sqsubseteq ComputerScience	\forall advancedKnowledge.ProcessManagement	
C++ \sqsubseteq OOP	LearningObject \equiv \exists BackgroundKnowledge \sqcap	\exists Description \sqcap
Java \sqsubseteq OOP		
HTML \sqsubseteq MarkupLanguages		
MarkupLanguages \sqsubseteq ComputerScience		
ClientServerProtocol \sqsubseteq ComputerScience		
InternetDevelopment \sqsubseteq ComputerScience		
WebDesigning \sqsubseteq InternetDevelopment		
TotalQualityManagement \sqsubseteq Skill		
VBScript \sqsubseteq ComputerScience		
Design \sqsubseteq Skill		
ComputerGraphics \sqsubseteq ComputerScience		
Engineering \sqsubseteq degree		

Let D be the set of knowledge requests composed by the tasks shown in the following together with their DL descriptions:

– Workers with basic knowledge of Internet Use and Markup Languages

$$T_1 = \exists \text{basicKnowledge} \sqcap \forall \text{basicKnowledge} . (\text{InternetUse} \sqcap \text{MarkupLanguages})$$

– Workers with advanced knowledge about Client Server Protocol and Process Management, expert in tools for Internet Development

$$T_2 = \exists \text{advancedKnowledge} \sqcap \forall \text{advancedKnowledge} . (\text{ClientServerProtocol} \sqcap \text{ProcessManagement}) \sqcap \exists \text{toolsKnowledge} \sqcap \forall \text{toolsKnowledge} . \text{InternetDevelopment}$$

– Graduated workers with more than three years of working experience and advanced knowledge about Total Quality Management, C++ and TCP/IP protocol

$$T_3 = \exists \text{advancedKnowledge} \sqcap \forall \text{advancedKnowledge} . (\text{TotalQualityManagement} \sqcap \text{C++} \sqcap \text{TCP/IP}) \sqcap \exists \text{hasMasterDegree} \sqcap \exists \text{hasExperience} \sqcap \forall \text{hasExperience} . (\geq 3 \text{ years})$$

Let now P be the set of available assignees composed by the following descriptions:

– Graduate, with less than 2 years of working experience and basic knowledge of Internet Use, Computer Graphics and HTML

$$P_1 = \exists \text{basicKnowledge} \sqcap \forall \text{basicKnowledge} . (\text{InternetUse} \sqcap \text{ComputerGraphics} \sqcap \text{HTML}) \sqcap \exists \text{hasMasterDegree} \sqcap \exists \text{hasExperience} \sqcap \forall \text{hasExperience} . (\leq 2 \text{ years})$$

– Managerial Engineer, with advanced knowledge about Internet Technologies and knowledge about Web Designing Tools

$$P_2 = \text{ManagerialEngineer} \sqcap \exists \text{advancedKnowledge} \sqcap \forall \text{advancedKnowledge} . \text{InternetTechnologies} \sqcap \exists \text{toolsKnowledge} \sqcap \forall \text{toolsKnowledge} . \text{WebDesigning}$$

– Engineer, with advanced knowledge about Total Quality Management, Client Server Protocol, VBScript and Object Oriented Programming

$$P_3 = \text{Engineer} \sqcap \exists \text{advancedKnowledge} \sqcap \forall \text{advancedKnowledge} . (\text{TotalQualityManagement} \sqcap \text{ClientServerProtocol} \sqcap \text{VBScript} \sqcap \text{OOP})$$

The matching process implements *MultipleAssign* [?] which performs an as-

signment problem [?] to minimize the objective function

$$Z = \sum_{i=1}^n \sum_{j=1}^n u_{ij} x_{ij}$$

where u_{ij} are the values of function U filling the suitability matrix in the following, if computed according to the close form proposed in [?]:

	T_1	T_2	T_3
P_1	0	1	0.88
P_2	0.6	0.11	0.6
P_3	0.6	0.56	0.4

Notice that x_{ij} are the decision variables such that in the solution $x_{ij} = 1$ if assignee i performs task j and $x_{ij} = 0$ otherwise. The algorithm returns the following assignment solution:

1. $x_{11} = 1$ ($T_1, P_1, H_{11} = \top, G_{11} = \top$)

P_1 represent a full match for T_1 according to the skills ontology ($H_{11} = \top$ means that no hypothesis is needed about knowledge not elicited in P_1 but required by T_1): even `MarkupLanguages` knowledge, apparently lacking, is in fact embedded in `HTML`. Moreover P_1 does not contradict T_1 , as shown by the result $G_{11} = \top$.

2. $x_{22} = 1$ ($T_2, P_2, H_{22} = \forall \text{advancedKnowledge.ClientServerProtocol}, G_{22} = \top$)

P_2 is compatible with T_2 according to the skills ontology: hypotheses H_{22} is anyway needed on the advanced knowledge about `ClientServerProtocol` while `InternetDevelopment` knowledge is embedded in `WebDesigning`. Moreover P_2 does not contradict T_2 , as shown by the result $G_{22} = \top$.

3. $x_{33} = 1$ ($T_3, P_3, H_{33} = \exists \text{hasExperience} \sqcap \forall \text{hasExperience}(\geq 3 \text{ years}) \sqcap$

$\forall \text{advancedKnowledge}(\text{C++} \sqcap \text{TCP/IP}), G_{33} = \top$)

P_3 is compatible with T_3 according to the skills ontology: hypotheses H_{33} have to be formulated on the work experience requirements and on advanced knowledge about `C++` and `TCP/IP`: in particular even if P_3 knows about `OOP`, her knowledge about `C++` can only be hypothesized because the second one is more specific than the first one. Moreover P_3 does not contradict T_3 , as shown by the result $G_{33} = \top$.

Both the second and the third solutions lead to a knowledge need which can be solved by the fruition of ad-hoc created coursewares. Respectively, such knowledge needs are represented as follows:

- $\rho_2 = \exists \text{BackgroundKnowledge} \sqcap \forall \text{BackgroundKnowledge}.P_2 \sqcap \exists \text{Description} \sqcap \forall \text{Description}.H_{22}$

- $\rho_3 = \exists \text{BackgroundKnowledge} \sqcap \forall \text{BackgroundKnowledge}.P_3 \sqcap \exists \text{Description} \sqcap \forall \text{Description}.H_{33}$

Suppose the following learning objects are available as knowledge providers in the system repository:

- *Java and C++ course. Requirements: Object Oriented Programming knowledge*

$\lambda^1 = \text{LearningObject} \sqcap \forall \text{BackgroundKnowledge}(\forall \text{advancedKnowledge.OOP})$

$\sqcap \forall \text{Description}(\forall \text{advancedKnowledge}(\text{C++} \sqcap \text{Java}))$

- *Client Server protocol course. Requirements: CGI knowledge*

$$\lambda^2 = \text{LearningObject} \sqcap \forall \text{BackgroundKnowledge} . (\forall \text{advancedKnowledge} . \text{CGI})$$

$$\sqcap \forall \text{Description} . (\forall \text{advancedKnowledge} . \text{ClientServerProtocol})$$

– CGI Course. Requirements: Internet Technologies Knowledge

$$\lambda^3 = \text{LearningObject} \sqcap \forall \text{BackgroundKnowledge} . (\forall \text{advancedKnowledge} . \text{InternetTechnologies})$$

$$\sqcap \forall \text{Description} . (\forall \text{advancedKnowledge} . \text{CGI})$$

By applying *teacher* the two learning solutions in the following are proposed.

Assignee	Λ	UncoveredKnowledge	KnowledgeBonus
P_2	$\{\lambda^3, \lambda^2\}$	\top	\top
P_3	$\{\lambda^1\}$	TCP/IP	Java

Assignee P_2 needs the courseware $\Lambda = \{\lambda^3, \lambda^2\}$ to fully cover her knowledge need ($UncoveredKnowledge = \top$); in particular the fruition of λ^3 makes the background knowledge needed to learn λ^2 available. No additional knowledge is gained by courseware fruition ($KnowledgeBonus = \top$). Assignee P_3 can only partially cover her knowledge need by learning λ^1 : knowledge about TCP/IP continues to be uncovered after the fruition, but on the other side a knowledge bonus about Java is gained.

By subjectively analyzing the two proposed learning solutions, a company management adopting the system we present may evaluate the opportunity of initiating one or both of the suggested learning paths.

4 Related Work

In recent years competence based management has been investigated under a variety of perspectives and several systems and approaches have been presented in literature to cope with both knowledge and learning management [?], even though the number of approaches integrating the two research field is small if compared with that of proposals about each one of them [?]. Among system for competence management, ontology based solutions to task assignment and team composition are proposed in [?, ?, ?, ?]. By reverting to learning management, an ontology based solution to learning paths composition grounding on non standard inference services provided by \mathcal{ALN} has been proposed in [?]. A surprisingly similar approach, exploiting a less expressive DL was proposed in [?].

In [?] and [?] two ontology based systems integrating competence management and e-learning are proposed.

5 Conclusions

We presented a semantic based system to support personnel and learning needs in an organizational context. The system performs the assignment of professional profiles to tasks in three different multiplicity cases and exploits non standard

inference services by DLs to investigate on possible knowledge gaps to be solved by the fruition of ad-hoc learning paths automatically composed. In order to keep its decisional support nature, the system is not fully automated but it leaves to the user the possibility of a manual choice in each selection process involving subjective factors.

References

- [Baader et al., 2002] Baader, F., Calvanese, D., Mc Guinness, D., Nardi, D., and Patel-Schneider, P., editors (2002). *The Description Logic Handbook*. Cambridge University Press.
- [Berners-Lee et al., 2001] Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web. *Scientific American*, 248(4):34–43.
- [Colucci et al., 2003] Colucci, S., Di Noia, T., Di Sciascio, E., Donini, F., Mongiello, M., and Mottola, M. (2003). A formal approach to ontology-based semantic match of skills descriptions. *Journal of Universal Computer Science (J.UCS)*, 9(12):1437–1454.
- [Colucci et al., 2004] Colucci, S., Di Noia, T., Di Sciascio, E., Donini, F., Mongiello, M., and Piscitelli, G. (2004). Semantic-based approach to task assignment of individual profiles. *Journal of Universal Computer Science*, 10(6):723–730.
- [Colucci et al., 2005a] Colucci, S., Di Noia, T., Di Sciascio, E., Donini, F., and Ragone, A. (2005a). Automated task-oriented team composition using description logics. In *5th International Conference on Knowledge Management I-Know 05*, pages 229–236. Springer.
- [Colucci et al., 2005b] Colucci, S., Di Noia, T., Di Sciascio, E., Donini, F., and Ragone, A. (2005b). Semantic-based automated composition of distributed learning objects for personalized e-learning. In *The Semantic Web: Research and Applications. 2nd European Semantic Web Conf.*, volume 3532 of *LNCIS*, pages 633–648. SV.
- [Colucci et al., 2006] Colucci, S., Di Noia, T., Di Sciascio, E., Donini, F., and Ragone, A. (2006). Integrated semantic-based composition of skills and learning needs in knowledge-intensive organizations. In *Competencies in Organizational E-Learning: Concepts and Tools*, pages 270–298. Idea Group Publishing.
- [Colucci et al., 2007] Colucci, S., Di Noia, T., Di Sciascio, E., Donini, F., and Ragone, A. (2007). Semantic-based skill management for automated task assignment and courseware composition. *Journal of Universal Computer Science (J.UCS)*. to appear.
- [Cormen et al., 1990] Cormen, T. H., Leiserson, C. E., and Rivest, R. L. (1990). *Introduction to Algorithms*. The Massachusetts Institute of Technology.
- [Draganidis et al., 2006] Draganidis, F., Chamopoulou, P., and Mentzas, G. (2006). An ontology based tool for competency management and learning paths. In *Proc. I-KNOW '06*. to appear.
- [Draganidis and Mentzas, 2006] Draganidis, F. and Mentzas, G. (2006). Competency based management: A review of systems and approaches. *Information Management and Computer Security*, 14(1):51–64.
- [Draganidis and Mentzas, 2007] Draganidis, F. and Mentzas, G. (2007). Ontology-based competency management for corporate e-learning. In *Competencies in Organizational E-Learning: Concepts and Tools*, pages 311–324. Idea Group Publishing.
- [Hefke and Stojanovic, 2004] Hefke, M. and Stojanovic, L. (2004). An ontology-based approach for competence bundling and composition of ad-hoc teams in an organization. In *Proc. I-KNOW '04*.
- [N. Karam and Meinel, 2007] N. Karam, S. L. and Meinel, C. (2007). Semantic composition of lecture subparts for a personalized e-learning. In *In proceedings of European Semantic Web Conference (ESWC), Innsbruck, Austria*, pages 716 – 728.

Wikis as a Technology Fostering Knowledge Maturing: What We Can Learn from Wikipedia

Simone Braun

(FZI Research Center for Information Technologies
Simone.Braun@fzi.de)

Andreas Schmidt

(FZI Research Center for Information Technologies
Andreas.Schmidt@fzi.de)

Abstract: The knowledge maturing theory opens an important macro perspective within the new paradigm of work-integrated learning. Especially wikis are interesting socio-technical systems to foster maturing activities by overcoming typical barriers. But so far, the theory has been mainly based on anecdotal evidence collected from various projects and observations. In this paper, we want to present the results of a qualitative and quantitative study of Wikipedia with respect to maturing phenomena, identifying instruments and measures indicating maturity. The findings, generalized to enterprise wikis, open the perspective on what promotes maturing on a method level and what can be used to spot maturing processes on a technology level.

Keywords: wiki, knowledge maturing, work-integrated learning

Categories: H.3.5, H.3.7, H.4.0, J.4

1 Introduction

Current research approaches to supporting workplace learning have shifted their attention from provision of learning material towards supporting the active contribution of the learner and the collaboration with others. Working and learning are increasingly conceived as embedded into, interwoven with, and even indistinguishable from everyday work processes. On the one hand, learning support is increasingly seen as work performance support; on the other hand, collaborative activities are increasingly understood as collaborative learning activities.

One important element of this new perspective is that individual learning processes are no longer seen as isolated one-way processes, but rather as interlinked with others and loosely coordinated by organizational goals. The theory of the knowledge maturing process (as introduced in [Schmidt 05]) provides a conceptual model for this interlinkage and the corresponding change in the quality of involved knowledge. The model uses the metaphor of maturing and is structured into several phases: from (1) emergence of ideas, via (2) distribution in community, (3) formalization up to (4) ad hoc training and (5) formal training. Along this process, the characteristics of knowledge change: implicit contextualization decreases, explicit linkage increases as well as hardness, legitimation, and teachability [Maier & Schmidt 07]. Furthermore, the theory identifies typical barriers in maturing, such as between

phases (2) and (3) where knowledge usually is detached from its originator and (3) and (4) as the typical barrier between knowledge management and e-learning.

Wikis have been identified in [Maier & Schmidt 07] as a promising technology fostering knowledge maturing because it allows for the coexistence of and smooth transition between different maturing phases. They can capture initial ideas in a similar way as didactically prepared learning material, and there are no technical barriers between the different maturity phases. Maturing can emerge as part of the social interaction, producing explicit or implicit conventions.

This sounds plausible from everyday observation, but in this paper, we want to go one step further and actually validate this view, i.e., the hypotheses that (1) knowledge maturing as described in [Maier & Schmidt 07] actually takes place in wikis and (2) that socio-technical wiki systems have brought forth instruments overcoming barriers in the maturing processes. In order to be able to rely on a large body of data, we have conducted a qualitative (which problems and mechanisms for maturing have emerged) and quantitative study (which measures can be used to describe maturing) on Wikipedia (primarily focused on the German Wikipedia), which we consider the largest social experiment on knowledge maturing. We will also analyze the transferability of these results to enterprise wiki systems.

2 Qualitative Analysis of Wikipedia: Three Layers of Maturing

Our initial step to analyze Wikipedia in terms of knowledge maturing was to identify instruments within Wikipedia that affect maturing. The first insight gained from that qualitative analysis was that the majority of such instruments is not technically implemented at all in the underlying MediaWiki system or has been only successively implemented over time as a response to pressing needs to codify social rules. That is also reflected in a varying list of (more or less) official principles and guidelines. The second insight was that we need a more differentiated view on maturing in order to describe and relate the instruments with the knowledge maturing theory. The phenomenon of maturing can be divided into three layers:

Artefact Layer. The most obvious layer is concerned with the artefacts that are produced in the course of communicating with others about certain knowledge elements. In the case of wikis, these artefacts are the wiki pages. The most obvious instruments of Wikipedia on the artefact level are the predicates that are given to certain articles, which represent the legitimation by the Wikipedia community. They indicate their level of quality like good article (“lesenswert”), featured article (“exzellente Artikel”), but also “stub” and “needs review”. Wikipedia has also developed definitions for the level of quality, which correlate with the notion of maturity. Closely linked with these predicates are candidate lists implementing a limited and very loose form of transition workflow. These lists encourage maturing of articles by creating awareness for them: typically other authors become aware of the existence and the need for improvement of these articles.

Knowledge Layer. As artefacts are usually only a medium for communicating knowledge between individuals and communities, these do not accurately reflect the individual knowledge. Even when knowledge is mature, artefacts like Wikipedia articles usually start at an immature status. This does not deny the coupling between artefact and knowledge layer: producing mature artefacts requires a certain level of

maturity on the knowledge layer, and this production process usually implies a construction process on the knowledge layer. Wikipedia offers also instruments on the knowledge layer: the use of categories to organize articles, links among articles that are not directly tied to the occurrence of the article name, and disambiguation pages. All of these instruments contribute to improving the decontextualization process (and the explicit linking as an inverse criterion): knowledge is organized from a more global perspective, abstracting away the creation context.

Social Layer. As knowledge maturing is an inherently collaborative activity, it also requires social “meta competencies”, i.e., a socialization process in which conventions and adherence to rules are learnt. Besides the “Wikiquote”, which comprises some principles of etiquette on how to work together, these instruments target at organizing, moderating, and improving the social interaction process, but also at the individual’s motivation to engage in collaboration:

- Discussion pages give space to complex negotiation processes among potentially conflicting viewpoints (which can also be seen on the knowledge layer as it often leads to a more holistic view on the subject).
- Changes are visible in a global (e.g. on the special page “recent changes”) as well as in a local change log (each page has a version history allowing for tracking all changes and offering the possibility to roll back in case of destructive behaviour). This transparency also helps to socialize new community members by showing them how collaboration actually works and which form of behavior is acceptable.
- Users can have a user identity (with a user page for self-portrayal, e.g., listing contributions, or as elected administrator with extended rights, or as mentor), and watchlists that support the responsibility for certain articles. These promote the feeling of social esteem and allegiance to the Wikipedia community and mission, important for user motivation (e.g. [Schroer 07]).
- Wikipedia also limits activities by technical means in order to cope with problems where different opinions don’t converge naturally and result in so-called edit wars. Featured articles are write-protected (only the German Wikipedia), and the English Wikipedia has introduced the “3-revert-rule”.

As a summary, Wikipedia instruments are primarily – as to be expected – on the artefact level: they indicate and promote the maturity of articles. But a significant number of instruments also address the social layer, developing the wikipedians’ social competency to contribute collaboratively to maturing processes.

3 Quantitative Analysis of Wikipedia: Maturity Metrics

In our second step we have focused on a quantitative analysis of Wikipedia to find measures for maturity at the artefact level; i.e. the maturity of articles. As maturing is a process, it is in particular interesting to observe the development of the articles over the time and not just snapshots. Wikipedia provides a rich data set of articles with their detailed version history. The articles characterized as featured articles are furthermore interesting as it is the highest available distinction within Wikipedia.

3.1 Existing quantitative studies on Wikipedia

A number of quantitative studies for assessing the quality of Wikipedia articles have been conducted. These focused in particular on the growth of Wikipedia (e.g. [Buriol 06] or [Voss 05]) or the number of edits and editors (e.g. [Wilkinson & Huberman 07] or [Kittur 07]) in order to make some assessments. [Lih 04] focused on content construction and use processes from the perspective of participatory journalism. He proposed the total number of edits (rigor) and the total number of unique editors (diversity) as metrics for quality. [Emigh & Herring 05] studied the formality of language and compared Wikipedia articles with the traditional Comlumbia Encyclopaedia and with the also community-based Everything2 based on frequency of parts of speech measurement. [Stvilia 05] studied the English Wikipedia and applied readability metrics by Flesch and Kincaid (cf. [Flesch 48]). He has shown that the readability for featured articles is better than median values for readability of random articles.

The studies of [Viégas 04] and [Buriol 06] are the only ones taking into account temporal aspects. [Viégas 04] developed a tool called “History Flow” for visualizing the version history of articles. With this tool they could reveal collaboration patterns like *mass deletion*, *offensive copy*, *phony copy*, *phony redirect*, *idiosyncratic copy*. Against expectations, the length of articles does not stabilize over time. Instead, articles with more than 100 edits are continuously growing.

[Buriol 06] made a temporal analysis of the Wikigraph. They examined in- and outgoing links from articles within the Wikipedia, but they did not considered broken internal links, that means links referring to not yet existent entries (highlighted in Wikipedia as red links). They generated webgraphs based on 17 snapshots. They found out that newer articles grow slower than older ones and older articles get more edits, which they interpret as maturity of coverage but not of depth. Concerning the linkage, they could state a power-law distribution of the ingoing and the outgoing links with an increase of outgoing links from 7 to 16 over the observed period.

3.2 Measures and Metrics

For the purpose of our study, we have developed three metrics used as simple measure for maturity. These are prototypical indicators for estimating the maturity level of articles over time.

- **Readability**¹. The assumption is that the more an article matures the better it is readable and understandable. As indicated by [Stvilia 05], featured (i.e. more mature) articles have a higher readability score. We revert to Flesch’s Reading Ease Score (RES) adjusted to the German language by [Amstad 78]: $RES(German) = 180 - (\#words/\#sentences) - 58.5 * (\#syllables/\#words)$
- **Presentation and Format**. A good and mature article is not only characterized by its content at word level but also by its presentation and format. Articles are structured by and by; i.e. an article is parted in sections, headings are introduced and images for illustration are added. Therefore, we are looking at the number of words per headline and the number of words per image.

¹ Stvilia specified readability as *complexity*

- **Linkage.** The more knowledge matures the better it can be embedded into different contexts. If users extend articles and add contents, cross references can be set. These can be links to other Wikipedia pages but also to articles outside Wikipedia. Here, we are regarding the number of words per internal link as well as the link density comprising internal and external links.

3.3 Results

The basis of our analysis is the German Wikipedia. We are using the full XML dump generated on 2007-01-24, containing all article pages with their complete version history and corresponding discussion pages (also with history) as well as pages beyond the articles' name space (e.g. user profiles). Due to time constraints, the following results refer to the first 68854 pages within the XML dump. As the order within the dump is only semi-alphabetical, the obtained articles range from A to Z. Of 68854, we filtered 19245 pages containing user profiles, discussion pages and lists. Because of the abolition of the explicit "stub" tag for articles within the German Wikipedia in 12/2005, we defined stubs as articles with less than 200 words. Thus, we received 19852 "stub" articles, 612 good articles, 343 featured articles and 28763 "normal" articles without any markup.

	Stub	Normal	Good	Featured
#articles	19852	28763	612	343
#words per article	43 (4) 60	1196 (753) 1359	5386 (4580) 3428	6689 (5952) 3561
RES per article	35 (41) 32	50 (53) 35	55 (55) 7	54 (54) 6
#headlines p. art.	0 (0) 1	7 (5) 8	21 (18) 13	24 (22) 14
#images per article	0 (0) 0	2 (0) 6	8 (6) 10	11 (8) 13
#words p. headline	105 (96) 50	198 (157) 198	391 (229) 880	438 (249) 883
#words per image	132 (138) 45	747 (493) 827	1289 (722) 1897	1181 (647) 1633
#(unique) internal links	6 (1) 9 6 (1) 9	75 (55) 71 82 (58) 87	212 (170) 144 238 (188) 171	240 (213) 137 272 (233) 169
#words p. int. link	5 (3) 5	16 (13) 17	29 (21) 25	31 (24) 26
link density (%)	31 (33) 16	9 (8) 4	5 (5) 2	4 (4) 2

Table 1: Comparison of article quality classes (mean (median) σ)

In a first run, we compared the four article quality classes by applying the measures on the last versions of the articles (cf. Table 1). In total, the mean and median results correlate to the maturing process. As expected, the number of headlines and images per article increases on average from stub to normal and good to featured articles; as well as in the case of the linkage. Whereas the increase from stub to normal and from normal to good is evident, the difference between good and featured articles is not as big anymore. This can be traced back to the fact that there is no clear definition of the

differences between good and featured articles and the non-obligatory tagging of featured articles as good. However, the standard deviations are severe and do not decrease as expected; except for the RES where the differences between normal, good and featured are only small but the standard deviation decreases obviously.

These results are further confirmed regarding the evolution of featured articles (cf. Figure 1). In a second run, we analyzed the set of 343 featured articles and their history and looked at the measures' development over the time. Therefore, we applied the measures to every version of an article and uniformly aggregated the versions to an interval of 100 discrete scores. It is interesting that, contrary to the first comparison, the median RES is higher in the early phases and then decreases and stabilizes on a value of 54 (cf. Figure 1a). The words per headline, images and internal link develop as expected (cf. Figure 1 b-d). Although articles are always growing, the words per headline converge to about 250-300 words. There is no significant difference between the featured articles tagged as good beforehand and the ones directly tagged as featured. Regarding the standard deviations, they increase similarly unexpectedly high (except for the RES). Thus, on average maturing can be observed in the development of Wikipedia articles, but these simple measures do not allow the automated classification of a single article.

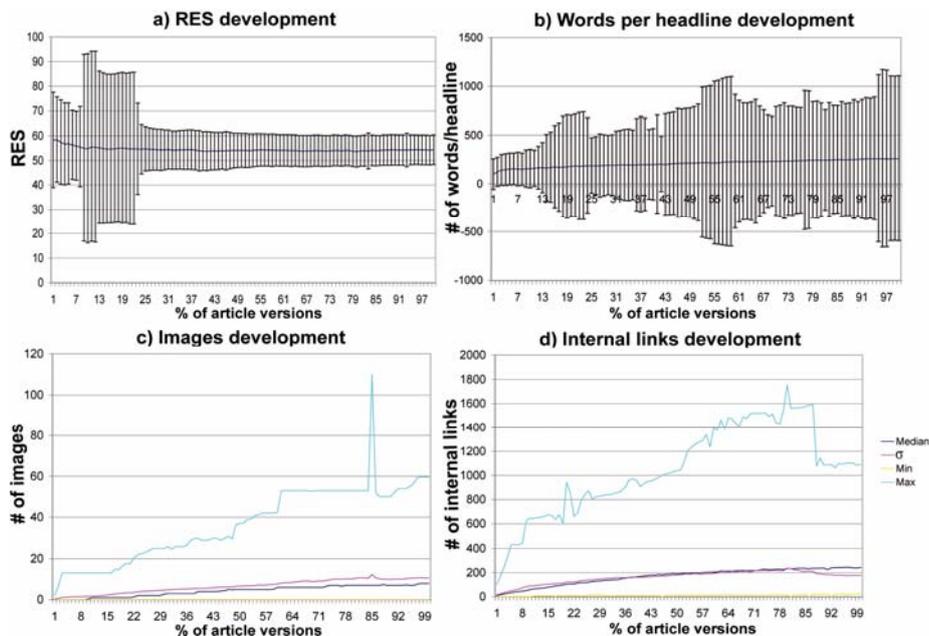


Figure 1: Measures development over the version history of featured articles

4 Wikipedia vs. Enterprise Wiki Systems

While Wikipedia as a large and transparent social experiment is ideal for empirical analysis, it should be clear that results from analyzing Wikipedia cannot be mapped

one to one on knowledge maturing phenomena in enterprise wikis. Concerning knowledge maturing, the most important difference lies in the systematic bias of Wikipedia to more mature knowledge: The criteria qualifying an article for inclusion in Wikipedia clearly favour more mature knowledge elements. On the other hand, enterprise wikis are mostly used to improve informal learning of immature elements, which are often still highly contextualized. Thus, the maturing focus of Wikipedia is clearly on the artefact level, whereas of enterprise wikis, being part of knowledge management initiatives, it is more on the knowledge level. Table 2 summarizes these and some more differences between the two. Still, the patterns of interaction remain the same: collaborative improvement and enrichment of content and socially negotiated interaction protocols. So results of our qualitative and quantitative study can also be applied to enterprise wiki systems, which is planned to be confirmed by specific enterprise wiki studies in the future.

	Wikipedia	Enterprise wikis
Size of community	large numbers	limited to (parts of) the organization
Goal	general public encyclopaedia	persistence and exchange of experiences
Type of knowledge	rather mature knowledge (at least ad-hoc training)	immature knowledge (mostly distribution in communities and formalization, up to ad-hoc-training)
Maturing focus	artefact level (knowledge is considered to be sufficiently mature)	knowledge level (artefacts are considered to be facilitating the collaboration and exchange)
Motivation	idealism and identification with Wikipedia goals, quest for social esteem	work process needs, professional esteem, organizational goals

Table 2: Differences between Wikipedia and enterprise wikis

5 Conclusions and Outlook

Our Wikipedia study has confirmed that maturing actually takes place in wiki systems and that social negotiation processes have brought forth a set of mechanisms that foster maturing. Furthermore, starting from the criteria developed in [Maier & Schmidt 07], a quantitative analysis has shown that there are measures indicating maturing of artifacts. But individual differences between articles (diagnosed by the high standard deviation) do not allow for determining maturity of artefacts in an automated way based on these simple measures. The results suggest that we need to go beyond artefact measures and rather consider creation and usage contexts.

We are aware that Wikipedia and enterprise wiki systems have several differences; most notably the difference in magnitudes of its user community and the systematic Wikipedia bias towards artefacts communicating more mature knowledge. We still argue that the mechanisms of Wikipedia and the identified measures can also be applied to enterprise wiki systems. Of course, this will be validated in a future study of an enterprise wiki system.

For improving knowledge maturing processes, we can derive from this study that we need to consider all layers of maturing (artefact, knowledge and social). These are mutually dependent – a fact which has often been neglected by systems focusing merely on the artefact level (i.e., on documents). Especially motivational aspects need to be taken into account: community members engage in maturing activities not only because it is useful (or they expect a future benefit), but also because they identify themselves with the socio-technical system as whole and specific parts of it (e.g., as caretakers). Apart from motivation, awareness of activities plays a key role (cf. [Lih 04]): only if others are aware of what's going on, what's there and what needs to be done, they will invest time in it. This social layer has to be analyzed in more depth.

From a technological point of view, an important lesson is that the instruments should not be the result of technical design, but rather of social negotiation. Technical systems should constrain collaboration processes as little as possible.

Acknowledgements

This work has been partially funded by the German Federal Ministry for Education and Research within the project Im Wissensnetz. We wish to thank Kerstin Schmidt and Henning Jacobs for their contributions to this work.

References

- [Amstad 78] Amstad, T.: "Wie verständlich sind unsere Zeitungen?", Dissertation, Zurich, Switzerland, 1978
- [Buriol 06] Buriol, L. S.; Castillo, C.; Donato, D.; Leonardi, S.; Millozzi, S.: "Temporal Analysis of the Wikigraph", In Proc. WI'06, 2006, 45-51
- [Emigh & Herring 05] Emigh, W.; Herring, S. C.: "Collaborative authoring on the web: A genre analysis of online encyclopedias", In Proc. HICSS'05, 2005
- [Flesch 48] Flesch, R.: "A new readability yardstick", *Journal of Applied Psychology*, Vol. 32, 1948, 221-233
- [Kittur 07] Kittur, A.; Chi, E. H.; Pendleton, B. A.; Suh, B.; Mytkowicz, T.: "Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie", In Proc. CHI'07, 2007
- [Lih 04] Lih, A.: "Wikipedia as participatory journalism: Reliable source? Metrics for evaluating collaborative media as a news resource", *Proc. Int. Symp. Online Journalism*, 2004
- [Maier & Schmidt 07] Maier, R.; Schmidt, A.: "Characterizing Knowledge Maturing: A Conceptual Process Model for Integrating E-Learning and Knowledge Management", In Proc. Conf. of Professional Knowledge Management, 2007, 325-334
- [Schmidt 05] Schmidt, A.: "Knowledge Maturing and the Continuity of Context as a Unifying Concept for Knowledge Management and E-Learning", In Proc. I-KNOW'05, 2005
- [Schroer 07] Schroer, J.; Hertel, G.: "Voluntary Engagement in an Open Web-based Encyclopedia: Wikipedians, and Why They Do It", In Proc. Congress of the European Association of Work and Organizational Psychology, 2007
- [Stvilia 05] Stvilia, B., Twidale, M. B., Gasser, L., Smith, L.C.: "Information quality in a community-based encyclopedia", *Proc. ICKM'05*, 2005, 101-113

[Viegas 04] Viegas, F.B.; Wattenberg, M.; Dave, K.: „Studying cooperation and conflict between authors with history flow visualizations”, In Proc. CHI'04, 2004, 575–582

[Voss 05] Voss, J.: “Measuring Wikipedia”, In Proc. ISSI'05, 2005

[Wilkinson & Huberman 07] Wilkinson, D.M.; Huberman, B.A.: “Assessing the value of cooperation in wikipedia”, *First Monday*, 12(4), 2007

Special Track on Knowledge Organisation and Semantic Technologies

Introduction

Johannes Liegl

(Know-Center Graz, Austria,
jliegl@know-center.at)

Mathias Lux

(Klagenfurt University, Information Technology, Austria,
mlux@itec.uni-klu.ac.at)

Viktorija Pammer, Peter Scheir

(Knowledge Management Institute, Graz University of Technology, Austria
and Know-Center Graz, Austria,
{vpammer,pscheir}@know-center.at)

Michael Granitzer, Stefanie Lindstaedt

(Know-Center Graz, Austria,
{mgrani,slind}@know-center.at)

Although semantic technologies are becoming more and more widespread, there are still a number of challenging research topics, ranging from intelligent search engines and ontology driven systems in general to ontology engineering issues. The common point of interest lies in semantic technologies and methods for sharing and retrieving information and in organising knowledge in models and in information systems.

Succeeding the Special Track on Advanced Semantic Technologies (AST 06) at the I-Know 2006, KOST 2007 puts the focus more widely on the knowledge contained in system and views semantic technologies as a means of expressing and using this knowledge.

Based on the reviewers' reports, 5 from 11 submitted contributions were accepted for presentation within the special track and for publication within the conference proceedings.

The accepted contributions range from the one end of the spectrum, ontology engineering, to the other ends, namely application of semantic technologies and evolution of formal ontologies towards integrating community-oriented aspects.

Kremen and Kuba contributed work on explanations for modeling errors in semantic web ontologies. Valarakos, Spiliopoulos, Kotis and Vouros present a framework to facilitate the development of ontology mapping tools. These works support our view, that as semantic technologies evolve, complex methods supporting the actual ontology engineering will gain in importance.

Tifous, El Ghali, Dieng-Kuntz and Giboin present an ontology of communities of practice, including experiences on collaborative ontology engineering. Grahl, Hotho and Stumme disregard ontologies but use folksonomies instead for categorising bookmarks.

Mödritscher, Hoffmann and Klieber finally present a knowledge artefact repository that uses ontologies as a way of organising the content of the repository.

Programm Committee

- Wernher Behrendt (Salzburg Research Forschungsgesellschaft, Salzburg, Austria)
- Wernher Behrendt (Salzburg Research, Austria)
- Paul Buitelaar (German Research Center for Artificial Intelligence, Germany)
- Sven Meyer zu Eissen (Bauhaus-Universität Weimar, Germany)
- Andreas Faatz (SAP Research, Germany)
- Joanna Guss (EADS Corporate Research Center, France)
- Andreas Hotho (BibSonomy & University of Kassel, Germany)
- Dietmar Jannach (Klagenfurt University, Austria)
- Harald Kosch (University of Passau, Germany)
- Annett Mitschick (Dresden University of Technology, Germany)
- Elena Simperl (Freie Universität Berlin, Germany)
- Marc Spaniol (RWTH Aachen, Germany)
- York Sure (University of Karlsruhe, Germany)
- Douglas Tudhope (University of Glamorgan, Pontypridd, United Kingdom)

Special Track Organization

- Johannes Liegl, Know-Center Graz (Austria)
- Mathias Lux, Klagenfurt University (Austria)
- Viktoria Pammer, Knowledge Management Institute, Graz University of Technology & Know-Center Graz (Austria)
- Peter Scheir, Knowledge Management Institute, Graz University of Technology & Know-Center Graz (Austria)
- Michael Granitzer, Know-Center Graz (Austria)
- Stefanie Lindstaedt, Know-Center Graz (Austria)

Incremental Approach to Error Explanations in Ontologies

Petr Křemen, Zdeněk Kouba

(Czech Technical University in Prague, Czech Republic
{kremen,kouba}@labe.felk.cvut.cz)

Abstract: Explanations of modeling errors in ontologies are of crucial importance both when creating and maintaining the ontology. This work presents two novel incremental methods for error explanations in semantic web ontologies and shows that they have significantly better performance than the state of the art black-box techniques. Both promising techniques together with our implementation of a tableau reasoner for an important OWL-DL subset *SHIN* are used in our semantic annotation tool prototype to explain modeling errors.

Key Words: ontologies, knowledge management, error explanations

Category: I.2.4, I.2.1, H.3.1, H.3.4, H.5.2

1 Introduction

The problem of error explanations turned out to be of high importance in ontology and semantic annotation authoring tools. Users of such tools need to be informed both about inconsistencies in the modeled ontology and about reasons for the inconsistency to occur. In this work we present two novel incremental algorithms for error explanations. These algorithms are universal enough to be reused with wide variety of reasoners, which seems to be of high importance especially in the dynamic field of semantic web languages.

We have tested the proposed algorithms with our implementation of *SHIN* description logics [Baader and Sattler 2001] (which is a significant subset of the OWL-DL language for semantic web) tableau algorithm. Both the explanation algorithms and the reasoner are used in the core of our novel annotation tool prototype to explain unsatisfiability of concepts (ontology classes).

2 Error Explanation Techniques - State of the Art

The mainstream of error explanations for description logics knowledge bases tries to pinpoint axioms in the knowledge base to localize the error. The notion of *minimal unsatisfiability preserving subterminology* (MUPS) has been introduced in [Schlobach and Huang 2005], to describe minimal sets $\{S_i\}$ of axioms that cause given concept to be unsatisfiable. Removing a single axiom from each of these sets turns the concept satisfiable. Similarly to defining MUPSes for concept satisfiability, in [Kalyanpur 2006] the notion of *justification* for arbitrary axiom

entailments has been presented. These justifications allow for explaining knowledge base inconsistencies, in our case annotation errors. Notions of MUPSeS and justifications are dual, as for each concept an axiom can be found, for which the set of justifications corresponds to the set of MUPSeS of the concept. From now on, we will use w.l.o.g. only the notion of MUPSeS and concept satisfiability.

At present, there are two approaches for computing explanations of concept unsatisfiability: black-box (reasoner-independent) techniques and glass-box (reasoner-dependent) techniques. The former ones can be used directly with an existing reasoner, performing many satisfiability tests to obtain a set of MUPSeS. The latter ones require smaller number of satisfiability tests, but they heavily influence the reasoner internals, thus being hardly reusable with other reasoning algorithms.

2.1 Black-box Techniques

There are plenty of black-box techniques that can be used for the purposes of error explanations. In [Banda et al. 2003], several simple methods based on *conflict set tree* (CS-tree) are shown. CS-trees allow for non-redundant searching in the power set of a given axiom set. Each node in a CS-tree is labeled with two sets, a set D of axioms that necessary belong to a MUPSeS and a set P of axioms that might belong to a MUPSeS. Each node represents the set $D \cup P$ and it has $|D \cup P|$ children, each one lacking an axiom from $D \cup P$. The method (denoted as *allMUPSeSbb*) introduced in [Banda et al. 2003] effectively searches the CS-tree in the depth-first manner, pruning necessarily satisfiable nodes.

An interesting black-box approach [Kalyanpur 2006] is based on a method for computing a single MUPSeS (denoted as *singleMUPSeSbb*) of a concept for given axiom set. In the first phase, this algorithm starts with an empty set K and fills it with all available axioms one by one until it becomes unsatisfiable. In the second phase, each axiom is conditionally removed from K . If the new K turns satisfiable, the axiom is put back. To obtain an algorithm for all MUPSeS the general purpose Reiter's algorithm [Reiter 1987] for computing hitting sets of given conflict set is used. This algorithm generates a tree, each node of which is labeled with the knowledge base and a MUPSeS computed for this knowledge base using *singleMUPSeSbb*. Starting with an arbitrary root MUPSeS, each of its children is generated by removing one of the MUPSeS axioms from the knowledge base and computing a single MUPSeS for the new knowledge base. The search terminates when all leaves of the tree are satisfiable. The advantage of this approach is that it provides also repair solutions that are represented by axioms of minimal (w.r.t. set inclusion) paths starting in root. These paths correspond to hitting sets of the set of MUPSeS. Due to the lack of space we refer to the works [Banda et al. 2003], [Kalyanpur 2006], [Schlobach and Huang 2005] for detailed algorithm descriptions.

2.1.1 Glass-box Techniques

A fully glass-box technique for axiom pinpointing in the description logic \mathcal{ALC} [Baader and Sattler 2001] is introduced in [Schlobach and Huang 2005]. However, for more expressive languages, like OWL-DL, OWL-Lite or OWL 1.1, there is no full adaptation of this approach. A partially glass-box method for searching a single MUPS in OWL-DL is presented in [Kalyanpur 2006]. This method provides only supersets of MUPSeS which justifies its usage as a preprocessing step in the first phase of the *singleMUPSbb* algorithm described above.

3 Incremental Approach to Error Explanations

High number of very expensive tableau algorithm runs required for black-box methods, as well as lack of glass-box methods, together with their poor reusability, has given rise to the idea of using incremental techniques for axiom pinpointing. These methods require the reasoner being able both to provide its current state, and to apply given axiom to given state. There is, however, no other interaction with the reasoner. These features place incremental methods somewhere between black-box and glass-box approach.

The following sections introduce two novel incremental methods for axiom pinpointing. Both of these methods need a reasoner state to be stored. Although the incremental approach can be used with a wide range reasoning algorithms, we have tested its feasibility with a tableau algorithm for \mathcal{SHIN} . The tableau reasoner state consists of two parts: a completion graph repository storing partially expanded completion graphs, and an axiom set used for expanding the completion graphs so far.

3.1 Computing a single MUPS

In this section, a novel incremental algorithm for computing a single MUPS is presented. This algorithm (see Algorithm 1) starts with an axiom list P and an empty state e of the reasoner. Axioms from P are tested one by one with the current reasoner state until the incremental test fails. The axiom $P(i)$ causing the unsatisfiability is put into the single MUPS core D , the rest of the axiom list is pruned and the direction of the search in the axiom list changes. The algorithm terminates, when all axioms are pruned.

Correctness.

Correctness of the algorithm is ensured by the following invariant. Before each direction change, D contains axioms that, together with some axioms in the non-pruned part of the axiom list, form a MUPS. Whenever an axiom i causes unsatisfiability, there must exist a MUPS that consists of all axioms in D , axiom

Algorithm 1 An Incremental Single MUPS Algorithm

```

1: function SINGLEMUPSINC( $P, e$ )                                ▷  $P \dots$  initial axioms,  $e \dots$  initial state.
2:    $lower, i \leftarrow 0$ 
3:    $upper \leftarrow length(P) - 1$ 
4:    $D \leftarrow \emptyset$ 
5:    $sD, last \leftarrow e$ 
6:    $direction \leftarrow +1$ 
7:   while  $lower \leq upper$  do
8:     if  $i \geq length(P)$  then
9:       return  $\emptyset$ 
10:    end if
11:     $(incState, result) \leftarrow test(P(i), last)$ 
12:    if  $result$  then
13:       $last \leftarrow incState$ 
14:       $i = i + direction$ 
15:    else
16:       $D \leftarrow D \cup \{P(i)\}$ 
17:       $(sD, result) \leftarrow test(P(i), sD)$ 
18:       $last \leftarrow sD$ 
19:      if  $direction = 1$  then
20:         $upper \leftarrow i - 1$ 
21:      else
22:         $lower \leftarrow i + 1$ 
23:      end if
24:       $direction \leftarrow -direction$                                 ▷  $+1 \dots$  right,  $-1 \dots$  left
25:    end if
26:  end while
27:  return  $D$ 
28: end function

```

i and some axioms in the previously searched part of the axiom list. This MUPS cannot be affected by pruning the axiom list tail that has not been explored in this iteration.

Example 1. Let's have an ontology containing six axioms numbered 1...6, where the unsatisfiability of a concept is caused by these axiom sets (MUPSes) $\{\{1, 2, 4\}, \{2, 4, 5\}, \{3, 5\}, 6\}$. The *singleMUPSInc* algorithm works as follows :

direction	input list	mups core
	[1, 2, 3, 4, 5, 6]	$D = []$
→	[1, 2, 3, 4, 5, 6]	$D = [4]$
←	[1 , 2, 3, 4, 5, 6]	$D = [4, 1]$
→	[1 , 2 , 3, 4, 5, 6]	$D = [4, 1, 2]$

Each line corresponds to a direction change. Whenever an unsatisfiability is detected, the search direction is changed, "overlapping" axioms are pruned (emphasized by strikeout) and the last axiom that caused the unsatisfiability (in bold) is put into the MUPS core D .

3.2 Computing all MUPSes

An incremental algorithm that can be used to search for all MUPSes (let's denote this algorithm as *allMUPSInc1*) is presented in [Banda et al. 2003]. This algorithm assumes, that a state of the underlying reasoner depends on the order of axiom processing. However, tableaux algorithms [Baader and Sattler 2001] are adopted in almost all current semantic web reasoners (for example Pellet). In case of tableau algorithms, two different permutations of an axiom set shall result in two equivalent states. Exploiting this fact, we modified the original algorithm

to decrease the number of redundant calls to the testing procedure, resulting in Algorithm 2 (`allMUPSInc2`).

The original algorithm `allMUPSInc1` manages three axiom lists D , T and P . At the beginning of each recursive call, D contains axioms that must belong to all MUPSeS searched in this recursive call, P represents possible axioms that might belong to some of these MUPSeS and T represents a list of already tested axioms. The first while cycle adds axioms from P to T while T remains satisfiable. If an axiom that causes unsatisfiability is detected, the execution is branched. The first recursive call tries to remove this axiom and go on adding axioms from P to T , while the second branch tries to add the axiom to the MUPS core D found so far. If D turns unsatisfiable, a MUPS has been found. If A does not contain a subset of this MUPS, it is inserted into A , and A is returned.

Our modification of the original algorithm avoids executing some redundant tests – both in the while cycle and in testing whether D turns unsatisfiable. For this purpose, we store the position of the first unsatisfiability test in P in the parameter *cached*. Let's denote $T = \{t_1, \dots, t_a\}$ and $P = \{p_1, \dots, p_b\}$. Then *cached* is such an index to P , that $\{t_1, \dots, t_a, p_1, \dots, p_{\text{cached}}\}$ is unsatisfiable and each of its subsets is satisfiable.

Algorithm 2 Modified version of *allMUPSInc1*

```

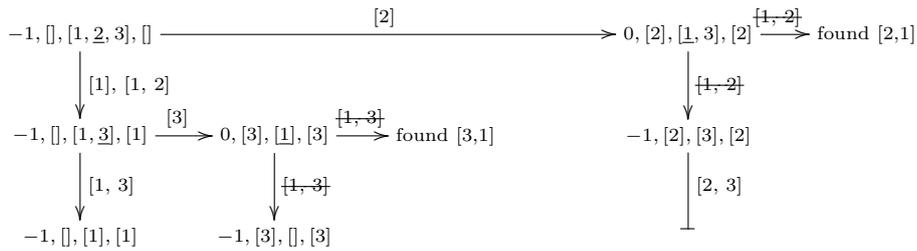
1: function ALLMUPSINC2( $D, sD, P, T, sT, A, \text{cached}$ )           ▷  $sD$  ( $sT$ ) is the state for  $D$  ( $T$ ).
2:    $result \leftarrow true$ 
3:    $i \leftarrow -1$ 
4:   while  $result \wedge \exists c \in P$  do
5:      $i \leftarrow i + 1$ 
6:     if  $c \notin T$  then
7:        $T \leftarrow T \cup \{c\}$ 
8:        $lT \leftarrow sT$ 
9:       if  $i = \text{cached}$  then
10:         $result \leftarrow false$ 
11:        break
12:       else
13:         $(result, sT) \leftarrow test(c, sT)$ 
14:       end if
15:     end if
16:   end while
17:   if  $result$  then
18:     return  $A$ 
19:   end if
20:    $A \leftarrow allMUPSInc2(D, sD, P \setminus \{c\}, T \setminus \{c\}, lT, A, -1)$ 
21:    $D \leftarrow D \cup \{c\}$ 
22:   if  $i = 0 \wedge d = t$  then
23:      $result \leftarrow false$ 
24:   else
25:      $(result, sD) \leftarrow test(c, sD)$ 
26:   end if
27:   if  $\neg result$  then
28:     if  $\neg \exists a \in A$  such that  $a \subset D$  then
29:        $A \leftarrow A \cup \{D\}$ 
30:     end if
31:     return  $A$ 
32:   end if
33:   return  $allMUPSInc2(D, sD, P \setminus \{c\}, D, sD, A, i - 1)$ 
34: end function

```

Correctness.

Correctness of the algorithm is ensured by the same invariant as for *allMUPSInc1* presented in [Banda et al. 2003] and the fact that, the variable *cached* uses the information of the last successful test performed on *T* only in the second recursive call, where $D = T$. In the first recursive call, the information cannot be used, as the sets *D* and *T* differ.

Example 2. To show how *allMUPSInc2* works, assume an axiom set $\{1, 2, 3\}$. MUPSeS for unsatisfiability of a concept are $\{\{1, 2\}, \{1, 3\}\}$. The algorithm runs as follows :



Each node in this graph represents a call to the procedure *allMUPSInc2*, with the signature *cached, D, P, T*. The search starts in the node $-1, [], [1, 2, 3], []$ and is performed in the depth first manner preferring up-down direction (first recursive call) to the horizontal one (second recursive call). Axioms that cause unsatisfiability in the given recursive call are underlined. Edges are labeled with the tests that have been done before the unsatisfiability is found and struck axiom sets represent the tests that are not performed, contrary to *allMUPSInc1*. In this example *allMUPSInc2* requires 6 tests contrary to 10 tests executed by *allMUPSInc1*.

4 Experiments

First, the discussed methods have been compared with respect to the overall performance. Two ontologies have been used for tests: the miniTambis ontology (30 unsatisfiable concepts out of 182) and the miniEconomy ontology¹(51 unsatisfiable out of 338). As shown in Tab.4, the performance of incremental methods is significantly better than the fully black box approach. Furthermore, combination of Reiter’s algorithm and *singleMUPSInc1* is typically 1-2 times worse than the fully incremental approaches. However, the main advantage of the *singleMUPSInc1* in comparison to the fully incremental approaches is that it allows direct computation of hitting sets of the set of MUPSeS (i.e. generating repair diagnoses), which makes them more practical.

It can be seen that our modification *allMUPSeSInc2* of *allMUPSeSInc1* provides a slight increase in the performance. To evaluate the difference in more detail, see Fig.1. This figure shows the significance of caching for different MUPS configurations. The highest performance gain (over 30%) is obtained for ontologies containing a lot of MUPSeS with approx. half size of the ontology size.

¹ To be found at <http://www.mindswap.org/2005/debugging/ontologies>

	miniTambis (time [ms])	miniEconomy (time [ms])
allMUPSbb	> 15min.	> 15min.
Reiter + singleMUPSbb	67481	> 15min.
Reiter + singleMUPSinc	19875	19796
allMUPSInc1	8655	14110
allMUPSInc2	7879	12970

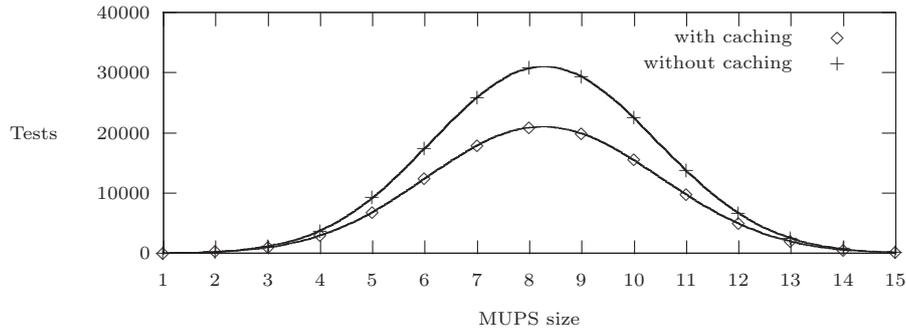
Table 1: Comparison of incremental and black-box algorithms.

Figure 1: Comparison of incremental algorithm with caching and without it. Different configuration of MUPSes of 15 axioms, say $\{1, \dots, 15\}$, were tested. For each $1 \leq k \leq 15$ (the x-axis), the set of all MUPSes is generated, so that it contains all axiom combinations of size k , thus containing $\frac{15!}{k!(15-k)!}$ MUPSes. For example, for $k = 2$, the set of MUPSes is $\{\{1, 2\}, \{1, 3\}, \dots, \{1, 15\}, \{2, 3\}, \dots\}$.

5 Conclusions and Future Work

Two novel incremental algorithms for finding minimal sets of axioms responsible for given modeling error in an ontology have been introduced. The first one is an incremental algorithm that searches for one such minimal axiom set (MUPS). The second one is an extension of the fully incremental algorithm presented in [Banda et al. 2003] used for searching all minimal axiom sets.

The introduced incremental methods seem promising and more efficient than the fully black box approaches in the context of error explanations. Although the fully incremental approaches are more efficient than the combination of single MUPS testers and Reiter's algorithm, they do not allow to compute diagnoses directly. This justifies our focus on both approaches. A problem of generating diagnoses by the fully incremental methods efficiently is still an open issue.

The explanation service for concept satisfiability based on the discussed incremental methods is implemented in our new annotation tool prototype (see Fig.2), providing the user with concept unsatisfiability explanations [Křemen 2007].

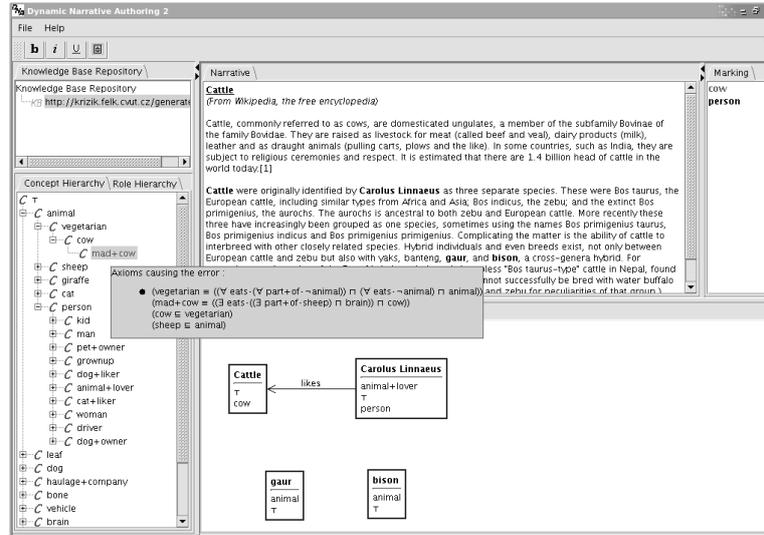


Figure 2: Semantic Annotation Tool Prototype. To detect and explain modeling errors our implementation of the *SHIN* reasoner and presented incremental algorithms are used. The annotation graphs (bottom pane) represent annotations created from the current document (top pane) and the ontology (left pane).

Acknowledgements

This work has been supported by the grant No. MSM 6840770038 “Decision Making and Control for Manufacturing III” of the Ministry of Education, Youth and Sports of the Czech Republic.

References

- [Baader and Sattler 2001] Baader, F., Sattler U.: “An overview of tableau algorithms for description logics”; *Studia Logica*, 69:5-40, 2001.
- [Baader et al. 2001] Baader, F., Calvanese D., McGuinness D., Nardi D., Patel-Schneider P. (eds.): “The Description Logic Handbook, Theory, Implementation and Applications”; Cambridge, 2003.
- [Banda et al. 2003] De la Banda M. G., Stuckey P. J., and Wazny J.: “Finding All Minimal Unsatisfiable Subsets”; In *PPDP’03C*, 2003.
- [Kalyanpur 2006] Kalyanpur A.. ”Debugging and Repair of OWL Ontologies”; PhD thesis, University of Maryland, 2006.
- [Křemen 2007] Křemen P. ”Inference Support for Creating Semantic Annotations”; Technical Report GL 190/07, CTU FEE in Prague, Dept. of Cybernetics, 2007.
- [Schlobach and Huang 2005] Schlobach S. and Huang Z. ”Inconsistent Ontology Diagnosis: Framework and Prototype”; Technical Report, Vrije Universiteit Amsterdam.
- [Reiter 1987] Reiter R.: “A Theory of Diagnosis from First Principles”; *Artificial Intelligence*, 32(1):57-96, April 1987.

AUTOMS-F: A Java Framework for Synthesizing Ontology Mapping Methods

Alexandros G. Valarakos

(AI Lab, Inf. & Comm. Systems Engineering Dept., University of the Aegean, Samos, Greece
alexv@aegean.gr)

Vassilis Spiliopoulos

(AI Lab, Inf. & Comm. Systems Engineering Dept., University of the Aegean, Samos, Greece
and Inst. of Informatics & Telecomm., NCSR "Demokritos", Greece
vspiliop@aegean.gr)

Konstantinos Kotis

(AI Lab, Inf. & Comm. Systems Engineering Dept., University of the Aegean, Samos, Greece
kotis@aegean.gr)

George A. Vouros

(AI Lab, Inf. & Comm. Systems Engineering Dept., University of the Aegean, Samos, Greece
georgev@aegean.gr)

Abstract: Although ontologies promise an effective technology for information integration, it is often the case that two or more information providers do not share the same ontology. Several (semi)-automated ontology mapping methods have been developed towards solving this problem. This paper presents AUTOMS-F, a framework implemented as a Java API, which aims to facilitate the rapid development of tools for automatic mapping of ontologies by synthesizing several individual ontology mapping methods. Towards this goal, AUTOMS-F provides a highly extensible and customizable application programming interface. AUTOMS is a case study ontology mapping tool that has been implemented using the AUTOMS-F framework, and has been successfully evaluated in the international OAEI 2006 contest.

Keywords: Ontology Mapping, Information Integration, Semantic Web, Semantic Technology, Java Framework

Categories: I. 2. 4, D.2.7

1 Introduction

Ontologies have been realized as the key technology to shaping and exploiting information for the effective management of knowledge and for the evolution of the Semantic Web and its applications. In such a distributed setting, ontologies establish a common vocabulary for community members to interlink, combine, and communicate knowledge shaped through practice and interaction, binding the knowledge processes of creating, importing, capturing, retrieving, and using knowledge. However, it seems that there will always be more than one ontology even for the same domain. In such a setting, where different conceptualizations of the same domain exist, information services must effectively answer queries, bridging the gaps between conceptualizations of the same domain. Towards this target, networks of semantically related information must be created at-request. Therefore mapping of ontologies is a

major challenge for bridging the gaps between agents (software and human) with different conceptualizations.

“Simple” cases of heterogeneity include ontologies that use different lexicalizations of the same ontology element (e.g., “car” and “road-vehicle”). More complicated situations appear in cases where ontologies are structured (in terms of concepts’ relations) in completely different ways.

Information integration and effective management of information will be admittedly achieved through reaching an agreement, by producing a single, commonly-agreed and shared reference ontology, or by achieving coordination so that each party uses its own ontology, but with it also establishes concept and relation mappings with other ontologies. In any case, tools for supporting the ontology mapping task are of paramount importance. Specifically, given two ontologies O_1 and O_2 , establishing a mapping between them involves computing pairs of ontology elements (one from O_1 and one from O_2) that have the same intended meaning.

AUTOMS-F (AUTomated Ontology Mapping through Synthesis - Framework) is a Java application programming interface (API) that aims to facilitate the development of integrated tools for automatic one-to-one mapping of domain ontologies. The main concern of AUTOMS-F is the provision of facilities for the synthesis of several ontology mapping methods. The ultimate goal is to provide synthesized approaches realized as integrated tools that produce better results and performance measures than each of the synthesized individual mapping methods alone. The framework has been used for the implementation of the AUTOMS mapping method [Kotis, 06] which is described as a case study in the fourth section of this article.

The paper is structured as follows: Section 2 presents the ontologies mapping problem, the requirements and the assumptions made towards implementing AUTOMS-F. Section 3 describes AUTOMS-F in detail. Section 4 presents AUTOMS, a specific mapping tool implemented using AUTOMS-F, as a case study of using the proposed framework. Section 5 presents related work, and section 6 concludes the paper, sketching our future plans.

2 Problem Statement

A mapping between two ontologies is mainly expressed by a one-to-one function between ontology elements (i.e., the concepts and the properties of ontologies). Therefore, establishing a mapping between ontology elements [INTEROP, 04], involves the computation of pairs of elements whose meaning is assessed to be similar. Similarity in meaning can be computed using a number of metrics that exploit ontology elements’ features. It is important to note that the *mapping process* does not modify the involved ontologies: It produces, as output, a set of *mapping pairs* together with their computed similarity measure.

The majority of the mapping methods/tools can be described by the generic mapping process [Ehrig, 04] depicted in Figure 1. The discrete steps of this process are as follows:

1. *Feature Engineering*: Ontologies are transformed into an internal representation. This step selects a fragment of the ontology to be processed.
2. *Search Step Selection*: Element pairs from the two input ontologies are being selected, with the one element belonging to the first ontology and the other

to the second. Depending on the mapping method, all element pairs or only a subset of them may be considered. The set of pairs constitute the search space of the method.

3. *Similarity Computation*: This step computes the similarity of the previously selected pairs. Many different similarity metrics may be utilized by a single method.
4. *Similarity Aggregation*: In this step all similarity metrics are aggregated into a single one.
5. *Interpretation*: This step concludes to a set of mapping pairs by exploiting the aggregated similarities computed in the previous step (e.g., this can be done using threshold value(s)).
6. *Iteration*: The whole process may be repeated several times, by propagating and updating the assessed similarities, taking into account the structure of the input ontologies.

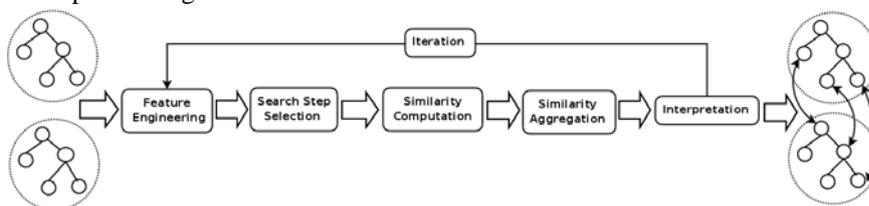


Figure 1: The generic and commonly accepted discrete steps of a mapping process.

We conjecture that any framework that aims to facilitate the development of ontology mapping tools must support the development of the generic steps shown in Figure 1. AUTOMS-F, aiming to the provision of a generic framework for the development of mapping tools, in accordance to the steps proposed, poses a number of requirements:

1. According to the *Feature Engineering* step, a mapping method may utilize only a subset of the available information provided by the input ontologies. Different mapping methods should be able to use different sets of features.
2. According to the *Search Step Selection* step, a method may examine only a subset of the possible mapping pairs, while different methods should be able to select different subsets of pairs, using well-defined conditions.
3. According to the *Similarity Computation* step, different mapping methods may need to compute different similarity measures for the assessment of mapping pairs.
4. According to the *Similarity Aggregation* step, the synthesis of mapping methods and the aggregation of their corresponding similarity measures should be robust, expandable and be easily supported by the mapping framework.
5. According to the *Interpretation* step, the mapping pairs may be produced based on the aggregated similarity values of custom mapping methods.

3 AUTOMS-F: Architecture and Implementation

AUTOMS-F is an open-source application programming interface (API) implemented in Java which aims to provide a basic framework towards synthesizing ontology mapping methods.

3.1 Framework's Conceptualization

The main concept of AUTOMS-F is the *mapping method*. A *mapping method* exploits all the information concerning the mapping process: The pair of ontologies involved and functional information regarding the execution of the mapping steps specified in section 2. This is done by linking the mapping method to:

- one or more *mapping methods*. When a mapping method is associated with at least another mapping method or another association of mapping methods, then this association constitutes a “task” (or synthesized mapping method). A task specifies the synthesis of different (atomic or synthesized) methods. Tasks, due to their recursive definition specify a hierarchical tree of arbitrary complexity, which is named the “*mapping association tree*”.
- a *parser*. This collects the appropriate elements of the ontologies that are involved in the mapping process. This collection defines the potential mapping pairs and results in an $(n \times m)$ similarity matrix, where n and m are the number of elements of the target and source ontology, respectively. The value of each matrix entry specifies the similarity of the specific pair to which the entry corresponds. A parser is associated to a task and is applied to its subsequent methods. Moreover, a parser is inherited to subsequent tasks (i.e., tasks lower in the hierarchy). In order to support parser inheritance property, we assume that tasks in the mapping association tree use parsers that collect pairs of ontological elements that are super-sets of the one collected by subsequent tasks' parsers.
- a *similarity method*. Such a method specifies the way similarities between pairs of ontological elements are being computed. Such method results to an $(n \times m)$ similarity matrix.
- an *aggregation operator*. This is responsible for specifying the way similarity matrices are being combined. Aggregation operators, as well as results, are associated with the task that combines the specific methods.
- a *pairs filter*. This defines the criteria for selecting the best mapping pairs from the resulted similarity matrix.
- a *results renderer*, which is responsible for the presentation of the *mapping pairs*.

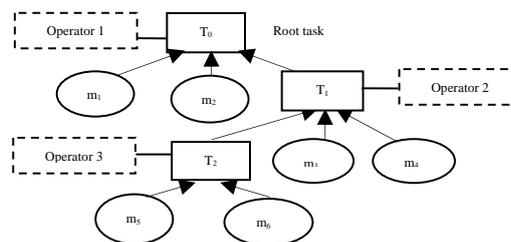


Figure 2: An example of a mapping association tree.

According to the above, the synthesis of mapping methods in AUTOMS-F is supported in two ways: a) by allowing a mapping method to have direct access to the similarity matrix computed by another method, and b) by combing the similarity matrices of mapping methods using specific aggregation operators. Figure 2 depicts an example of a mapping association tree in which aggregation operators are shown with dashed line rectangulars: They are attached to tasks (rectangulars), and each task aggregates one or more methods (shown in ovals).

The “mapping association tree” specifies also the execution order of the mapping methods and the application of the operators and parsers. The top-most task in the tree is named the “root task”. This is the T_0 task in Figure 2. We assume a right-to-left bottom-up execution order of the methods in a tree. Hence, according to Figure 2, the method m_5 follows the execution of method “ m_6 ”. “ m_4 ” executes and may exploit the aggregated results that are produced by the methods “ m_6 ” and “ m_5 ”. The aggregation of “ m_6 ” and “ m_5 ” methods is being preformed by an aggregation operator associated to task T_2 .

Parsers are associated to *tasks* and apply to tasks’ subsequent methods. However, different *parsers* can be defined at any level of the tree. Because of that, different methods may exploit different collections of element pairs: Generally, the similarity matrix of a method “contributes” to the computation of the similarity matrix of the root task, which always contains the super-set collection of ontological element pairs.

The root task comprises a *pair filter* that selects the pairs that specify “best mappings” (e.g., whose similarities are above a specific threshold value). Finally, the renderer presents the *mapping pairs*.

3.2 Implementation Issues

AUTOMS-F has been developed using the Jena Framework [Jena, 07]. It has been implemented in Java for ensuring platform independency and uses various well-established design patterns [Brandon, 02] for ensuring usability, reuse, extensibility and abstraction.

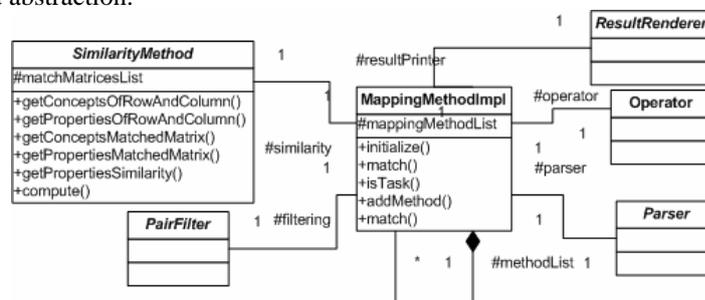


Figure 3: UML diagram of the main AUTOMS-F classes, attributes and operations.

Figure 3 depicts a UML diagram of the main classes of the framework according to its conceptualization (section 3.1). The “MappingMethodImpl” class is linked through an aggregation relation with itself and it aggregates at least one “MappingMethodImpl” class. Also, the same class is linked with exactly one of the following abstract classes: “SimilarityMethod”, “PairFilter”, “Parser”, “Operator” and “ResultRenderer”. The “MappingMethodImpl” class stores a list with the methods-tasks to which it is linked

using the “mappingMethodList” attribute. This method is responsible for doing the necessary initializations (“initialize” operation) and for performing the mapping operations (the “match” operation of the “MappingMethodImpl” class). The “SimilarityMethod” class supports various manipulations of the similarity matrices to support the synthesis of methods. Due to space restrictions we present only some of the attributes and operations of the system. All the classes, except the “MappingMethodImpl” class, constitute hot spots for the framework, hence are the classes that can be further extended.

The Strategy pattern - behavioral design pattern - is used in the “MappingMethodImpl” class to support the creation of different mapping methods. The template method pattern in the “SimilarityMethod” class - a behavioral pattern - is used for the computation of the similarity of a pair of ontological elements. Thus, instantiating the framework, one can define - override - the methods that measure the similarity between a pair of ontological elements and leave the construction of the similarity matrix to the “SimilarityMethod” class. Also, the composition pattern - a structural pattern - is exploited for the specification of the “mapping association tree”. AUTOMS-F, as it exploits Jena’s model loader, can handle ontologies that are implemented in RDF, RDFS, OWL and DAML+OIL formalisms. The ontologies can be read from the local disk or be accessed through their URLs. An ontology element can be any of Jena’s ontology class (OntClass) or property (OntProperty) objects. Hence, a method can retrieve any information about an ontology element, i.e., label, super-concepts, class properties etc.

AUTOMS-F contains samples of all the extensible classes resulting in a default mapping method. More advanced mapping methods can be developed by extending the “SimilarityMethod” class and overriding the methods that measure the similarity between ontology elements. Also, someone may integrate a method to the framework by extending the “SimilarityMethod” class and overriding only its “compute” operation, which is responsible for executing the similarity method. This means that the new class computes the mapping and the similarity matrix, which is defined in the extended class. Following the latter case, special attention should be given to the collection of the ontological elements pairs used: This must be consistent with the collection of the T_0 task. To ensure this consistency, we recommend the use of a framework-based defined parser and not the use of any user-specific manipulation of the ontological elements pairs.

4 A Case Study: The AUTOMS tool

AUTOMS-F has been used for the development of the AUTOMS ontology mapping tool which synthesizes 6 mapping methods [Kotis, 06]: The *lexical*, *semantic*, *simple structural*, *properties-based*, *instances-based* and the *iterative structural* method. Figure 4 depicts the association tree of AUTOMS and the position of the mapping methods in it. The lexical and semantic methods are executed first. Then the structural matching method follows by exploiting the results of the previously run methods, whose results have been aggregated by task T_2 . Afterwards, AUTOMS executes the properties-based and instances-based mapping methods, and finally, the iterative structural matching method is being executed by exploiting results from the other methods in its level as well as from the task that aggregates results from lower levels.

AUTOMS uses the same parser and aggregation operator in any of its tasks. The parser is defined in the T_0 task and the operator of each task unifies the matrices of constituent methods.

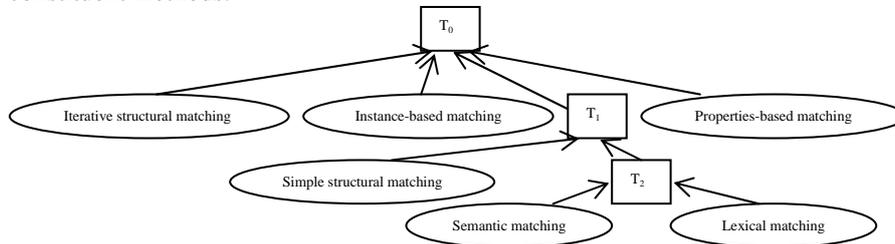


Figure 4: AUTOMS's mapping association tree.

The requirements of AUTOMS have been satisfied by the flexibility and extensibility provided by AUTOMS-F. The learning curve of the framework was rather short. In some cases, AUTOMS developers needed to extend the framework for capturing OAEI contest's requirements [OAEI, 06]; however this did not effect the development of AUTOMS, and AUTOMS-F proved to be a very robust and flexible framework. AUTOMS developers have been provided with an optimized version of AUTOMS-F, resulting to quite short (comparing to other tools of the OAEI contest) execution times. Scalability was a weak point at the time AUTOMS-F was used to develop AUTOMS, since very large ontologies (~30MB) provided by the OAEI organizers could not be loaded and parsed. AUTOMS tool was evaluated in the OAEI 2006 contest among 10 other systems, achieving very high precision and recall measures, providing evidence for the potential of AUTOMS-F to this direction. For a detailed view of contest's results and AUTOMS performance please visit OAEI 2006 results Web page at <http://oaei.ontologymatching.org/2006/results/benchmarks/>.

5 Related Work

To the best of our knowledge the only work that is related to the presented one is the Alignment API [Alignment API, 07]. It is used for the evaluation of the ontology mapping tools that participated in the OAEI contest [OAEI, 06]. It has been implemented using Java and provides an API for incorporating, evaluating and presenting the results of different tools.

AUTOMS-F and the Alignment API are based on different technologies. AUTOMS-F uses the Jena framework whereas the Alignment API uses the OWL API [OWL API, 07]. Moreover, the Alignment API executes mapping methods in a pipe line, in contrast to AUTOMS-F which defines an execution structure of the mapping methods - the "mapping association tree" - facilitating the effective synthesis of different mapping methods as well as their parallel execution by exploiting the thread mechanism. The Alignment API supports the combination of two methods by means of fixed operators i.e. compose, join, inverse and meet which combine the result matrices of the constituent methods. However, these operators have not been fully implemented as far as the version 2.5 is concerned. On the other hand, using AUTOMS-F one has the flexibility to define her own aggregation operators,

combining more than two matrices. Also, AUTOMS-F supports the use of different parsers for collecting ontology elements. Different parsers can be applied in the context of a specific method or task. At the current version, AUTOMS-F does not provide any evaluation utilities, whereas the Alignment API does. In general, AUTOMS-F provides more hot spots than the Alignment API, thus making itself more extensible and customizable. Alignment API is under LGPL license. AUTOMS-F will be available soon in www.icsd.aegean.gr/ai-lab under GPL licence.

6 Concluding Remarks and Future Work

AUTOMS-F addresses the ontology mapping problem by providing advanced facilities for the development of synthesized ontology mapping methods. AUTOMS is an evaluated case of the framework's potential. Although the full automation of ontology mapping is still a challenge, AUTOMS-F provides a robust framework for the synthesis of different mapping methods, increasing the benefits of deploying state of the art mapping technology.

We plan to extend AUTOMS-F in several ways. Firstly, execution threads will be added to the methods of a task at each task level, in order to decrease the execution time. Secondly, we will introduce a consistency checking method that will ensure consistency between the mapping result pairs according to ontologies specification semantics. Thirdly, we will investigate a way to introduce iterative execution in the framework, hence satisfying all the steps of the generic mapping process shown in Figure 1. Finally, we will investigate scalability issues and evaluation utilities.

Acknowledgements

This work was supported by the ONTOSUM project (www.ontosum.org).

References

- [Alignment API, 07] Alignment API, <http://alignapi.gforge.inria.fr/>
- [Brandon, 02] G. Brandon, *The Joy of Patterns*, Addison-Wesley, ISBN 0-201-65759-7, 2002.
- [Ehrig, 04] M. Ehrig, S. Staab, QOM - Quick Ontology Mapping, GI Jahrestagung (1), 2004.
- [INTEROP, 04] INTEROP - Network of Excellence, State of the art and state of the practice including initial possible research orientations. Project n. 508011, 2004.
- [Jena, 07] Jena Java Framework, <http://jena.sourceforge.net/>
- [Kotis, 06] K. Kotis, et al., AUTOMS: Automating Ontology Mapping through Synthesis of Methods, OAEI 2006 contest, Ontology Matching International Workshop, USA, 2006.
- [OAEI, 06] OAEI: Ontology Alignment Evaluation Initiative. <http://oaei.ontologymatching.org/2006/>
- [OWL API, 07] OWL API, <http://owl.man.ac.uk/api/readme.html/>

O'CoP, an Ontology Dedicated to Communities of Practice

Amira Tifous

(INRIA, Edelweiss Team, Sophia Antipolis France
Amira.Tifous@sophia.inria.fr)

Adil El Ghali

(INRIA, Edelweiss Team, Sophia Antipolis France
Adil.ElGhali@sophia.inria.fr)

Alain Giboin

(INRIA, Edelweiss Team, Sophia Antipolis France
Alain.Giboin@sophia.inria.fr)

Rose Dieng-Kuntz

(INRIA, Edelweiss Team, Sophia Antipolis France
Rose.Dieng@sophia.inria.fr)

Abstract: The Palette project dedicated to learning in Communities of Practice (CoPs) aims to offer several services for CoPs, in particular Knowledge Management (KM) services based on an ontology dedicated to CoPs, the so-called O'CoP. Built from information sources about the Palette CoPs, O'CoP aims both at modelling the members of the CoP and at annotating the CoP's knowledge resources. The paper describes the structure of O'CoP, its main concepts and relations, and it reports some lessons learnt from the cooperative building of this ontology.

Keywords: Community of Practice, Knowledge Management, Ontology

Categories: I.2.4, H.4.3, H.5.3

1 Introduction

CoPs are “groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis” [Wenger et al., 02].

The objectives of the Palette IST project (<http://palette.ercim.org/>) are to develop services for CoPs: information, knowledge management (KM), and mediation services. Eleven pilot CoPs are involved in the participatory design of the Palette services. These CoPs, located in various European countries (Belgium, France, Greece, Switzerland, UK), belong to three different domains: (i) teaching (e.g. @pretec, a CoP of Belgian teachers playing the role of resources-persons to support the use of Information and Communication Technologies (ICT) in schools), (ii) management (e.g. ADIRA, a French professional association gathering executives from medium to large IT companies in Rhône-Alpes region), and (iii) engineering (e.g. UX-11, a CoP composed of 150 IT engineer-students practicing GNU/Linux). The CoPs' size varies from less than ten members to more than a hundred of members.

KM services aim at supporting CoPs' management of their knowledge resources, so as to improve: (i) the access, sharing, and reuse of existing knowledge, and (ii) the

creation of new knowledge. A knowledge resource can be either a *document* materialising the knowledge made explicit by CoPs' members when cooperating, or a *person* holding tacit knowledge. The KM services will be based on Semantic Web technologies: they will rely on an ontology (describing concepts useful about a CoP, its actors and their competences, its resources, its activities, etc.) and on the semantic annotation of the CoPs' knowledge resources w.r.t. this ontology. In [Vidou et al., 06], we proposed generic models useful for understanding a group activity, collaboration, competencies, learners' profiles, and lessons-learned. A CoP being a specific kind of a group, the CoP-dedicated ontology, so-called O'CoP, is based on these generic models. It consists of CoP-relevant concepts and relations with which the CoPs' resources can be annotated. These CoP-relevant concepts and relations are specialisations of the high-level ontology constituted by the generic concepts used to represent the generic models. The CoP-oriented KM services will rely on the O'CoP ontology.

After summarising our ontology development method and the ontology structure (section 2), we will describe its main concepts (section 3), the lessons learnt from its building (section 4), before concluding (section 5).

2 Ontology Development Method and O'CoP Ontology Structure

Our method for developing the O'CoP ontology includes the following steps:

- *Information sources collection*: selecting three main sources to be used either as *corpora* where picking out candidate terms, or as *grids* for extracting candidate terms: (i) Rough-data documents (audio records/files of CoPs' interviews, transcriptions and minutes of these interviews; the interviews were performed by Palette members that played the role of *mediators* between some specific CoP and the knowledge engineers); (ii) Analysed-data documents (e.g., syntheses of interviews, vignettes and scenarios structuring the CoPs' activities); (iii) Methodological and theoretical documents (e.g., our generic models and existing ontologies or thesaurus);
- *Contextualised lexicon construction*: selecting from the corpora and w.r.t. the grids (i) the *terms* relevant for describing the CoPs and (ii) their respective contexts (i.e. the text surrounding the terms) to help understand the terms.
- *Vocabulary identification* i.e. refining the contextualised lexicon once validated by the CoPs' mediators and producing, for each term, a definition and some examples of use.
- *Hierarchy building*: (i) identifying the terminological concepts and relations, and (ii) structuring them, and eventually adding new higher-level concepts.
- *Ontology formalisation* in RDF/S, the formal language agreed in Palette.

A tool, called ECCO, supports these iterative steps and provides the user with mechanisms enabling to keep the traceability of the sources of the candidate terms.

The resulting O'CoP ontology is structured into three main layers (see Figure 1):

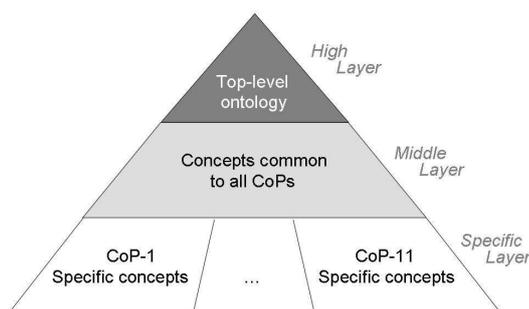


Figure 1: Structure of the O'CoP ontology

- A *high layer* (or top-level ontology) including the concepts and relations needed to represent the generic models presented in [Vidou et al., 2006]; they served as a grid for analysing the corpora and building the other layers of the ontology.
- A *middle layer* including the concepts common to all CoPs. These concepts correspond to terms confirmed by the mediators as common to all CoPs. They are specialisations of the high-level ontology concepts. Note that some concepts stemming from literature on CoPs could be included in this common layer, provided that they are attested by at least the CoP corpora.
- A *specific layer* including the concepts specific to each CoP: these concepts correspond to terms confirmed by the mediators as specific to a given CoP or to very few CoPs.

3 Description of the Main Concepts of the Ontology

3.1 Community

The main concepts related to the community in the O'CoP ontology are:

- *Community*: it can be a *community of interest*, a *community of learners*, a *goal-oriented community* or a *community of practice*. In the interviews, interviewees acknowledged that the group of persons they are in (be it so-called a *community of teachers*, a *network of teachers*, a *resource-persons community*, an *association of companies*, etc.) is a (kind of) CoP.
- *Domain* and *Field*: as defined in [Wenger, 2004], the *Domain* is the area of knowledge that brings the community together, gives it its identity and defines the key issues that the CoP's members need to address. It is the "focus" of the CoP and evolves over its life span in response to new, emerging challenges and issues [Henri, 06]. As for the *Field*, it is the "context" of the CoP; it can be referred to as the "discipline" or the "branch of knowledge" of the CoP's members (e.g. the *Domain* of ePrep¹ is the Educative use of ICT and its *Field* can be Mathematics, Physics, etc.).

¹ A CoP gathering teachers of French "Classes préparatoires aux Grandes Ecoles" interested in ICT (cf. http://www.eprep.org/Presentation_GB.htm#CoP).

- *Objective*: related to the CoP as a whole, or to a part of it (a group, a project, a team, etc. depending on the CoP’s organisation and functioning modes), an objective can be *Permanent* or *Temporary*.
- *CoP’s characteristics*: the CoP’s identity is characterised by (i) the *Membership*: is the CoP open to any person interested in it or are there some conditions (e.g., competency, cooptation, etc.) for entering the CoP? (ii) the *Cultural Diversity* (from homogeneity to heterogeneity) of the CoP’s members w.r.t. the nationality, profile, organisational culture [Langelier and Wenger, 2005].
- *Organisational structure*: the organisation of CoPs varies from formal and structured (e.g. the CoP ADIRA², which is based on a “board of governors”) to informal (e.g., the DL³ which is based on “informal subgroups”).

Figure 2 shows some relations concerning communities.

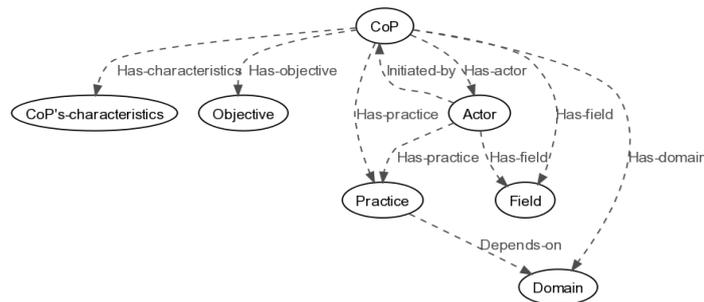


Figure 2: Relations concerning a community

3.2 Actors

We define an *Actor* as “an Individual or a Legal entity intervening in the CoP”. The *Actors* of a CoP are not only the CoP’s members, but also the entities interacting with the CoP (also called the CoP *environment*). A *Legal entity* can be a *Professional organisation* or an *Institution* (*Companies* and *Educational institutions*). *Actors* can be involved in the CoP as *Members*, *Contributors* (*Individuals* participating in particular activities or during specific periods of the CoP’s life) or *Partners* (*Legal entities* supporting the CoP).

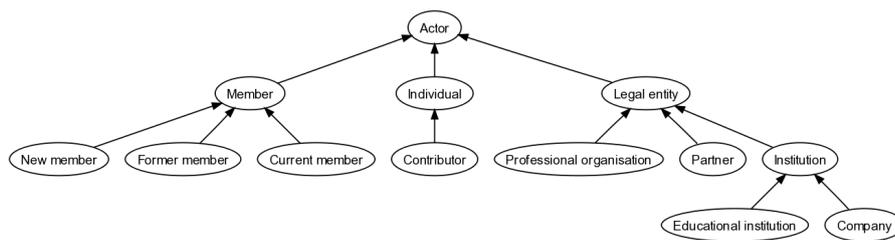


Figure 3: Concepts describing Actors in a CoP

² <http://www.adira.org/>

³ Doctoral Group Lancaster : <http://domino.lancs.ac.uk/>

Moreover, the *Actors* of a CoP can be defined according to their:

- *Role in the CoP*: it represents the *Actor*'s position in the CoP, which can be divided in two types:
 - *Governance role*: in order to interact, learn and share knowledge effectively, the CoP's actors (e.g. the members) need a support, which can be provided by: (i) *Facilitator*: s/he encourages the participation of the members, facilitates the interactions among them. (ii) *Coordinator*: s/he organises and coordinates the activities and events of the CoP. We distinguish the *Individual coordination* (ensured by one main coordinator) and the *Collective coordination* (in the case of a CoP organised per groups or teams, where individual local coordinator belongs to a coordination group or team). (iii) *Animator*: s/he guides and manages the community, ensures its development, relevance and effectiveness. An *Animator* thus plays both roles of *Facilitator* and *Coordinator*.
 - *Peripheral role*: represents knowledge providers and receivers. They are more or less involved or active in the CoP, their participation depends on the *Actors* who play these roles (personality, motivation, period, activity, etc.).
- *Individual profile*: identifies a CoP's *Individual* inside and outside the CoP. It comprises the concepts of *Competency* and *Occupation*.
- *Practice*: CoP's *members* are practitioners in an *Institution*, outside the CoP. They meet physically or virtually, through the CoP, which constitutes a channel for them to exchange about their shared Practice (e.g. teaching practice).
- *Behaviour*: the Attitude of the member towards the CoP gives more information about his/her degree of engagement in the CoP.

3.3 Competency

A *Competency* is defined as a set of *Resources* provided or to be acquired by an *Actor* (who plays a particular *Role* in some *Environment* or *Situation*) so that the Actor can perform, or help some other Actor to perform some *Activity*. Figure 4 gives a partial view of the Competency-Resource component and shows some relations concerning the Competency concepts.

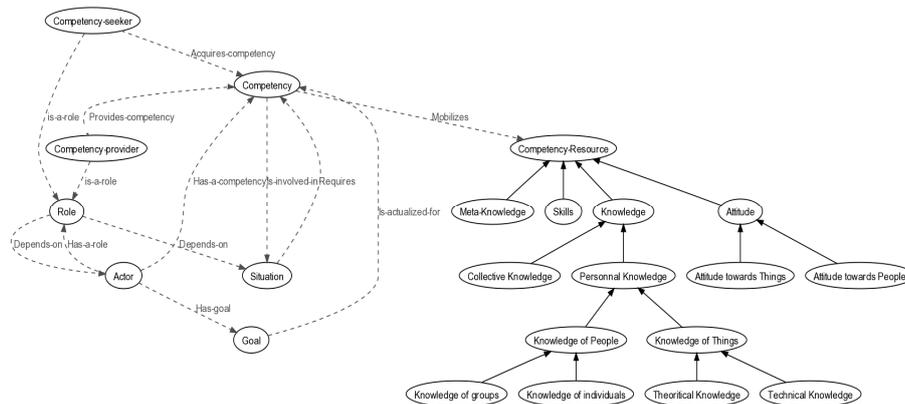


Figure 4: Excerpt of Competency concept hierarchy and some relations describing it

3.4 Resources

The *Resources* handled by a CoP are subdivided into:

- *Tools* defined according to the needs of community and their functionalities. A hierarchy describes the categorisation of these tools answering recurrent needs of a CoP including knowledge capturing (*Knowledge portal*), knowledge storage and sharing (*Repository*), collaboration (*Workspace, Agenda*, etc.).
- *Materialised resources* including documents, discussion. This last type of resources in the CoPs is associated to the interactions that hold within the CoP. These discussions can be *synchronous* (chat, audio and video conferences, etc.) or *asynchronous* (mail, forum, etc.). Almost all Palette CoPs are interested in easy access to these interaction traces and in archiving them.

On the other hand, these resources can be characterized w.r.t. many dimensions, for instance, Figure 5 shows some of the concepts needed to deal with the ownership of resources in a CoP.

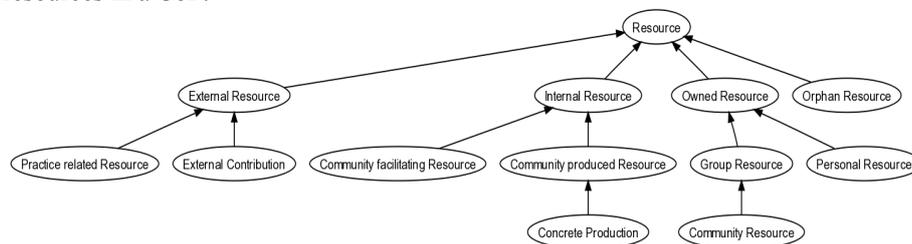


Figure 5: Categorisation of resources ownership

4 Lessons Learnt from the O'CoP Ontology Development

Concerning information sources collection, *the relevance of the terms extracted from the corpora strongly depends on the relevance of the corpora*. CoPs' mediators had focused their interviews on the organisation of CoPs, and had scarcely asked questions about CoPs' practices. As a consequence, the transcriptions of interviews contained very few terms related to practices. This leads to an ontology in which concepts related to practices are not very numerous.

During the terminological analysis, we found several terms common to some CoPs but used to evoke different concepts: e.g. the term "*platform*" was used to designate, depending on the CoP: a website, a workspace for the CoP, that may contain its documents and where the discussions of members are hosted, or yet a dedicated software e.g. e-learning platform. Some terms were also used ambiguously to designate concepts: e.g. CoPs use different terms to designate the persons in charge of particular tasks in the CoPs ("*coordinator of the project*", "*local coordinator*", "*manager*", etc.), whereas these tasks are not well described and identified. Finally, some CoPs use different terms to designate the same concepts, these synonyms must be associated to the same concept in their ontologies in order to avoid redundancy. For example, the terms *Journal* and *Logbook* are used to designate the record of activities or practices of a CoP's member. More generally, the synonym terms (either in the same CoP or in several CoPs) were recognised by the validators, during the

phase of vocabulary identification and term validation. In the implementation of the ontology, the synonym terms corresponding to a given concept were indicated through the RDF/S label of this concept.

The different CoPs adopted different terminologies, sometimes quite specific to the CoP and rather different from the terminology usually found in literature on CoPs. Therefore, we did not include in the common layer of the ontology the concepts offered by literature (e.g. the taxonomy of facilitation tasks for CoPs proposed by [Tarmizi and de Vreede, 2005]) if they were not attested by the Palette CoPs' information sources.

The O'CoP ontology building was a distributed, cooperative process between: (a) 6 ontologists focusing on different parts of the ontology since each one was guided by one generic model, (b) 11 CoPs' mediators validating from the CoPs' viewpoints. This led to the need of integration of different viewpoints. The different ontologists had various ways of modelling knowledge: e.g. the concept of *Activity* was needed for modelling *Competency* and *Resource*. Concepts related to *Activity* were thus modelled with various detail grains and various perspectives, requiring more integration work. Moreover, the integration between different concepts developed by different ontologists was often performed through the introduction of relations linking such concepts (e.g. relation between an *Actor* and an *Activity*, etc.). Notice that such kinds of relations were emphasised in the generic models that guided us. But they needed to be refined for more specialised concepts.

Our approach was both bottom-up (relying on a deep analysis of the information sources on the CoPs) and top-down (guided by our generic models).

5 Conclusions

This paper presented an original ontology composed of more than 800 concepts and 80 relations, dedicated to CoPs, and more precisely aimed at enabling to annotate the CoPs' members and the CoPs' resources. The link between CoPs and ontologies was studied in some recent related work. In [O'Hara et al., 02], the authors present a method based on analysis of the relationships between instances of a given ontology in order to identify potential CoPs in an organisation. In [Bettahar et al., 06], the authors develop an ontology aimed at enabling services among a civil servant CoP; [Floyd and Ulena, 05] studies the design of situated ontologies for knowledge sharing in a CoP. In comparison to this related work, the O'CoP ontology is original through: (i) the method used to build it cooperatively from analysis of several real CoPs, (ii) its objective of enabling to annotate CoPs' resources in addition to modelling the notion of CoP, and. (iii) its 3-layered structure, with a generic layer, a middle layer gathering concepts common to all CoPs and a low layer specific to a given CoP. O'CoP was for example used by the @pretic CoP, in order to annotate the mails exchanged by the members of the CoPs about their problems in the use of ICT in schools. Our work can also be partially compared to the typology of virtual CoPs (i.e. CoPs interacting through ICT) proposed by [Dubé et al., 06] or to the typology of CoPs based on their knowledge characteristics [Klein et al., 05] but these typologies are not materialised through ontologies.

More generally, the O'CoP ontology can be specialised for a new CoP. The high and middle layers are generic and can thus be reused for any CoP. If the new CoP is

similar to one of the Palette CoPs, the low layer corresponding to this CoP can be reused. But if no Palette CoP is relevant, concepts more specific to the new CoP can be added in the low layer, possibly by relying on our method described in section 2.

As a further work, after achieving the current validation of the integrated O'CoP ontology by the CoPs' mediators, we will make the ontology available to all the Palette CoPs and develop several KM services based on it: knowledge creation, annotation, retrieval, presentation, evaluation, and evolution services.

Acknowledgements

We thank very much C. Evangelou and G. Vidou that worked on subontologies not reported in this paper, the CoPs' mediators for their intensive and fruitful work of validation, P. Durville, the ECCO tool developer, F. Gandon for his methodological support, and the European Commission for funding the Palette project.

References

- [Bettahar et al., 06] Bettahar, F., Moulin, C., Barthès, J.P.: Ontologies Support for Knowledge Management in E-Government Environment. In Proc. of ECAI'06 Workshop on Knowledge Management and Organizational Memories, Riva del Garda, Italy (2006).
- [Dubé et al., 06] Dubé, L., Bourhis, A., Jacob, R.: Towards a Typology of Virtual Communities of Practice Interdiscipl. Journal of Information, Knowledge and Management, 1(2006):69-93.
- [Floyd and Ulena, 05] Floyd, C., Ulena, S.: On Designing Situated Ontologies for Knowledge Sharing in Communities of Practice. 1st Int. Work. PHISE'05,(CAiSE'05) Porto (2005).
- [Henri, 06] Henri, F.: CoPs: Social Structures for the Development of Knowledge. PALETTE Kick-off Meeting, Lausanne, March 13-15. Also in <http://www.licef.telug.uqam.ca> (2006).
- [Klein et al., 05] Klein J.H., Connell, NAD, Meyer, E.: Knowledge characteristics of communities of practice. Knowledge Management Research & Practice, 3,2(2005): 106-114.
- [Langelier and Wenger, 05] Langelier, L., Wenger, E. (eds.): Work, Learning and Networked, Quebec, CEFRIO (2005).
- [O'Hara et al., 02] O'Hara, K., Alani, H., Shadbolt, N.: Identifying Communities of Practice: Analysing Ontologies as Networks to Support Community Recognition. IFIP 02, (2002).
- [Tarmizi and de Vreede, 05] Tarmizi, H., de Vreede, G.-J.: A Facilitation Task Taxonomy for Communities of Practice, Proc. 11th Americas Conf. on Information Systems, Omaha (2005).
- [Vidou et al., 06] Vidou, G., Dieng-Kuntz, R., El Ghali, A., Evangelou, C.E., Giboin, A., Tifous, A. and Jacquemart, S.: Towards an Ontology for Knowledge Management in Communities of Practice. PAKM06, Vienna, Austria, Springer LNAI 4333, p. 303-314. (2006).
- [Wenger, 98] Wenger, E.: Communities of Practice: Learning as a Social System. Systems Thinker (1998).
- [Wenger et al., 02] Wenger, E, McDermott, R., Snyder, W.M.: Cultivating Communities of Practice, Harvard Business School Press (2002).
- [Wenger, 04] Wenger, E.: Knowledge management as a doughnut: Shaping your knowledge strategy through communities of practice. Ivey Business Journal, 68,3 (2004).
- [Wenger et al., 05] Wenger, E., White, N., Smith, J.D., Rowe, K.: Technology for Communities. CEFRIO Book Chapter (2005).

Conceptual Clustering of Social Bookmarking Sites

Miranda Grahl

(Knowledge & Data Engineering Group, University of Kassel, Germany
grahl@cs.uni-kassel.de)

Andreas Hotho

(Knowledge & Data Engineering Group, University of Kassel, Germany
hotho@cs.uni-kassel.de)

Gerd Stumme

(Knowledge & Data Engineering Group, University of Kassel, Germany
stumme@cs.uni-kassel.de
Research Center L3S, Hannover, Germany
<http://www.l3s.de>

Abstract: Currently, social bookmarking systems provide intuitive support for browsing locally their content. A global view is usually presented by the tag cloud of the system, but it does not allow a conceptual drill-down, e. g., along a conceptual hierarchy. In this paper, we present a clustering approach for computing such a conceptual hierarchy for a given folksonomy. The hierarchy is complemented with ranked lists of users and resources most related to each cluster. The rankings are computed using our FolkRank algorithm. We have evaluated our approach on large scale data from the del.icio.us bookmarking system.

Key Words: folksonomies, knowledge management, data mining

Category: H.3.7, H.5.4

1 Introduction

Social resource sharing systems are a way of collaboratively organising collections of resources, and are thus a promising alternative to classical knowledge management approaches -at least in domains where stronger structured approaches like ontologies could not take hold yet, or where their maintenance is too costly. This will hold especially in domains where people with no experience in data modelling have to deal with the tools.

The underlying structure of social resource sharing systems are so-called *folksonomies*, i. e., taxonomies created by the folk. A folksonomy consists of the *personomies* of its users. A personomy is the collection of all resources of a user, combined with a set of *tags*, which are catchwords that can be chosen arbitrarily by the users. Navigation in social resource sharing systems goes along hyperlinks which allow, for instance, to visit for a given tag, a web page listing all resources which have been tagged with this tag by at least one user. These systems allow thus for direct search of relevant entries. They also allow for serendipitous

browsing, by following links to tags, users, and/or resources in a more or less random way. In a nutshell, folksonomy based systems are tuned for search and local navigation. With their tag clouds, social resource sharing systems also provide a simple mean to discover the overall content of their folksonomy. However, when the set of all tags becomes too large, one is looking for more structured ways of presenting the folksonomy's content.

In this paper, we present a conceptual, hierarchical clustering approach for folksonomies, and discuss its value for structuring of a folksonomy. First, we make iterative use of partitioning clustering (using two times the KMeans algorithm in our setting) on the set of tags. This step is followed by an application of our FolkRank algorithm [9] to discover resources and users that are related to the leaf clusters of the resulting cluster hierarchy, leading to the discovery of communities of interest. The generation of the cluster hierarchy is completely automatic, and may serve as input for a manual creation of a concept hierarchy (ontology) in a subsequent step. The result is a three-level hierarchy of sets of tags (see Fig. 1). The clusters on the lowest, most detailed level are complemented by lists of related resources and users.

In order to evaluate our approach, we have analyzed the large-scale popular social bookmarking system del.icio.us.¹ which is a simple-to-use interface that allows users to organize and share bookmarks on the internet.

State of the Art. The most similar feature to our approach that is implemented in del.icio.us is its tag cloud.² It provides a first, global overview over the content of the system, but does not allow further conceptual drill-down into the data, since a click on one of the tags leads directly to an unstructured list of bookmarks.

Clustering of tags is one approach to support browsing and search within social bookmarking sites and is therefore of large interest. The scientific work that is most similar to ours is presented by [2]. A graph based clustering approach called Metis [1] is used on the weighted tag co-occurrence graph to find tag clusters in del.icio.us and RawSugar data.

Simpler clustering approaches are using only the weights of the tag co-occurrence network to split the graph into independent subgraphs. Every subgraph is then considered as a cluster [10]. The exploration of the network along the co-occurrence graph is discussed in the blog of Rashmi Sihna.³ Simple probabilistic methods or association rule mining approaches are used to extract relation between tags in [12] and [11]. A hierarchical clustering approach is applied on the weighted tag graph in [8] in order to compute a tag hierarchy. Similar clustering approaches are used to construct a hierarchy of tags for blogs in [5]. There are many more

¹ <http://del.icio.us>

² <http://del.icio.us/tag/>

³ http://www.rashmisinha.com/archives/05_02/tag-sorting.html

potential clustering approaches [3] (e. g., text clustering), but their applicability on folksonomy data still remains to be clarified.

2 Dataset, Notations, and Algorithms

In the next section, we briefly present the used dataset for our experiment, and introduce some notations. Then we recall the basics of the clustering and the ranking algorithm that we used.

Dataset and Basic Notations. We have evaluated our approach on the social bookmarking system del.icio.us. Between July 27 and 30, 2005, we crawled del.icio.us and obtained a set U of 75,085 users, a set T of 456,666 tags, and a set R of 3,006,114 resources [9]. There were in total 7,281,940 posts, i. e., triples of the form (u, S, r) , indicating that user $u \in U$ has assigned all tags contained in $S \subseteq T$ to resource $r \in R$. The set $Y \subseteq U \times T \times R$ of all tag assignments, i. e., of all (user, tag, resource) triples that show up in at least one post, consisted of 17,362,082 tag assignments

KMeans – a Clustering Algorithm. We used the well known cluster algorithm KMeans [7] for our experiments as it provides in many cases good results. For KMeans, objects have to be represented in an n -dimensional vector space. As we will be working with a tag-tag-co-occurrence matrix [see Section 3], our objects will be tags, as well as each dimension (feature) of the vector space. The principle of KMeans is as follows: Let k be the number of desired clusters. The algorithm starts by choosing randomly k data points of D as starting centroids and assigning each data point to the closest centroid (with respect to the given similarity measure; in our case the cosine measure). Then it (re-)calculates all cluster centroids and repeats the assignment to the closest centroid until no reassignment is performed. The result is a non-overlapping partitioning of the whole dataset into k clusters.

Each cluster is described by its centroid. Usually one considers only the top n features of each centroid, i.e. those n dimensions of the vector space which have the highest values in the vector. A large set of alternative clustering approaches exists [3]. To build a concept clustering hierarchy, an obvious solution would be to apply an hierarchical agglomerative clustering approach. But preliminary experiments showed that the distribution of the cluster size is strongly skewed, i.e. the majority of tags is assigned to one heterogenous cluster. KMeans, on the other hand, provided clusters with balanced sizes, which are more suitable to human perception.

Folkrank – a Ranking Algorithm. To compute the users and resources that are most related to clusters of tags, we use our Folkrank algorithm [9]. Given a set of preferred tags, users, and/or resources of a folksonomy, Folkrank computes

a topic specific ranking which provides an ordering of the elements of the folksonomy in descending importance with respect to the preferred elements. FolkRank applies a two step approach to implement the weight-spreading ranking scheme on folksonomies. First, we transform the hypergraph between the sets of users, tags, and resources into an undirected, weighted, tripartite graph. On this graph A , we apply a version of PageRank [4] that takes into account the edge weights. Our FolkRank algorithm computes a topic-specific ranking in a folksonomy by using a differential approach.

3 Constructing the Conceptual Hierarchy

For generating the conceptual hierarchy, we first removed, in a preprocessing step, some spam, and computed a vector space representation of the set of tags. Then we clustered the remaining set of tags, resulting in the lowest, most fine grained level of clusters. For each cluster, we extracted one tag as description. These descriptions were clustered again, yielding the middle layer of the conceptual hierarchy. Finally, we computed pairs of tags as descriptions of these ‘meta-clusters’, yielding the highest, most general level of the hierarchy. These steps are described in detail in the remainder of this section, together with the FolkRank based computation of sets of related users and resources for each cluster on the most fine grained level.

Data Preprocessing. In del.icio.us, posts with more than 50 tags are usually spam. As they strongly bias the co-occurrence network of tags, we removed all these posts in a first step. Then we computed, for each pair t_i, t_j of tags, their co-occurrence: $W(t_i, t_j) := |\{(u, r) \in U \times R \mid (u, t_i, r) \in Y \wedge (u, t_j, r) \in Y\}|$. Each tag $t_i \in T$ is now represented in the $|T|$ -dimensional vector space by the vector $\mathbf{t}^i := (\mathbf{t}_j^i)_{j=1, \dots, |T|}$ with $\mathbf{t}_j^i := W(t_i, t_j)$ if $W(t_i, t_j) \geq 50$ and $i \neq j$, and $\mathbf{t}_j^i := 0$ else. The threshold of 50 turned out to be necessary to concentrate on the significant relationships between tags. Further experiments showed that lower thresholds resulted in clusters, contains a wide variation of unrelated tags. We removed all tags that were represented by the $\mathbf{0}$ vector, as they are only peripheral to the folksonomy (see also [6]). The set T was thus reduced to 6356 tags. The remaining ‘core’ tags were then clustered.

Iterated Clustering. The conceptual hierarchy is computed as follows from bottom to top.

(i) We cluster the set T of tags with k -Means with $k = 300$, resulting in a clustering $\mathbb{C} = \{C_1, \dots, C_{300}\}$ where each cluster contains 21.18 tags in average. (Five of these clusters are displayed completely at the lowest layer of Figure 1.) For each cluster C_i , we extracted from its centroid the tag \hat{t}_i with the highest

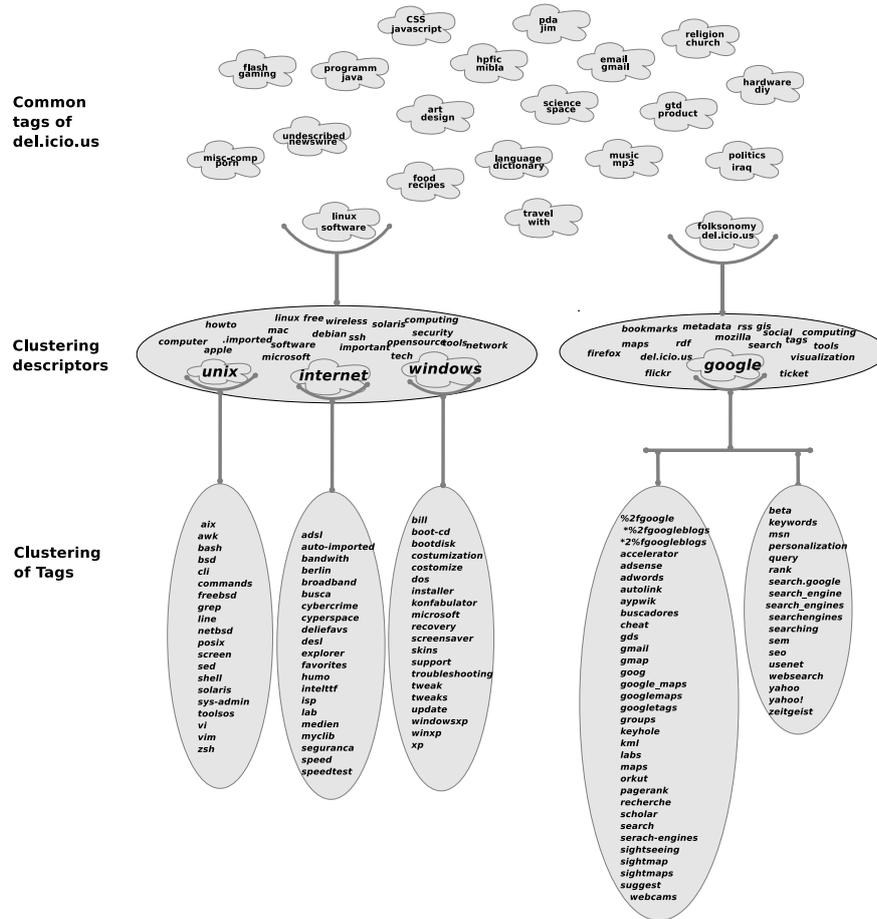


Figure 1: Conceptual hierarchy of the social bookmarking system del.icio.us.

value as description of the cluster.⁴ (The descriptors of 41 clusters are displayed in the middle layer of Figure 1.) We denote the set of all descriptors by $\hat{T} := \{\hat{t}_i | i = 1, \dots, 300\}$. Note that $|\hat{T}| = 274$ instead of 300, as one descriptor can have the highest value in more than one centroid (e.g., the tag ‘google’).

(ii) While the set \hat{T} provides already a good overview over the clusters computed above, it is still too large to be studied at a glance. Therefore, we clustered this set again with KMeans, this time with $k = 20$. We denote the result $\hat{C} = \{\hat{C}_1, \dots, \hat{C}_{20}\}$. (Two of the resulting clusters are displayed at the middle layer of

⁴ Usually, one is taking more than one entry of the centroid as description, e.g., ten, but in our case, already the first tags turned out to be highly descriptive, contributing in average 67,41% to the centroid.

Figure 1.) Again, we extracted for each cluster a description from its centroid. This time, however, the most central tag in the centroid is not significant enough, as it contributes in average only 14.45% to the centroid. Therefore, we extracted the two most central tags from each centroid. (All 20 resulting tuples are shown in the top layer of Figure 1.) These tuples are a condensed summary of the current content of del.icio.us and enable the user to start a top-down navigation of the system.

The choice of $k = 300$ and $k = 20$, resp., results from the aim to obtain clusters with about 10-20 elements each. Slightly smaller/larger choices of k did not significantly affect the results; larger differences resulted in too small or too large clusters that were of no use to the user.

Computing Related Users and Resources. After having structured the set of tags in a conceptual hierarchy, we complement now the most fine grained level of tag clusters with those users and resources which are most related to each cluster. I. e., for each of the clusters C_1, \dots, C_{300} , we compute a ranking of users and resources, resp., according to their relevance to the tags contained in the cluster. To this end, we have applied, for each cluster C_i , the FolkRank with $s = 0.85$ and with a preference vector \mathbf{p}^i composed as follows: $\mathbf{p}_j^i := 1 + 10,000$ if tag $t_j \in C_i$ and 1 else.

4 Results

We have applied the process to the data of the del.icio.us system⁵. Part of the resulting hierarchy is shown in Fig. 1. From top to bottom, the hierarchy allows us to explore the folksonomy in more and more detail. The top layer of the hierarchy is displayed completely in Fig. 1. Its 20 pairs of tags provide a first overview over the content of del.icio.us, and are thus comparable to the tag cloud.⁶ A main difference, however, is that the tag cloud of del.icio.us does not allow further drill-down along a conceptual hierarchy. When clicking on a tag in the tag cloud, del.icio.us directly presents all resources tagged with this tag. In our approach, we can navigate further down two more levels of the conceptual hierarchy.

For two selected entries, we have displayed the complete next level of the hierarchy. Let us first consider the subtree spanned by the tags ‘linux’ and ‘software’. The next level shows a cluster consisting of 23 tags, including the tags ‘linux’, ‘unix’ and ‘windows. The fact that only ‘linux’ made its way further up to the top level (because it had the largest contribution to the centroid of the cluster)

⁵ We also applied the approach to our BibSonomy system (<http://www.bibsonomy.org>). The results are not shown here due to space restrictions.

⁶ <http://del.icio.us/tag/>

indicates the preference of the del.icio.us users concerning operating systems. For three of the 23 tags, we have also displayed the full next (and last) level of the hierarchy. We see for instance that the tag ‘unix’ on the intermediate level summarizes many unix commands and tools, such as ‘vi’ or ‘awk’. The tag ‘windows’ on the intermediate level represents a cluster which contains the microsoft operating systems ‘xp’ (in three variants) and ‘dos’, and issues like ‘customization’, ‘installer’ or ‘update’. The tag ‘internet’ on the intermediate level shows a wider variety of topics on the subsequent level, as might be expected.

The second branch of the hierarchy, spanned by del.icio.us, also provides some interesting insights. First, we observe that one tag in the subsequent layer is ‘google’, which, in contrast to the popularity of this search engine, did not make it to the top level of the hierarchy. This may again indicate a bias of the del.icio.us users. A second observation is more of a technical nature: We observe that ‘folksonomy’ is the most central component of the centroid of the rightmost cluster on the middle layer, even though the tag itself is not contained in the cluster itself. This effect results from the fact that we have set $\mathbf{t}_i^i = 0$ in the vector space representation. A third observation is that the tag ‘google’ in the intermediate layer leads to two different clusters in the fine grained layer. This results from the fact that both clusters displayed in the lower right of Fig. 1 have a centroid in which ‘google’ is the largest entry. In fact, both clusters are related to Google, but address different aspects. The rightmost cluster is about search engines in general, including ‘seo’ [= search engine optimization], and the web applications Yahoo! and Zeitgeist. The second cluster from the right consists mainly of Google services, like Google Maps or Google Blog Search. This cluster contains also, to a lesser extent, research related tags: ‘scholar’, ‘pagerank’. If one is interested in the users or resources that are related to a cluster, one can use the results of PageRank. For the two Google clusters, the rankings are shown in Tab. 1. We see for instance, that the del.icio.us user ‘ubi.quito.us’ is the most relevant contributor to the Google service cluster, and the second most relevant contributor (behind user ‘fritz’) to the search engine cluster. The top URL of the clusters are shown in Tab. 1. Our conceptual hierarchy leads us thus also to (implicit) communities of interest among the del.icio.us users, and to clusters of related web pages.

5 Summary and Conclusion

In this paper, we have shown a way of building a conceptual hierarchy on the set of tags of a folksonomy by using partitioning clustering algorithms. The leaves of the tag hierarchy have been extended with corresponding clusterings of the sets of resources and users, resp., to allow for accessing all dimensions of the folksonomy. Concerning the evaluation of the presented approach, it is difficult

Table 1: Users and resources that are most related to the two Google clusters. The upper table relates to the left and the lower to the right cluster.

rank	user	rank	URL
0.002	ubi.quito.us	2.6E-4	http://www.keyhole.com/kml/kml_tut.html
0.002	kof2002	1.9E-4	http://www.googlesightseeing.com/
0.001	idealisms	1.9E-4	http://scholar.google.com/
6.4E-4	dajdump	1.9E-4	http://webaccelerator.google.com/
3.1E-4	dymphna	1.8E-4	http://www.shreddies.org/gmaps/
2.6E-4	laugharne	1.8E-4	http://www.arnebrachhold.de/2005/06/05/google-sitemaps-generator-v2-final
2.5E-4	konno	1.7E-4	http://www.google.com/webhp?complete=1&hl=en
2.5E-4	preoccupations	1.6E-4	http://www.keyhole.com/kml/kml_doc.html
2.1E-4	josquin	1.5E-4	http://serversideguy.blogspot.com/2004/12/google-suggest-dissected.html
2.1E-4	wxpbofh	1.3E-4	http://www.google.com/help/cheatsheet.html

rank	user	rank	URL
0.001	fritz	1.5E-4	http://www.vaughns-1-pagers.com/internet/google-ranking-factors.htm
3.8E-4	ubi.quito.us	1.4E-4	http://www.google.com/press/zeitgeist.html
2.9E-4	kof2002	1.1E-4	http://www.philb.com/whichengine.htm
2.8E-4	triple_entendre	1.0E-4	http://inventory.overture.com/d/searchinventory/suggestion/
2.2E-4	cemper	9.9E-5	http://www.google.com/
1.7E-4	juanjo	8.8E-5	http://www.buzzle.com/editorials/6-10-2005-71368.asp
1.5E-4	konno	7.8E-5	http://findory.com/
1.4E-4	tomohiromikami	7.4E-5	http://www.betanews.com/
1.2E-4	relephant	7.3E-5	http://clusty.com/
1.2E-4	masaka	7.1E-5	http://cgi.cse.unsw.edu.au/collabrank/del.icio.us/

to find an objective measure of quality of clusters, as there does not exist any gold-standard for folksonomy clustering. This will be part of our future work.

Acknowledgement

This work has partially been supported by the European Commission within the research project ‘TAGora – Emergent Semantics in Online Social Communities’.

References

1. H. Ayad and M. S. Kamel. Refined shared nearest neighbors graph for combining multiple data clusterings. In M. R. Berthold, H.-J. Lenz, E. Bradley, R. Kruse, and C. Borgelt, editors, *IDA*, volume 2810 of *Lecture Notes in Computer Science*, pages 307–318. Springer, 2003.
2. G. Begelman, P. Keller, and F. Smadja. Automated Tag Clustering: Improving search and exploration in the tag space. *WWW2006, May*, pages 22–26, 2006.
3. P. Berkhin. Survey of clustering data mining techniques. Technical report, Accrue Software, San Jose, CA, 2002.
4. S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.
5. C. H. Brooks and N. Montanez. Improved annotation of the blogosphere via autotagging and hierarchical clustering. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 625–632, New York, NY, USA, 2006. ACM Press.
6. C. Cattuto, C. Schmitz, A. Baldassarri, V. D. P. Servedio, V. Loreto, A. Hotho, M. Grahl, and G. Stumme. Network properties of folksonomies. *AI Communications Special Issue on "Network Analysis in Natural Sciences and Engineering" (to appear)*, 2007.

7. E. Forgy. Cluster analysis of multivariate data: Efficiency versus interpretability of classification. *Biometrics*, 21(3):768–769, 1965.
8. P. Heymann and H. Garcia-Molina. Collaborative creation of communal hierarchical taxonomies in social tagging systems. Technical report, Stanford InfoLab, 2006.
9. A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In Y. Sure and J. Domingue, editors, *The Semantic Web: Research and Applications*, volume 4011 of *LNAI*, pages 411–426, Heidelberg, June 2006. Springer.
10. P. Mika. Ontologies are us - a unified model of social networks and semantics. In *Proceedings of the 4th International Semantic Web Conference (ISWC)*, 2005.
11. C. Schmitz, A. Hotho, R. Jäschke, and G. Stumme. Mining association rules in folksonomies. In *Proc. IFCS 2006 Conference*, Ljubljana, July 2006.
12. P. Schmitz. Inducing ontology from flickr tags. In *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland*, May 2006.

Integration and Semantic Enrichment of Explicit Knowledge through a Multimedia, Multi-source, Metadata-based Knowledge Artefact Repository

Felix Mödritscher

(Institute for Information Systems and Computer Media,
Graz University of Technology, Austria, fmoedrit@iicm.edu)

Robert Hoffmann

(Institute for Information Systems and Computer Media,
Graz University of Technology, Austria, rhoff@iicm.edu)

Werner Klieber

(Know-Center Graz, Austria,
wklieber@know-center.at)

Abstract: Explicit knowledge is often stored in various repositories within a company. As APOSDLE aims at considering the whole intellectual capital of companies, we are coping with the task to integrate the content of several and different repositories within the APOSDLE platform. Further, other modules in the APOSDLE system also require functionality to semantically enrich knowledge artefacts like documents or multimedia objects. In this paper, we present our approach to these challenges, namely the “Multimedia, Multi-source, Metadata-based Knowledge Artefact Repository”. After presenting the basic idea and critical issues of this solution, we point out related technological approaches as well as their limitations. Finally, we report about a first prototype realising this idea and our experiences gained so far.

Keywords: Digital Object Repository, Data Layer, Knowledge Artefact, Metadata Standards

Categories: H.2.4, H.2.5, H.2.8, H.3.2, H.3.7, H.5.1

1 Introduction

Companies face the problem of continuous growth and increasing amounts of knowledge relevant for their core businesses [Borghoff and Pareschi, 97]. Particularly, explicit knowledge, by means of reports, articles, manuals, patents, images, video, software, and the forth, underlies a high degree of fluctuation. As key resources are often hidden in different heterogeneous knowledge repositories [Dzbor et al., 00], many companies already focus on managing and integrating these repositories into a knowledge management system in order to provide access their contents [Kühn and Abecker, 97].

One of the key issues in the APOSDLE project [APOSDLE, 07] comprises the idea to consider all parts of a company’s intellectual capital and make it available to the knowledge workers. On the one hand, implicit knowledge should be accessible via tools recommending experts within the working context and supplying collaborative features to support knowledge exchange. On the other hand, the documents relevant for a certain task should be captured by the APOSDLE platform and utilised by means of different application scenarios, like providing the knowledge workers with these resources, embedding them into learning events or making discussions possible.

Against this background, this paper introduces our idea of a “Multimedia, Multi-source, Metadata-based Knowledge Artefact Repository” (M3KAR), which ought to be part of APOSDLE’s underlying knowledge infrastructure and deals with managing so-called Knowledge Artefacts. In context of the APOSDLE system, M3KAR aims at providing the functions which allow other modules to semantically enrich digital resources and exploit this semantics. In the following the basic concept of our solution approach is described and aspects relevant for practice are outlined. Thereafter, an overview of related systemic types is given, and differences to M3KAR are highlighted. Finally, the implementation of our idea within the first APOSDLE prototype is described, and experiences are summarised.

2 The Basic Concept of M3KAR

As mentioned before, our task in the APOSDLE project is to guarantee a homogeneous access to all documents stored in different repositories within a company. Further, different modules and tools within the APOSDLE platform should also have the possibility to retrieve these documents on basis of a unique identifier, add or modify semantics to the digital objects and utilise this meta-information or the content for their purposes. Therefore, two important issues have to be considered: (1) the Knowledge Artefacts themselves and (2) the systemic architecture of M3KAR.

2.1 Knowledge Artefacts and Their Lifecycle

Our first definition involves the so-called Knowledge Artefact (KA) which can be understood as a piece of (digital) information relevant for a certain working context and enriched with semantic information in terms of metadata (see [Ley, 06]). Regarding related fields like standardisation of digital objects (e.g. with the Dublin Core Metadata Element Set) or content management, Knowledge Artefacts underlie a lifecycle within the APOSDLE platform, as shown in Figure 1.

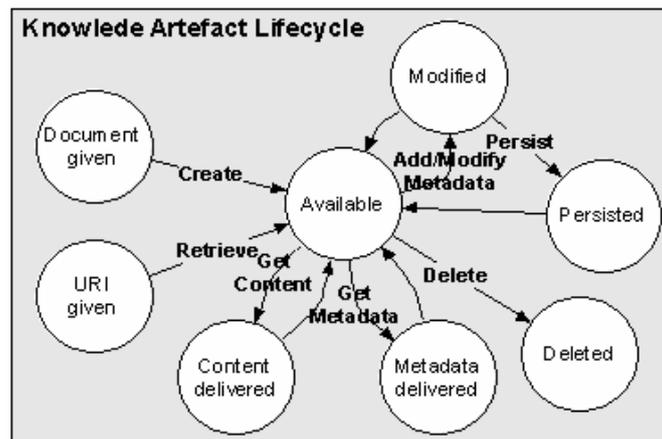


Figure 1: The Knowledge Artefact Lifecycle

Concerning a company's intellectual capital, the following scenarios are of relevance for M3KAR. On the one hand, Knowledge Artefacts might be available in some repository and, therefore, might be retrieved by unique identifiers. On the other hand, KAs can be also created within certain APOSDLE scenarios, e.g. during learning or collaboration. After retrieving or creating a KA, it is made available for other systemic components. The tools and modules can now access the content and metadata of a KA, add or modify attributes or even delete it. As a consequence of this lifecycle, the following six issues are of importance for realising a Knowledge Artefact Repository:

1. Due to the distributed architecture of APOSDLE, a Knowledge Artefact can be considered as unique within the whole platform. Addressing design patterns, [Pant and Ondo, 02] outline the special challenges of the Singleton pattern in distributed environments.
2. The main application scenario for M3KAR comprises the retrieval of Knowledge Artefacts, whereby high-level retrieval methods are provided by another module. Thus retrieval is restricted to the principles of data retrieval, e.g. by retrieving KAs according to their identifier or supplying fuzzy queries on certain attributes to locate the relevant document within the integrated repositories. Therefore, KAs must provide a set of system-managed and read-only attributes, such as a unique URI (Uniform Resource Identifier), in order to be accessible. Additionally, certain metadata fields have to be searchable.
3. Furthermore, other components should be able to add or modify metadata attributes of Knowledge Artefacts. Besides, it is intended that some modules might even create whole KAs within the platform. Thus, changes of the KA need automatically to be persisted. Moreover, also the modifications of the original content within the repositories have to be detected and regarded in order to provide accurate information to the knowledge workers.
4. Concerning privacy, KAs underlie a role-based access control mechanism [Sandhu et al., 96], which, in our case, is primarily based on the permissions within the document repositories integrated into the APOSDLE platform. Therefore, the access permissions have to be resolved from these repository systems, e.g. by utilising mapping techniques.
5. Another requirement of the APOSDLE approach deals with the type and granularity of resources to be described by metadata attributes. Precisely, APOSDLE has to support multimedia content objects, but also virtual objects or parts of documents. Further, the definition of relations to other KAs is necessary.
6. Finally, M3KAR should also tie up to interoperability aspects, i.e. by being capable of supporting existing standards in the field of digital object repositories or even allowing the export of digital objects together with a standard-based description.

Overall, these aspects derived from the Knowledge Artefact lifecycle lead to four dimensions of metadata sets relevant for our purposes: (1) system-managed (read-only) vs. user/application-managed, (2) searchable vs. non-indexed ones, (3) altering vs. constant ones, and (4) temporary vs. persistent ones. In any case, one achievement of the APOSDLE approach is to automatically generate metadata – no matter of what

category. Thus, system-managed attributes, like the URI or permissions, should only be created and updated by M3KAR. Other attributes might automatically be set and modified by other modules or tools.

2.2 Functionality and Architecture of the Knowledge Artefact Repository

Beside the depicted aspects of Knowledge Artefacts, the focus of our approach addresses the architectural design as well as the technical realisation of M3KAR. The component itself represents the data layer of the APOSDLE platform, as visualised in Figure 2. Yet, structural and semantic information is provided by other modules, i.e. by the Semantic Repository Manager, the Classification Service or the Semantic Service. The primary task of the Knowledge Artefact Repository comprises the management of Knowledge Artefacts and their provision to the other components of the APOSDLE platform.

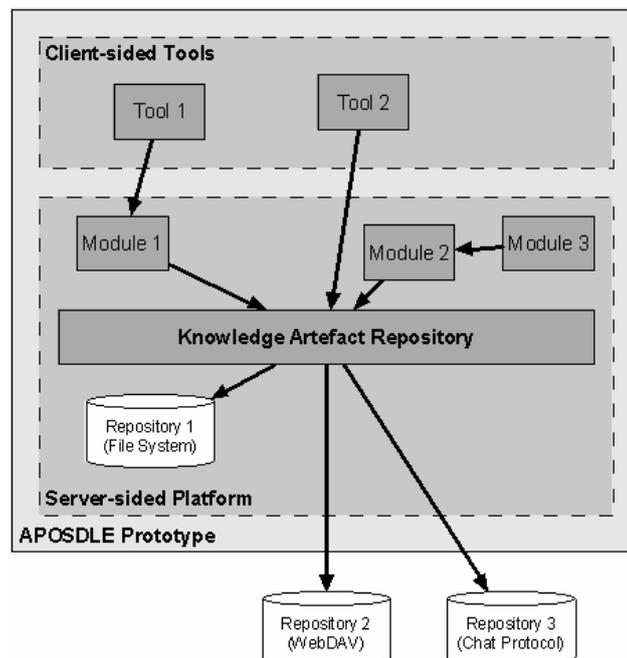


Figure 2: The Concept of a Knowledge Artefact Repository

Figure 2 shows the basic functionality of the M3KAR idea, which deals with the following requirements:

1. A so-called Repository Manager manages the repositories which are integrated into M3KAR. These repositories might be located in or outside the platform, whereby different types of repositories (File System, WebDAV, Chat Protocol) are supported. Technologically, the Product Factory design pattern [Metsker and Wake, 06] is applied within the Repository Manager. As a first step, all necessary operations on a repository

are defined by an interface class. For each repository type, an own implementation of this interface is required. Finally, the repositories can be initialised by a configuration file, but also instantiated on-the-fly without restarting the platform.

2. The repository implementations also support the management of documents on-the-fly, so that documents can be inserted via the platform and, subsequently, new Knowledge Artefacts are made available within the APOSDLE platform. Moreover, also the modification and deletion of KAs are allowed, as already stated in the last subsection. Metadata as well as content imported by other components need to be stored within an own, internal data repository, i.e. a database management system.
3. Due to these first two issues, a notification service is required, so that other modules are informed about changes of and within the repositories. For instance, the Classification Service should be notified in order to initialise the knowledge extraction process and guarantee an accurate index for high-level retrieval functions.
4. Addressing data integration, we face the problem that a digital entity (like a document) can be retrieved from different repositories. As a first and efficient approach to this problem, we propose the calculation of a checksum for digital artefacts and compare their contents, if the checksum is equal. A more sophisticated solution would require semantic comparison of the content in order to detect different versions of artefacts.
5. The logging of the events within M3KAR is important for different reasons, e.g. for examining the usage of KAs in some application scenario or for improving the performance of the overall platform.
6. Finally, performance issues might be of interest, not only for M3KAR, but also for the overall APOSDLE system. Thus, statistical or probabilistic methods [Przybylski, 90], e.g. on basis of access statistics logged within M3KAR, could be utilised to cache the content of KAs which have been very often requested or are predictably of relevance.

Each single aspect is well documented in literature and, further, a lot of systemic types exist in this context. Yet, none solution exactly fulfils the requirements given by the APOSDLE research project, as highlighted in the upcoming section.

3 Overview and Limitations of existing Solution Approaches

Amongst a broad range of solution approaches, relevant types of data layer approaches are inspected in order to outline the differences to the M3KAR idea.

The first category of solutions to be mentioned in this context comprises database management systems itself, which, evidently, lack of the object model for providing operations to other components. Thus, all methods required within the APOSDLE platform must be mapped to a relational database model, which might cost performance and complicates the realisation of other concepts, like caching or session-based metadata. Hence, database technology is necessary for M3KAR no matter whether to integrate a company's explicit knowledge or to manage data objects internally, but it is not sufficient for realising the data layer of APOSDLE.

Secondly, middleware approaches and GRID environments focus on the management of digital objects and their description with metadata, as stated by [Stockinger et al., 01] or with respect to concrete examples like Dinopolis [Haselwanter, 03]. Particularly, the Dinopolis Object supports the issues concerning the KA-lifecycle, while this middleware system is not fully compliant to the server-sided technologies utilised in the APOSDLE project, namely the Spring Framework. In context of M3KAR, an object persistence layer is advantageous and covers many of our requirements on a Knowledge Artefact Repository. Therefore and due to the excellent technological compatibility to the Spring Framework, we decided to apply the Hibernate object persistence layer.

Thirdly and on a higher abstraction level, document and content management systems (and other kinds of knowledge management systems) include such data layer approaches. For instance, the Hyperwave Information Server [Hyperwave, 07] supplies features to manage digital objects and describe them with metadata. Moreover, it is also possible to integrate external data sources by indexing their contents with the full-text indexing engine. Thus, it is possible to make external resources accessible. Furthermore, the Hyperwave Information Server also provides an internal Java-based API to create and manage such Knowledge Artefacts in the way depicted in the last section. However, the commercial character and the functional overheads of such a solution are strong arguments against the utilisation of full-feature products in this scope. Even open source solutions, if customisable to our needs, can not be considered as applicable for the data layer of APOSDLE.

Finally, two research approaches related to M3KAR have to be outlined here: One the one hand, [Kebbell and Campbell, 04] report about a framework which is capable of managing digital objects and their metadata. Although the overall idea is rather similar to M3KAR, it is realised for a small area of a digital library infrastructure, including only one (internal) object repository and not being made available for other projects. On the other hand, [Eriksson and Bang, 06] present a document repository based on semantic documents. Again, this approach addresses only one repository and neglects problems like data integration.

In conclusion of the afore-mentioned, the necessity for realising the idea of the multimedia, multi-source, metadata-based knowledge artefact repository is absolutely given within the APOSDLE research project.

4 Current Implementation of M3KAR and Experiences Gained

At this stage, some aspects of M3KAR have been implemented as a part of the first APOSDLE prototype. This initial version of the M3KAR has been named "Homogeneous Access" (HA) and already realises some of the aspects described in section 2. Yet, this first version of M3KAR still is restricted in many ways:

- Currently, the creation of KAs is not supported. Thus, the KA lifecycle only comprises the retrieval of KA from existing repositories on basis of the URI and the adding of semantic information by means of metadata. System-driven attributes are created on the first access, while all kinds of metadata can be modified. The persisting of changes is not possible, as the first APOSDLE prototype only represents a demonstrator. Consequently, it is not

necessary to deal with the Singleton pattern. KAs are created as multi-state objects, whereby each module uses an own instance of each KA.

- Moreover, HA neither supports KAs on different granularity levels nor multimedia artefacts. Access control on the basis of the permissions from the repositories is also not implemented so far, while access to Knowledge Artefacts is always granted, if a repository is included into the APOSDLE platform. Standard-compliance was non-issue for the first prototype at all.
- The first version of M3KAR already allows the integration of multiple repositories, which was the main objective of Homogeneous Access. At this stage, three types of repositories are supported: (1) file system, (2) WebDAV, and (3) Chat protocols of the Concert Chat Server (the collaboration tool of APOSDLE).
- All repositories have to be configured before start-up of the platform in order to utilise their content. Adding new repositories is not yet supported. The contents of the repositories are being captured once at the system start-up, as a module of the APOSDLE platform, namely the Classification Service, browses through all repositories and indexes their documents. From this moment on, the KAs can be accessed by other components.
- Logging is done to the debug console for debugging reasons only. An own log mechanism for further analysis of the usage of KAs and its exploitation for other purposes is not realised.
- Finally, performance is not considered at all, although it would be an important issue – not only for M3KAR, but also for the overall usability of the APOSDLE system. Concrete performance tests indicated that it took only about one second to retrieve 1000 KAs without content, while it lasted up to 3.5 minutes to deliver a KA including a document with a size of about 1MB thousand times in a row. In this context, it is recommended to apply at least statistical caching methods to improve the systemic performance for multiple deliveries of contents.

As shown with this first version of M3KAR, some of the concepts depicted in section 2 are already realised, others might be of interest or even necessary for the next APOSDLE prototype. Particularly, the persistence of KAs and further aspects of their lifecycle as well as caching approaches are required to guarantee an efficient, usable APOSDLE system.

5 Conclusions and Future Work

In summarising this paper, we have to state that our idea of a “Multimedia, Multi-source, Metadata-based Knowledge Artefact Repository” represents not only a data-layer for APOSDLE but also comprises an advantageous approach to integrate and semantically enrich a company’s explicit knowledge which might be distributed over several repositories. Moreover, M3KAR is a key component for a usable APOSDLE platform. By now, we have implemented a first version of M3KAR, examined it and identified necessary or even critical issues to be considered within the next APOSDLE prototype. On that account, we have already started to plan and implement these new features.

Acknowledgements

APOSDLE is partially funded by the FP6 of the European Commission within the IST work programme 2004 (FP6-IST-2004-027023).

References

- [APOSDLE, 07] APOSDLE, “learn@work”, Project Website, 2007, <http://www.aposdle.tugraz.at/> (2007-05-16).
- [Borghoff and Pareschi, 97] U.M. Borghoff, and R. Pareschi, “Information Technology for Knowledge Management”, *Journal of Universal Computer Science*, 3(8), 1997, pp. 835-842.
- [Dzbor et al., 00] M. Dzbor, J. Paralic, and M. Paralic, “Knowledge Management in a distributed organization”, Technical Report KMI-TR-94, Knowledge Media Institute, Open University, 2000.
- [Eriksson and Bang, 06] H. Eriksson, and M. Bang, “Towards Document Repositories Based On Semantic Documents”, *Proceedings of International Conference on Knowledge Management*, 2006, pp. 313-320.
- [Haselwanter, 03] E. Haselwanter, „Aspects of component composition in distributed frameworks”, master’s thesis, Graz University of Technology, 2003.
- [Hyperwave, 07] Hyperwave, “Company’s Website”, 2007, <http://www.hyperwave.com/e/> (2007-06-16).
- [Kebbell and Campbell, 04] A. Kebbell, and D. Campbell, “Managing digital objects and their metadata: challenges and responses”, *Proceedings of International Conference on Dublin Core and Metadata Applications*, 2004.
- [Kühn and Abecker, 97] O. Kühn, and A. Abecker, “Corporate Memories for Knowledge Management in Industrial Practice: Prospects and Challenges”, *Journal of Universal Computer Science*, 3(8), 1997, pp. 929-954.
- [Ley, 06] T. Ley, “APOSDLE Glossary: Workplace Learning, Work-integrated Learning”, Know Center, 2006, http://www.aposdle.tugraz.at/weblog/aposdle_glossary_workplace_learning_work_integrated_learning (2007-05-16).
- [Metsker and Wake, 06] S.J. Metsker, and W.C. Wake, “Design Patterns in Java”, 2nd edition, Boston: Pearson, 2006.
- [Pant and Ondo, 02] Y. Pant, and K. Ondo, “Thread Specific Singleton: Handling singleton pattern errors in multi-threaded applications and their variations”, *Journal of Object Technology*, 1(2), 2002, pp. 155-169.
- [Przybylski, 90] S.A. Przybylski, “Cache and Memory Hierarchy Design: A Performance Directed Approach”, San Francisco: Morgan Kaufmann Publishers, 1990.
- [Sandhu et al., 96] R.S. Sandhu, E.J. Coyne, H.L. Feinstein, and C.E. Youman, „Role-Based Access Control Models”, *Computer*, 29(2), 1996, pp. 38-47.
- [Stockinger et al., 01] H. Stockinger, O.F. Rana, R. Moore, and A. Merzky, “Data Management for Grid Environments”, *Proceedings of International Conference on High-Performance Computing and Networking*, 2001, pp. 151-160.

KVD'07- Special Track on Knowledge Visualisation and Knowledge Discovery 2007

Vedran Sabol

(Know-Center Graz , Austria
vsabol@know-center.at)

Wolfgang Kienreich

(Know-Center Graz, Austria
wkien@know-center.at)

Michael Granitzer

(Know-Center Graz, Austria
mgrani@know-center.at)

Keith Andrews

(Graz University of Technology, Austria
kandrews@icm.edu)

Many researchers in the field of Knowledge Discovery and Knowledge Visualisation are aware of the indicators on information growth provided in the study “How Much Information 2003”¹ conducted by the University of California at Berkeley. In early 2007, a new study conducted by IDC² provided new numbers on information growth for 2006 and estimates for 2010. The conclusions of this study, which appropriately was titled “The Expanding Digital Universe”, are easily summarized: The amount of digital information surrounding us continues to grow exponentially.

In 2003, the amount of stored digital information was estimated to be 5 Exabyte. A real-world analogy equated this to a ten meter high pile of books per human being on earth. In 2006, a new real-world analogy had to be found: Each human being now owned around six tons of books, an amount barely manageable by an everyday delivery truck, as the amount of stored digital information has risen to about 161 Exabyte. In 2010 – only 4 years from now – IDC estimates that we will face about 6 times more information, namely 998 Exabyte.

So why bother with those numbers? The estimated amount of information points towards the development of new technical solutions as well as the application of existing algorithms in new scenarios to allow humans to manage parts of the digital universe instead of getting lost in a black (information) whole. Today’s tools are

¹ <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>

² http://www.emc.com/about/destination/digital_universe/pdf/Expanding_Digital_Universe_IDC_WhitePaper_022507.pdf

barely able to scale with the amount of information, yielding to decreased usability and reduced efficiency of knowledge workers.

New intelligent and intuitively usable techniques for analysing and mining the digital universe are required in order to provide future proofed work places. Besides intelligent, semantically enhanced algorithms and techniques, provided by Knowledge Discovery, Data Mining and Information Retrieval techniques, new interaction techniques and presentation paradigms are required to bridge the growing gap between user and machine and ultimately between user and user.

Our special track on Knowledge Visualization and Knowledge Discovery addresses the emerging fields of knowledge and information visualization, knowledge discovery and information retrieval. We focus on the discovery, interaction with and visual presentation of abstract knowledge and information spaces, with the goal to facilitate the rapid assimilation and understanding of knowledge items and to support communication between users. Over the last three years, we have experienced a significant increase in interest, which manifested in 21 submissions this year, wherefrom 12 high quality papers have been accepted for presentations. The papers address a broad range of topics ranging from plagiarism analysis and text mining to new visualisation algorithms and visualisation of semantic structures.

The following papers have been accepted for presentation and publication:

- *Usability of Diagrams for Group Knowledge Work: Toward an Analytic Description*, Sabrina Bresciani, Martin J. Eppler
- *Visual tools decipher historic artefacts documentation*, Iwona Dudek, Jean-Yves Blaise
- *Spatiotemporal Knowledge Visualization and Discovery in Dynamic Social Networks*, Ralf Klamma, Yiwei Cao, Marc Spaniol, Yan Leng
- *GlobeMash: A Mashup for Accessing GLOBE*, Joris Klerkx, Erik Duval
- *Text Mining for Indication of Changes in Long-Term Market Trends*, Petr Kroha, Thomas Reichel, Björn Krellner
- *Sky-Metaphor Visualisation for Self-Organising Maps*, Rudolf Mayer, Khalid Latif
- *Printf in 4D user interfaces*, Alexander Oberbichler, Laszlo Böszörmenyi
- *An Interactive Visualization Model for Competence Management: an Integrative Approach*, Lena Pareto, Ulrika Lundh Snis
- *Visualizing Semantic Structures to Support Reading* Yongqiang Qiu
- *On the Generalized Box-Drawing of Trees*, Benno Stein, Frank Benteler
- *Visualization of Rule Bases - The Overall Structure*, Valentin Zacharias
- *Service oriented information supply model for knowledge workers*, Bilal Zaka, Hermann Maurer

The submissions have been peer reviewed by an international program committee, namely:

- Jean-Yves Blaise (UMR CNRS/MCC MAP 694 gamsau, Marseille)
- Remo Burkhard (Competence Center Knowledge Visualization, University St. Gallen)
- Yiwei Cao (Aachen University)
- Mario Döllner (Universität Passau)
- Victor M. García-Barrios (Graz University of Technology)
- Michael Grabner (Graz University of Technology)
- Christian Gütl (Graz University of Technology)
- Katharina Kaiser (Vienna University of Technology)
- Marián Képesi (Graz University of Technology)
- Ralf Klamma (RWTH Aachen)
- Joris Klerkx (Katholieke Universiteit Leuven)
- Werner Klieber (Know-Center Graz)
- Mark Kröll (Know-Center Graz)
- Bruno Pinaud (Université de Nantes)
- Benno Stein (Universität Weimar)
- Simone Stumpf (Oregon State University)
- Albert Weichselbraun (Wirtschaftsuniversität Wien)
- Gerhard Wohlgenannt (Wirtschaftsuniversität Wien)

The organizers hereby thank the program committee for their valuable reviews and remarks. Without their qualified judgments we would not be able to achieve such a high special track quality.

Michael Granitzer, Vedran Sabol, Wolfgang Kienreich and Keith Andrews

Graz, September 2007

Printf in 4D User Interfaces

Alexander Oberbichler

(University Klagenfurt, Austria
alex@itec.uni-klu.ac.at)

Laszlo Böszörményi

(University Klagenfurt, Austria
laszlo@itec.uni-klu.ac.at)

Abstract: The visualization of temporal information should not be seen as a special case. A lot of applications take advantage of the time factor (e.g. capturing user events). In general recorded activities on a workstation can be seen as a temporal database and so can profit from a time-based visual output system too. We introduce a model which classifies the time as a "first class citizen" in today's operating systems. Such an invention would not only improve and standardize the capturing of temporal events but it would be of benefit for a temporal visualization system too. Within a 3D visualization space and a global time axis we introduce a *printf4D()* method. With this method it is possible to display images, videos, text strings or any other kinds of information in an automatically animated way. As a first proposal we demonstrate this function in a "flow of information" metaphor where information-objects are not displayed all at once but in a flowing manner over a certain period of time. Additionally we will show that *printf4D()* is not limited to temporal data. It can be extended automatically to any kind of static information.

Key Words: 4D UI, animation, global time axis, flow of information metaphor

Category: H.4.3, H.5.1, H.5.2, I.3.7, I.3.8

1 Introduction

The well known face of today's Graphical user interfaces (GUI) will change in the near future. Approved 3D technologies are embedded in the process of human computer interaction (HCI) more frequently. Sun's Looking Glass [Lg3d, 06], Microsoft's Task Gallery [Robertson, 00] or Novell's Xgl 3D Desktop are only some popular representatives of the future UI generation. For the croquet project [Smith, 04] led by Alan Kay the future working place is more a kind of virtual reality where every user can see the current working steps of the colleagues in the network.

In that relation the visualization of time based information is an important issue. Projects like Time-Tube [Chi, 98], Life-Lines [Plaisant, 96] or the well known Perspective-Wall [Mackinlay, 91] introduce good methods to present temporal information on the monitor. Such techniques should not only be used for the visualization of temporal databases. The power of 3D presentation in combination with temporal visualization techniques can be of benefit for many

applications. Project checkpoints, the next service dates of the car or even the workstation-tasks are possible information which can profit from a so called 4D UI. Not time critical knowledge like images in a photo gallery can profit from an animated visualization system too. For that reason the time factor is essential for every content. We introduce a system wide global time axis to categorize the content with time-information. Based on this global temporal representation the resulted visualization will depend on the time factor very strongly.

A lot of applications take already care of the time dimension (e.g. recording of user-inputs or -activities). But without a global and unique time system to record such events every application has to implement a proprietary solution. The time dimension should be embedded as a kind of "first class citizen" in today's operating systems. This enables an easy and unified exchange of temporal information between applications. Even if the programmer is not interested in the 4D visual output system it would be possible to take advantage of the generally available global time axis.

A core function of our 4D visualization framework will be a *printf4D()* method. With it's functionality it will be possible to display an unlimited set of information in an animated way over a certain period of time.

2 4D framework

Basis for our 4D framework is a global time axis and this must fulfil certain properties. Above all, 2 points have to be considered. On the one hand to abstract all kinds of time based events, the time model must support different granularity levels. On the other hand, the time model should be able to manage different lifelines concurrently because in a real 4D environment every created object has an own history of events and so an own lifeline. We rely on the multi dimensional and multi granularity time model of Keravnou [Keravnou, 99]. Although it was developed for the medical sector it is designed abstractly enough to fit all requirements for a 4D global time axis.

Based on this time model we have developed a programming interface to ease the development of 4D applications. One could argue that the temporal aspect could be simply implemented on the top of OpenGL or based on frames as in FLASH. However, this method would expand the whole software development process considerably. Temporal aspects should be integrated into a single framework as much as possible. As can be seen in figure 1, the programmer has only to distinguish which information is displayed over a period of time (encapsulated in a class we called dynamic page) or "forever" (static page). We describe our 4D framework in our technical report [Oberbichler, 07] in more detail. We call the whole programming layer a book-abstraction. This should serve as an intuitive help to think of writing a book with dynamic- and static-pages in the whole 4D

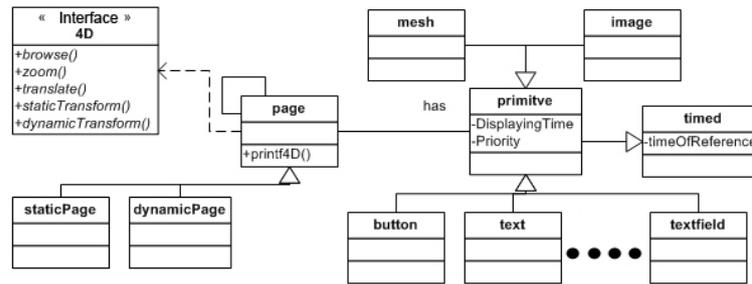


Figure 1: Section of 4D book abstraction

development circle. Main focus for the temporal visualization is the way how the dynamic-page class is implemented. The idea of this class is that like printing text on the monitor it should be possible to render any kind of contents in an animated way over a certain period of time. In the following chapters we will discuss the preconditions and mode of action of such a *printf4D()* method.

3 Dynamic page and time point critical data

If an animated output is needed, the corresponding content (*primitive*; see figure 1) has to be added to a dynamic page. It is only necessary to specify how many primitive objects are displayed concurrently.

Every primitive has a connection to the global time axis (the corresponding parameter is called *TimeofReference* which specifies when the state of an object changes. For instance an appointment in a calendar becomes obsolete. Such an information is time point critical.), a *DisplayingTime* and a *Priority*. With these values it is possible to adjust the result of the *printf4D()* method.

3.1 Rendering dynamic-pages

Starting from the situation where every object has a time of reference on the global time-axis and a priority level, we are concerned of visualizing these temporal connections.

In our 4D framework the visualization part depends on the actually used book-implementation and in temporal aspects on the dynamic page implementation. It is possible to realize this part with well known temporal visualization techniques (e.g. Perspective wall, Time Lines, Time Tube, ...). As our main focus is the temporal (or animated) presentation of any kind of content we talk about a "flow of information" metaphor. The time and so our visualization is in continuous motion. Figure 2 demonstrates this analogy by rendering the time as lotic water. From the left side future events (e.g. dates in a calendar) are running

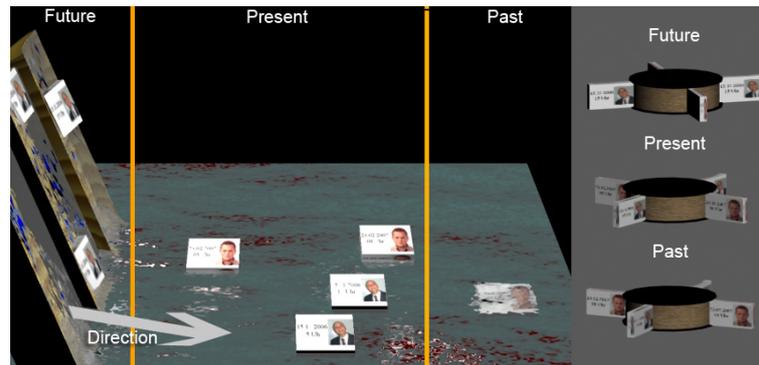


Figure 2: Flowing information metaphor and rotating cylinders

down a waterfall until they reached the lake at the bottom which presents the near present. At the right end of the monitor objects are disappearing when they are becoming "too" old (this can be animated as a sinking effect).

In general the time references of the given primitive objects are not in a constant distance to each other. Also the number of elements in the three temporal sectors can differ very much. For that reason we have to consider 2 points:

1. Not all objects should be displayed concurrently. Each of the three time-sectors should use *printf4D()* methods to show only a cut-out. Since the displayed objects get replaced by other ones every object will be displayed anyway.
2. If we display e.g. 10 objects in the present-sector with nearly the same time-references these objects would swim very close to each other in the lake. To prevent such a phenomenon it is necessary to scale the displayed time-spans appropriate for each sector (e.g. the future-sector displays the next 2 days meanwhile the present-sector displays ± 5 minutes from the present moment of time). This is also necessary to guarantee a viewable motion of the primitive objects.

The flowing information visualization is not the only way to render dynamic pages. Our model is abstract enough to define the content and the presentation once and to render it with different book implementations. We can e.g. inherit from a kind of "cylinder book" implementation which renders dynamic pages as rotating cylinders (figure 2 right side). After a certain period of time appointments are handed over to the next "time sector" cylinder.

4 Adding static information on the global time-axis

Static pages represent the classical visual viewing concept (displaying every content at once without any animated effects). To generate such an output the number of maximum concurrently displayed objects and the displaying time for every primitive must set to ∞ . In that case the *printf4D()* implementation has not to be changed. In static pages the *TimeofReference* parameter is of no relevance, all objects are remaining in their original positions.

A transformation from a static- to a dynamic-page can be useful to animate a set of information automatically. For that reason every primitive object has to be equipped with a *TimeofReference* and so connected to the global time axis. One possibility is to use the creation time of every primitive but thereby we will have a system full of past events. That is not the wished behavior in the most cases. Our aim is that our visualization system distinguishes between time point critical and uncritical data. For static objects the global time axis has to be seen as programming interface for the resulted animation. If every object gets a fixed *TimeofReference* they will become older at a certain time, and so the problem to have a set of past events is only delayed. So we introduce a period length for all static elements. After this time the objects are shifted on the global time axis (e.g. if an image is displayed for a specific time it will be moved to a future time point depending on the period length. This process guarantees a continually animated presentation of future and past objects). For time point critical data like appointments in a calendar this process would not make sense.

In the following we discuss possibilities to generate *TimeofReferences*. Starting point for this automation is the order how primitive objects are added to the pages. To categorize the added objects a priority level is assigned too.

1. Incremental:

All primitive objects are of the same value and get a constant time point on the time axis in equidistant offsets, equal period lengths and priority levels. This is of course the easiest method. But for many applications it is not important to distinguish between the displayed objects (e.g. animated gallery which shows a set of pictures over a certain period of time).

2. Order dependent:

Quite often important elements are defined before other ones. In that case it would be possible to generate priority levels automatically relying on the order of creation. So first added objects should get higher priorities than later added ones. This task can be achieved with many functions, for instance the $\frac{1}{t}$ function:

$$\forall t \geq 1 : \textit{priority}(t) = a \cdot t^{-1} \quad (1)$$

When all primitive objects are saved in a queue the variable t specifies the index of the object. So first added primitive objects get a higher priority p than later added ones.

3. Random:

A set of information which should be presented in any order and in any intervals should be categorized by a random function. In comparison to the incremental method the random function results in a more realistic presentation of the primitive objects.

5 Display time for every object

The *Printf4D()* method does not show all objects at once in dynamic pages. After a certain period of time some objects are exchanged by other ones. But for that reason we have to distinguish how long every object is displayed. Of course the programmer can specify the time value manually for every object but it is also possible to automate this process.

– Weighted function:

A priority level is assigned to every primitive object which is added to a dynamic page. Based on this value it is possible to calculate a weighted displaying time for every object.

$$p \geq 0 \wedge p \leq 1 : d_{displaytime}(p) := t_{min} \cdot (p + 1) \quad (2)$$

The only parameter one should be specified manually is a minimum displaying time (t_{min}). This is necessary to prevent an exchange of animated objects after an unacceptable short time. The parameter p is the priority of the displayed object. So this weighted function would double the displaying time for important objects at the worst case.

– *TimeofReference* dependent:

Objects which *TimeofReference*-values are close to the actual moment of time are preferred in getting more displaying time.

– Constant or random functions:

If the displaying time is not of relevance for the aimed application a constant value can be assigned for every primitive object. It is also quite straight forward to use random functions for a diversified presentation.

6 Test implementation

We implemented the discussed time- and abstract book-model in C++. As 3D graphic engine we use the open source library Ogre3D [Junker, 06]. As a first

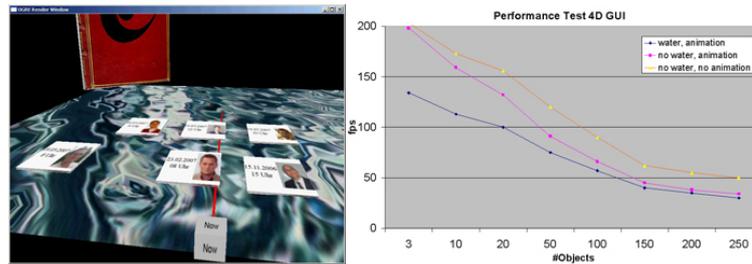


Figure 3: 4D calendar and performance test

test application we implemented a simple 4D calendar (figure 3) which uses the properties of the dynamic page class to display all saved appointments with the *printf4D()* method. The appointments are rendered as a low polygon cubes with a mapped texture and specific textual information about the date.

Among other things we have tested the performance of this sample application on a low-budget system. Between 150 and 250 appointments can be animated concurrently without overloading the system [Oberbichler, 07]. Today's CPUs and GPUs are capable to realize simple translation techniques which are required for a *printf4D()* method. We also got a first impression how the *printf4D()* can support the user. It is not a trivial task to display 250 appointments in a calendar clearly arranged. The power of the temporal output can be very useful for different kinds of applications.

In listing 1 it is possible to see a sample code how to use our book-abstraction.

Listing 1: "4D gallery" Sample-Code

```

1 Samplebook app;
2 Sampledynpage *mDPPage = new Sampledynpage("testpage");
3 for(int i=0; i < ARRAYLENGTH; i++)
4 {
5     mDPPage->staticadd(new Image(path[i]), 4D::RANDOM);
6 }
7 app.add(mDPPage);
8 app.open();
9 ...
10 app.pages("testpage").printf4D();

```

In general it is possible to create an animated picture gallery with these basic lines. All images are added to a dynamic page by a random algorithm and can be rendered with the *printf4D()* function call afterwards.

7 Conclusion

We introduced a basic print method in 4D user interfaces. Our main aim is to visualize an infinite set of primitive objects over a certain period of time. Based on a system wide time axis static- and dynamic-objects are categorized by priority levels which allows calculating different displaying times for more important objects. We demonstrated the *printf4D()* method on "flow of information" metaphor which is rendered as lotic water. Basic translation operations are not critical for today's PCs. Considering the exploding volume of information we have to find methods which do not only present this information but confront the users with possibly forgotten things too. *printf4D()* animates temporal data in a sequential way and static information as a user reminder as well.

References

- [Chi, 98] Chi E. H., Pitkow J., Mackinlay J., Piroli P., Gossweiler R., Card S. K.: "Visualizing the evolution of web ecologies"; Proc. of the Conf. on Human Factors in Computing Systems, CHI, Los Angeles (1998) 400-407.
- [Junker, 06] Junker G.: "Pro OGRE 3D Programming"; Springer-Verlag GmbH (2006)
- [Dyck, 03] Dyck J., Pinelle D. Brown, B., Gutwin C.: "Learning from games: Hci design innovations in entertainment software"; Proc. 2003 Conf. On Graphics Interfaces (GI'03), Halifax, Montreal (2003)
- [Goralwalla, 01] Goralwalla I. A., Leontiev Y., Ozsü M. T., Szafron D., Combi, C.: "Temporal granularity: Completing the puzzle"; Journal of Intelligent Information Systems, 16, 1 (2001), 41-63
- [Lg3d, 06] Project looking glass:lg3d - a comprehensive overview, 2006, <https://lg3d-core.dev.java.net>
- [Keravnou, 99] Keravnou E. T.: "A multidimensional and multigranular model of time for medical knowledge-based systems"; Journal of Intelligent Information Systems, 13, 1-2 (1999), 73-120
- [Mackinlay, 91] Mackinlay J. D., Robertson G. G., Card S. K.: "The perspective wall: detail and context smoothly integrated"; Proc. of the SIGCHI conference on Human factors in computing systems, ACM Press, New York (1991), 173-176
- [McFarlane, 99] McFarlane D.: "Coordinating the interruption of people in human-computer interaction"; HumanComputer Interaction - INTERACT, IOS Press Inc., The Netherlands (1999), 295-303
- [Oberbichler, 07] Oberbichler A., Böszörményi L.: "Framework for 4d user interfaces"; Technical Report TR/ITEC/07/2.01, Institute of Information Technology, University Klagenfurt (2007)
- [Palmiter, 91] Palmiter S., Elkerton J.: "An evaluation of animated demonstrations of learning computer-based tasks"; Proc. of the SIGCHI conf. on Human factors in computing systems, ACM Press, New York (1991), 257-263
- [Plaisant, 96] Plaisant C., Milash B., Rose A., Widoff S., Shneiderman B.: "Lifelines: Visualizing personal histories"; CHI '96: Proc. of the SIGCHI conf. on Human factors in computing systems, ACM Press, New York (1996), 221-227.
- [Robertson, 00] Robertson G., van Dantzich M., Robbins D., Czerwinski M., Hinckley K., Ridsen K., Thiel D., Gorokhovskiy V.: "The task gallery: a 3d window manager"; Proc. CHI '00, ACM Press, New York (2000), 494-501
- [Smith, 04] Smith D. A., Raab A., Reed D. P., Kay A. C.: "Croquet: A menagerie of new user interfaces"; Proc. C5 '04, IEEE Computer Society, Washington (2004), 4-11

Spatiotemporal Knowledge Visualization and Discovery in Dynamic Social Networks

Ralf Klamma, Yiwei Cao, Marc Spaniol, Yan Leng
(Lehrstuhl Informatik 5, RWTH Aachen University, Germany
{klamma|cao|spaniol|leng}@i5.informatik.rwth-aachen.de)

Abstract: In this paper, we introduce a so-called DyVT tool (**D**ynamic social network **V**isualization **T**ool) to support spatiotemporal knowledge visualization and discovery in dynamic social networks. The dynamic aspects of social networks refer to contextualized information such as spatial, temporal as well as users' personalized information. We also define an XML-based target language incorporating emerging formats like DyNetML, KML, and GraphML. It also provides means to express, store and interchange the dynamic aspects of complex dynamic social network data. Based on this language, users can animate and personalize spatiotemporal knowledge extracted from social network data like email threads or blogs. In addition, a Java based graphical user interface is also available to enable non-experienced users to customize knowledge visualization easily. A mashup with Google maps for spatiotemporal visualization is provided. With this tool spatiotemporal knowledge on an IBM DB2 Mailing list database containing 69 mailing lists and 56389 mails altogether is well explored.

Keywords: Dynamic social network, information visualization, spatiotemporal knowledge, XML, GraphML, DyNetML, KML, interchange format, Google Map, animation

Categories: E.1, E.2, H.2.3, H.2.8, H.3.1, H.3.2, H.3.3, H.3.5, J.4

1 Introduction

Social network data generated from computer mediated communication over the Internet has created an enormous interest in the analysis of social interaction using digital networks. More and more research efforts have been put into Social Network Analysis (SNA) [DeFo03]. Social network indicates a special kind of network where social relationships are viewed as nodes and tiers with values in a social context. One of the key characteristics of a social relationship is its dynamic nature. However, most past analyses of social networks are essentially static where the dynamic nature of social relationships indicating when and where social relations take place is discarded. Hence, more recently there is an upsurge of interest in studying dynamic social network in which social data sets change over space and time [BrCa03]. Understanding networks from a dynamic perspective is essential to understand the interaction in virtual worlds. It facilitates reasoning about real groups as complex dynamic systems that evolve over space and time. Spatiotemporal information visualization has been a key research issue in the area of spatiotemporal knowledge representation and reasoning.

Meanwhile, visualization has played a major role in SNA. Instead of creating intriguing pictures, visualization is essential for generating learning situations

[ViDo04]. A standing critique of the research in social network visualization focuses on a “structural bias” that implicitly denies much of the dynamic nature of social relations [Milg67]. Much recent interest in this area revolves around understanding how networks develop and change. So far, the research community has developed a number of powerful tools for collecting, analyzing and visualizing social network data. However, these existing tools are developed for different purposes. Some concentrate on temporal visualization using movies or mashup images, while some aim to deal with spatial data using different map formats. Unfortunately, there is no such a system existing to visualize complex dynamic social network data in a user-friendly tool. From this point of view, a meaningful way to capture this dynamic contextualized information in a highly collaborative way is needed.

This paper presents spatiotemporal knowledge discovery via a so-called DyVT tool supporting dynamic social network visualization. Furthermore, an XML-based language is specified to define the data model integrating the dynamic nature of complex social network data such as temporal data, and geospatial data into a uniform file. All the later visualization tasks are performed based on this uniform XML file. It uses animations showing how relationships emerge over time to discover temporal knowledge. Moreover, a graph with a map as background is effective to represent spatial knowledge. In addition, user-chosen personalized graph appearance can also be visualized dynamically.

The rest of the paper is organized as follows. After reviewing the related work in Section 2, Section 3 is devoted to system architecture and system implementation. In Section 4 we make a summary and give the perspective of future work.

2 Related Work

In order to get meaningful knowledge visualization, the first step is to design an interchange format for complex social network data with regard to spatiotemporal knowledge. Secondly, visualization methods can be applied. Thus both interchange formats for social network data and existing packages are discussed in this section.

Interchange formats for social network data can be classified into two categories: unstructured free text based and XML based formats. The typical examples of the former format are DL [UCINET07] and Pajek .net [NMBa05]. These formats are easy to implement by software designers. But they are ambiguously understandable by both users and machines. The interoperability is poor, too. The latter XML based formats can overcome these limitations. They are easy to read and to parse by providing powerful facilities for building expressive and stable data formats with DTDs and XML Schemas. Three languages among them are presented as follows to show how they represent social network data.

GraphML (Graph Markup Language) [BrEi02] is an XML language to deal with network data. Its characteristic feature enables users to add modules that implement special extensions or additional data, which affords GraphML a large degree of flexibility. Additional advantage of GraphML is being supported by a large number of graph analysis and drawing software such as JUNG and yFiles etc.

DyNetML is provided by CASOS at Carnegie Mellon University [TRCa04]. It represents dynamic network data as sets of time slices. A dynamic network element is

defined as a sequence of *MetaMatrix* elements representing a snapshot of the organization for one time period.

KML is used to display geographic data in some special browsers such as Google Earth, Google Maps, and Google Maps for mobile. KML has a tag based structure with names and attributes used for specific display purposes. Thus, Google Earth and Maps act as browsers for KML files [KML07].

Next, the state-of-the-art overview of the available packages to visualize dynamic social network are given. Animations and “movies” are often used to visualize temporal data. DyNet [BrCa03] at CMU, MOVIE MOL [MOVIE MOL07] provided by Linköping University and Uppsala University, as well as SoNIA [SoNIA07] at Stanford University all use this method. In addition, Mapnet [Mapnet07] and London Connections [FaOw07] have been developed to visualize spatial data on maps. However, these packages serve different purposes. Some focus on temporal data visualization, while some concentrate on spatial data visualization.

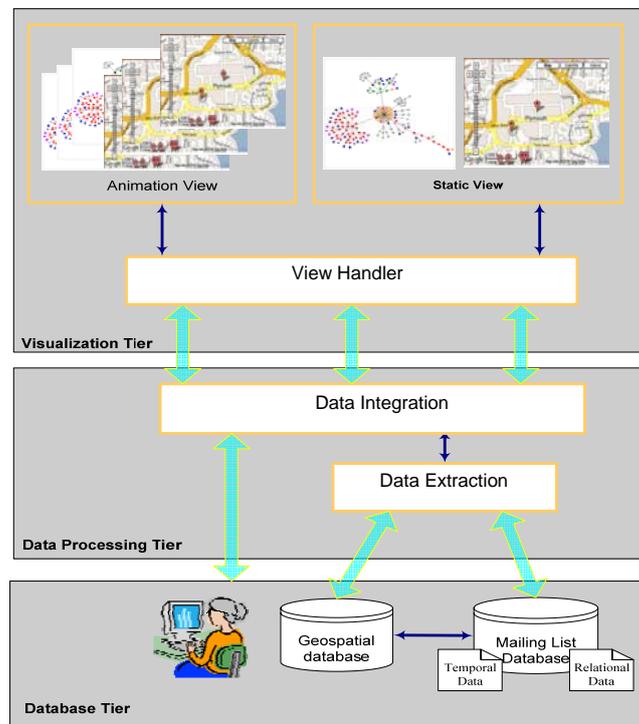


Figure 1: System architecture of DyVT

3 DyVT: Dynamic Social Network Visualization Tool

DyVT visualizes spatiotemporal knowledge within one common tool using the XML-based DyVTML. It employs the three-tier architecture based upon the information visualization reference model. The reference model divides the visualization process

into a series of steps ranging from data acquisition and modeling, the visual encoding of data, to the presentation of interactive displays [Chi00]. Figure 1 illustrates the system architecture of DyVT.

3.1 Database Tier

The database tier contains diverse social network data sources. In the prototype we use three data sources. The *IBM DB2 mailing list database* contains information of mailing lists, each of which can describe a social network. More detail about the definition of social relations can be found in [KSDe06]. If A sends an email to the mailing list, B and C reply it, then there exist directed relations between B and A as well as C and A. Besides the relational data for social networks, the temporal data can be obtained from it as well. *GeoLiteCity database* [MaxMind07] determines the Internet visitors' city information extracted from the IP address. Currently, it is a database with the size of 19 MB which generally contains IP address range and geographical data for all publicly assigned IPv4 addresses. *User-Chosen network appearance data* allows users to configure the graph appearance they prefer. After the attributes for nodes and edges are chosen, an XML-based ADML (*Appearance Data Markup Language*) file including all appearance data can be generated (cf. Figure 2).

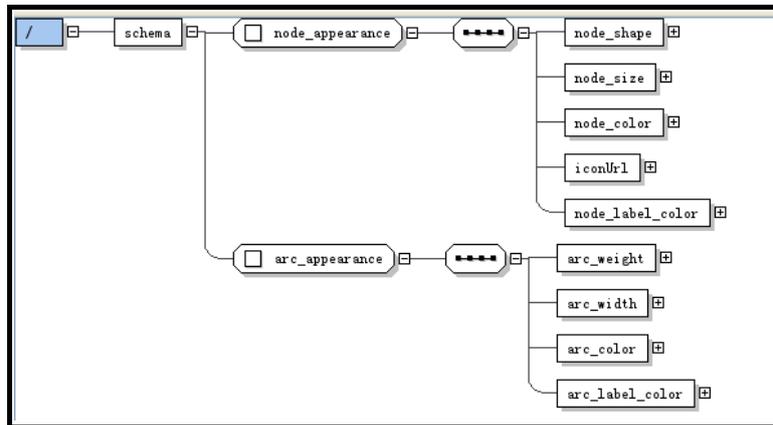


Figure 2: XML Schema for ADML

3.2 Data Processing Tier

The data processing tier bridges between the database tier and the visualization tier. It aims to collect all backing data needed in visualization tier. This process is developed through two steps: *Data Extraction* and *Data Integration*.

The step of data extraction is responsible for extracting relational social network data and dynamic nature of relations including spatiotemporal data from the mailing lists and the GeoLiteCity databases. Then the data is transformed into the XML-based DyVTML (*Dynamic Visualization Tool Markup Language*). It allows setting two constraints to extract the relational data with time stamps of social networks: *Choose a mailing list* and *Define time interval*. In addition, spatial data is extracted from the

email addresses. To obtain the relevant spatial data, the domain parts of the email addresses should be firstly extracted. Then the host names are mapped to the corresponding IP addresses. Finally, the IP addresses are converted into numeric representation which is used to look up and retrieve spatial data in GeoLiteCity database. The transformation process keeps the visualization tier independent from the varying formats of the data sources. Therefore, it reduces the execution time and is convenient for data migration.

In addition, GraphML, DyNetML and KML are suitable for being used as interchange formats. However, each of them serves different purposes. The starting idea of defining DyVTML is a logical integration of these three formats. The graph topology of GraphML is applied to represent relational data including the definition of nodes and edges. Since the spatial knowledge visualization is developed with the help of Google maps API, the supported KML is used. Temporal knowledge representation derives from DyNetML using sets of time slices. However, we do not use the expression of DyNetML directly, because in DyVT the time slices are able to be defined by users. Furthermore, we use the concept of time slices to list the time stamps for both nodes and edges.

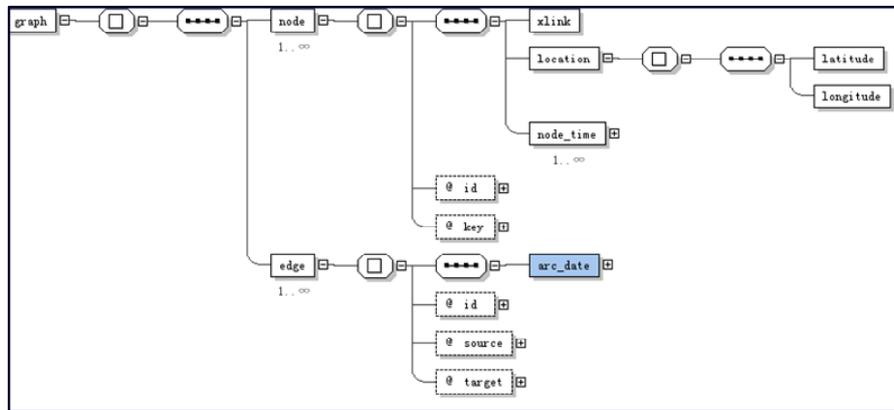


Figure 3: XML Schema for DyVTML

So the step of data integration contributes to parse two separate XML-based files. Appearance data ADML constructed in the database tier and the underlying social network data DyVTML. Both files instantiate series of nodes and edges which are data models including visual features for the visualization tier. The separation of appearance data from social network data allows different users to visualize the same underlying social network using different personalized appearance settings quickly.

3.3 Visualization Tier

The actual rendering of the data onto views is done in the visualization tier. To provide varying perspectives onto the dynamic nature of social network data, two handlers are built: *temporal view handler* and *map view handler*.

The temporal view handler is based on the SoNIA package. Images rendered within successive time slices are based on circle layout and the modified Kamada

Kawai layout [MMBe05] where the initial positions are fixed by nodes' position in the previous slice. These images are gathered for an animation to visualize social network within a certain time interval. DyVT enables users to specify the size of time slices which affects the number of nodes and edges shown in each frame of animation. Users can also choose the speed of moving time slices which affects how smoothly the network changes (cf. Figure 4).

In addition, the visualization results can be exported to multimedia formats such as JPEG, GIF, and SVG etc. The map view handler builds the mashup with Google maps [GoogleMap07]. Moreover, animated effects are applied to arrange the order of displaying nodes and edges based on their time stamps. Popup windows and name lists are developed for node information retrieval in order to improve the graph readability (cf. Figure 5).

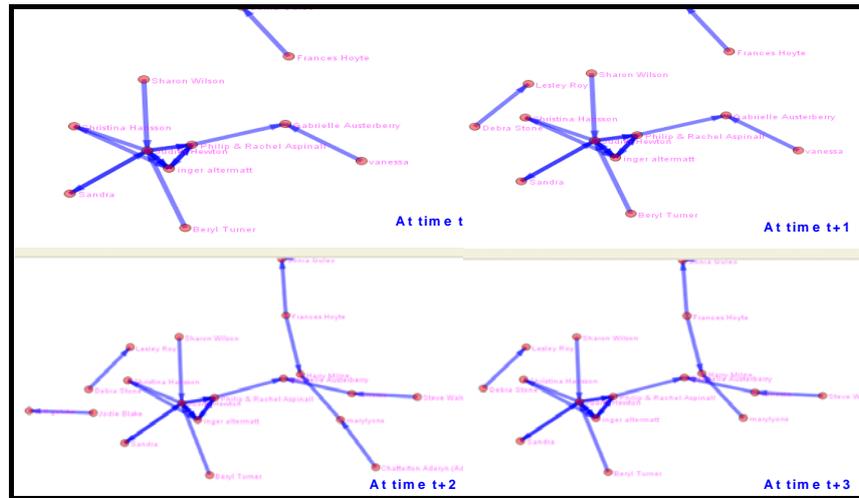


Figure 4: Screenshot of temporal knowledge representation in 4 successive time slices

4 Conclusions and Future Work

In summary, DyVT is a unique tool to capture and handle dynamic social networks. By defining two new XML based languages. DyVTML combines GraphML, DyVTML and KML for the integrated representation of spatiotemporal knowledge of dynamic networks. And ADML allows personalized views on networks visualization. DyVT allows straightforward animations of dynamic networks, mashups with Google maps and exportation of visualization results in available multimedia formats.

DyVT is still a prototype which provides a way to visualize multiple dynamic natures in a uniform platform. The future work will focus on pattern-based spatiotemporal knowledge exploration [AnAn06]. Furthermore, the interchange formats should support multi-relational social network in DyVTML by allowing different personalized appearances to each cluster within a network. Above all, DyVT is expected to be used in various Web 2.0 environments such as Wikis, Blogs, and

social networking software etc, in order to explore spatiotemporal knowledge of large dynamic social networks.

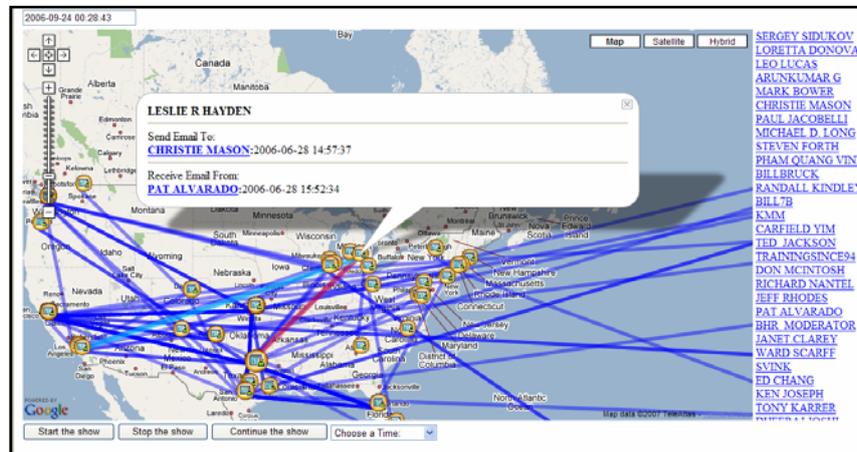


Figure 5: Screenshot of map view with a Popup Window and Name List

Acknowledgements

This work was supported by the excellence initiative of the German government within the research cluster “Ultra High-Speed Mobile Information and Communication (UMIC)”, by the German National Science Foundation (DFG) within the collaborative research center SFB/FK 427 “Media and Cultural Communication”, and by the 6th Framework IST programme of the EC through the NoE on Professional Learning (PROLEARN) IST-2003-507310.

References

- [AnAn06] N. Andrienko and G. Andrienko: Exploratory Analysis of Spatial and Temporal Data, Springer-Verlag Berlin Heidelberg, 2006
- [BrEi02] U. Brandes, M. Eiglsperger, I. Herman, M. Himsolt, and M. S. Marshall: GraphML progress report: Structural layer proposal. P. Mutzel, M. Junger, and S. Leipert (eds.) Proceedings of 9th International Symposium on Graph Drawing (GD '01), Springer Lecture Notes in Computer Science 2265, 2002, 501-512
- [BrCa03] R.L. Breiger and K.M. Carley: Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, National Academies Press, 2003, ISBN 0309089522
- [Chi00] E. H. Chi, A taxonomy of visualization techniques using the data state reference mode, Proceedings of IEEE Symposium on Visualization (InfoVis' 2000), October 9-10, 2000, Salt Lake City, USA, 69-76
- [DeFo03] A. Degenne and M. Forse: Introducing Social Networks, Sage Publications, 2003
- [GoogleMap07] Google Map API, <http://www.google.com/apis/maps/> {8.5.2007}

- [KML07] KML homepage: <http://earth.google.com/kml/index.html> {14.3.2007}
- [KSDe06] R. Klamma, M. Spaniol, and D. Denev: PALADIN: A Pattern Based Approach to Knowledge Discovery in Digital Social Networks, K. Tochtermann, H. Maurer (Eds.): Proceedings of I-KNOW '06, 6th International Conference on Knowledge Management, Graz, Austria, September 6-8, 2006, J.UCS (Journal of Universal Computer Science) Proceedings, Springer, 457-464
- [Mapnet07] Mapnet homepage: <http://www.caida.org/tools/visualization/mapnet/> {8.5.2007}
- [MaxMind07] MaxMind LCC, <http://www.maxmind.com/app/city> {14.3.2007}
- [Milg67] S. Milgram, The small world problem, *Psychology Today* 1, May 1967, 61-67
- [MMBe05] J. Moody, D. McFarland and S. Bender-deMoll: Dynamic network visualization. *American Journal of Sociology* vol. 110, no. 1, January 2005, 1206-1241
- [MOVIEMOL07] MOVIEMOL: "An easy-to-use molecular display an animation program", <http://www.fos.su.se/moviemol.html> {12.03.2007}
- [NMBa05] W. De Nooy, A. Mrvar, V. Batagelj, *Exploratory Social Network Analysis with Pajek*, Cambridge: Cambridge University Press, 2005
- [SoNIA07] SoNIA - Social Network Image Animator homepage: <http://www.stanford.edu/group/sonia/index.html> {14.3.2007}
- [FaOw02] R. Fawcett-Tang and W. Owen: *Mapping: An Illustrated Guide to Graphic Navigational Systems*, Rockport Publishers, 2002, ISBN 2880467071
- [TRCa04] M. Tsvetovat, J. Reminga, K. M. Carley: DyNetML: interchange format for rich social network data, Technical report, Institute for Software Research International School of Computer Science, Carnegie Mellon University, January 2004, CMU-ISRI-04-105
- [UCINET07] UCINET DL Format: <http://www.analytictech.com/downloadnd.htm> [Last access: 8.5.2007]
- [ViDo04] F. B. Viegas, J. Donath: *Social Network Visualization: Can We Go Beyond the Graph?* Workshop on Social Networks for Design and Analysis: Using Network Information in CSCW, 2004

GlobeMash: a Mashup for Accessing GLOBE

Joris Klerkx

(Katholieke Universiteit Leuven, Belgium
Joris.Klerkx@cs.kuleuven.be)

Erik Duval

(Katholieke Universiteit Leuven, Belgium
Erik.Duval@cs.kuleuven.be)

Abstract: In this paper, we present GlobeMash, a mashup web application that uses standardized data formats like XML, JSON, LOM, SVG, CAM and protocols like SOAP, HTTP, to enable users to access the repositories of the Global Learning Objects Brokered Exchange (GLOBE) consortium. GlobeMash uses the Google Maps API to display the repositories and results on a geographical map, the Timeline API of the SIMILE project and the federated search layer of the GLOBE infrastructure. It enables users to query all the repositories in GLOBE and to get an insight in their search history by visualizing the latter as a combination of an extended tag-cloud and a synchronized timeline.

Keywords: Hypermedia systems, web-based services, AJAX, information visualization.

Category: H3.3

1 Introduction

This paper presents GlobeMash, a mashup application that focuses on opening up a collection of learning object repositories. More specifically, we focus on the Global Learning Objects Brokered Exchange (GLOBE) consortium [GLOBE, 04], which aims to unlock the ‘deep’ web of the learning repository networks that its members (Ariadne, Edna, Merlot, Nime, LORnet and KERIS) maintain. GlobeMash enables users to

- find learning objects within the GLOBE network of repositories,
- obtain a visual insight in the search history of users.

These functionalities are necessary for users to be able to access learning objects of multiple repositories in an efficient and flexible way. Saving the history of a user is a functionality that resulted from usability research of user interfaces to access libraries [Dorner, et al., 03]. Furthermore, this history feature can be used to enable support systems for administration decisions. It can provide insight into the user’s behaviour thus enabling management to recognize needs and potential savings (e.g. cheaper learning material) earlier [Najjar, et al., 06].

A mashup is a web application that uses services from more than one source to create a new service. Services used in mashups are typically sourced from a third party via a public interface or API. GlobeMash uses

- the SQI API [van Assche, et al., 06] for searching within GLOBE,
- the Google Maps API [GoogleMaps, 07] to show geographic information about the GLOBE Repositories and their results, and

- the Timeline API of the SIMILE Project [SIMILE, 03] to visualize a history of keyword queries.

In section 2, we discuss the features of GlobeMash that distinguishes it from the existing interfaces. Section 3 goes deeper into the technical aspects of GlobeMash and we conclude this paper in section 4.

2 A Mashup for GLOBE

In this section we will present the mashup that was created to access GLOBE. First of all, we situate GlobeMash in the context of the existing web interfaces that each of the members of GLOBE has implemented. In the following sections, we will describe the features of GlobeMash one-by-one. We will not include keyword searching within GLOBE, as this is the most basic functionality that every interface should offer. We will only include those features that are specific for GlobeMash.

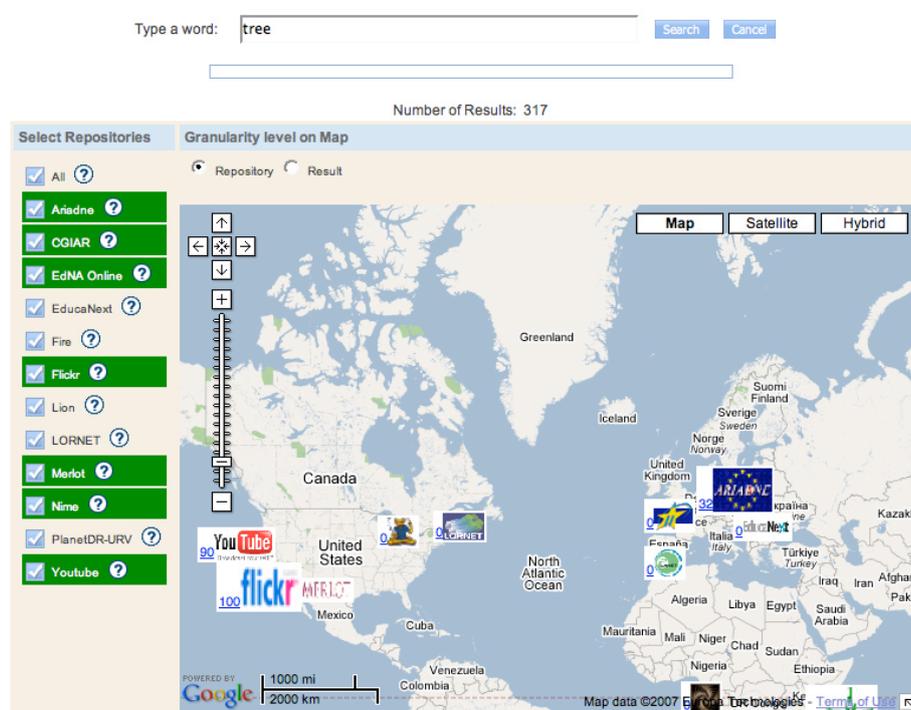


Figure 1: *GlobeMash: A Mashup Application to Search within GLOBE.*

2.1 Situation

Each member in GLOBE has implemented a web interface for querying the GLOBE repositories. They all have the following features in common:

- Users can search in all repositories by filling out a simple query in an electronic form.
- Search results are always presented as a human readable abstract of the metadata that describes the learning object.
- Users can consult a result from a particular repository through the web interface of the repository where the result was found.

Before researchers can investigate the behaviour of the users in GLOBE, they need to find a way to collect and merge the logs for all these separate applications. A possible solution would be to use the Contextualized Attention Metadata (CAM) scheme [Wolpers, et al., 06] for logging events so that the log events are standardized and easily distributable for researchers.

Title	Description	Repository
Phylogeny and Reconstructing Phylogenetic Trees	A java application that lets you investigate the problem of creating a phylogenetic tree, there is some explanation of the problem of drawing phylogenetic trees and a couple of versions of the applet. The applets let you randomly generate trees and also let you infer a phylogenetic tree with a couple of different methods. Number of species, mutation rate and the method for calculating the distance matrix can all be varied.	AMRELIC
Seed Threshing	A short video clip demonstrating the process of seed threshing. This is a part of Session 2 of the Tree Seeds for Farmers Toolkit.	COBAN
Seed Threshing	A short video clip demonstrating the process of seed threshing. This is a part of Session 2 of the Tree Seeds for Farmers Toolkit.	ARJANNE
Seed Sorting	A short video clip depicting the process of sorting seeds on a tarp. This is a part of Session 2 of the Tree Seeds for Farmers Toolkit.	COBAN
Seed Sorting	A short video clip depicting the process of sorting seeds on a tarp. This is a part of Session 2 of the Tree Seeds for Farmers Toolkit.	ARJANNE
Seed Progeny	An Image depicting species selection. Session 2 of the Tree Seeds for Farmers Toolkit	ARJANNE
From little acorns... An evaluation of The Knowledge Tree - an e-journal of flexible learning in VET	An evaluation of the Australian Flexible Learning Framework e-journal, The Knowledge Tree.	
Tree of Life	As described at the site, "The Tree of Life is a project containing information about the diversity of organisms on Earth, their history, and characteristics. The information is linked together in the form of the evolutionary tree that connects all organisms to each other." Biosciences section of the Directory of the Virtual Library, an expert-run catalog of sections of the	AMRELIC

Figure 2: Result list presentation of *GlobeMash*; results are sorted on the description field.

2.2 Result presentation

All existing interfaces in GLOBE present results of a keyword query as a list of objects, with a small number of metadata fields. These fields typically include the title, a description, the author(s) and the repository where the result was found. In figures 1 and 2, the reader can see that we integrated both such a list presentation and on top of that a googlemap with geographic information. In the following subsections, we discuss all methods of representing results within *GlobeMash*. Besides those presentation methods, a user can always look at the full metadata instance (LOM) that was returned by the *Globe* network, which is not always the case in existing interfaces.

2.2.1 Geographic information of the repository

When a user selects this level of granularity, all repositories that are searched on, are displayed on the map, as a logo at the geographic location of the headquarters of the organisation that maintains it. This can be seen in Figure 1. If the user issues a keyword query, a number is shown to the left of the icon that presents the number of search results that are returned. The size of the icon grows proportionally to the

number of results as a clear visual cue. This way of presenting the result information provides a clear visualization of the relation between the repositories and the results. We can for instance imagine that a user wants to see which repositories contain large numbers of learning resources on a specific topic like economy. The user can click on the repository logo of his choice to only see the results of that repository in the list (Figure 2) below the map so the map can also be used as a filtering mechanism.

2.2.2 Geographic information of the individual results

When a user selects this level, all results are indicated on the map. To be able to do so, we need a mapping from the metadata of the result to a location. We try to automatically deduce this location from the metadata of the result. First of all, if present, we map the address of the institution of the author(s) through an open geolocation database. If the metadata lacks the address of the institution, we look at the email address of the author(s) and map the domain in this address to an IP address. We can then again map this IP address to a location. Clearly, we cannot always do this. For instance, if the email address of the author(s) is not within the domain of his institution but rather from a domain like gmail.com or hotmail.com, the resulting location of the result will not be correct. Always available in the returned metadata is the location of the description of the result on the web in the form of a handle. This location is mapped to a geolocation in case that all previous ways fail.

With all results displayed on the map, one can get an impression about the distribution of learning objects on specific topics over the world.

2.2.3 List presentation of the results

A table of results is always shown when the user performed a query. The columns indicate the title, a description, the author(s) and the repository where the result was found. All the results can be sorted on every column as research proofs that this is a usability requirement for searching libraries and databases [Dorner, et al., 03]. This functionality is not available in the existing interfaces to find material in GLOBE. When the user selects a repository in the map, the table filters out the results of the other repositories so the list is synchronized with the map.

Another usability requirement is that a user should be able to merge results from different repositories [Dorner, et al., 03] in one list. Therefore we added the functionality to enable users to create their own list with results of their interests that came in from the network of repositories. The user can click on a result and look at the object itself without having to leave the current webpage. Results from multiple queries to the GLOBE network can be added to this list.

2.3 A history visualization

In order to create a feedback loop that enables learning from the way people actually use the search interface within GLOBE, it is essential to track the behaviour of users with the web application. It is needed to answer questions like

- Which keywords are used how many times?
- Which keywords are used that have many or no results?
- Which keywords have results in multiple repositories?

- Which keywords or topics have the attention of a particular user over time?
- What is the behaviour of users that search within the GLOBE community?
- What is the time spent with the web application?
- How many results does a user look at in detail after performing a query?
- When did a user perform a query and which keywords did she use?
- Etc.

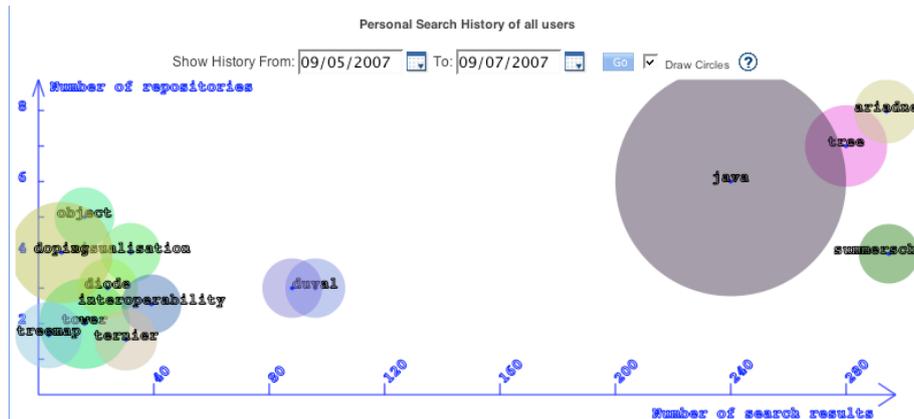


Figure 3: A Tag Cloud Visualization of Users' Search history

Therefore we created a visualization that enables users to get an insight in the questions above. Every event in the web application is logged in the Contextualized Attention Metadata (CAM) specification [Wolpers, et al., 06].

With the logged history, a visualization of the history of keyword queries from users is built and shown in GlobeMash. The visualization consists of a tag-cloud visualization, which we extended with extra information levels, and a timeline that displays all search events on it. This can be seen in figures 3 and 4. If a user does not login, the history of every user of the last time period is visualized without giving details on which query belongs to which user. If the user does login, he can see his own history. With the current software, we are able to visualize the events of multiple users. A user should be able to publish his extended tag-cloud to other users like people can do in the social bookmarking tool del.icio.us where they can publish their tag-cloud that is based on their bookmarks. This is however not yet functional GlobeMash but it will be added in the near future.

2.3.1 Extended Tag-Cloud

A tag-cloud is a list of the most popular tags, usually displayed in alphabetical order, and visually weighted by font size. In a tag-cloud, when a user clicks on a tag, one obtains an ordered list of tag-described resources, as well as a list of other related tags [Hassan-Montero, et al., 06]. In our version, the tags represent a keyword query of a user. If we would use an alphabetical order to place the keywords, users would not be able to get an insight in questions like

- Which keywords are used that have many or no results?

- Which keywords have results in multiple repositories?

Therefore, the keywords are not displayed in alphabetical order but are placed in a coordinate system where the X-axis represents the average number of search results and the Y-axis the number of repositories where results were found, both in relation to the performed keyword query. The scale of the X-axis and the Y-axis are automatically computed from the data that has to be drawn in the visualization.

On top of that, the visualization offers the functionality to draw circles around a tag or keyword. The bigger the radius of a circle, the higher the frequency of the keyword query through time. The color-coding of these circles can be random or it can be used to encode the number of different users that actually used the keyword query, depending on the selection of the user. Another idea for the color-coding would be the number of results that were downloaded and used by e.g. a teacher in his course.

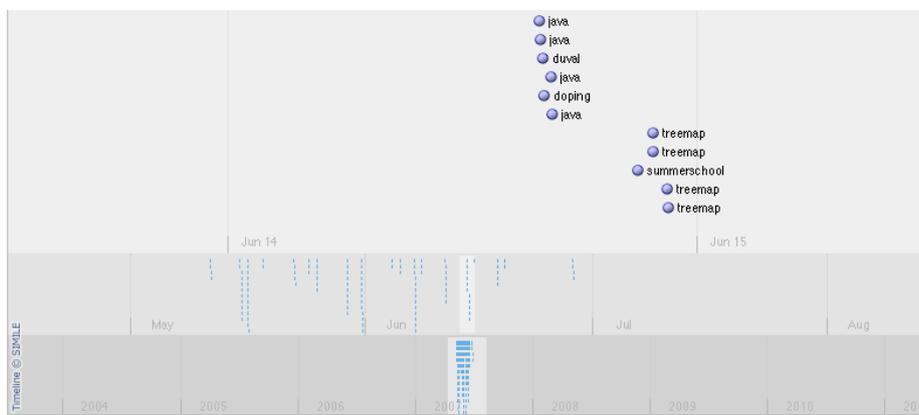


Figure 4: A Timeline visualization of users' search history

2.3.2 Timeline

The tag-cloud visualization gives an overview of the keyword used in queries in the federated network of repositories. It however does not show when a keyword is used. Therefore we integrated a timeline that we synchronized with the tag-cloud visualization. On the timeline, all keyword queries are displayed at the exact time and date that they were issued. In this way, a user can get an idea of the topics he was interested in over time. If a user scrolls the timeline to a specific date, he can choose to only show the keywords from a start date to the chosen date on the visualization. This can be used to filter out keywords that were used in a time period that does not have the interest of the user. If a user clicks on a keyword in the tag-cloud visualization, the timeline will smoothly scroll to the last timestamp that the keyword was used.

3 Underlying Technology

Our objective was to design and implement a mashup web application which is modular and uses standardized data formats (LOM [LOM, 04], XML [XML, 06], JSON [JSON, 06], SVG [SVG, 03], CAM [Wolpers, et al., 06]) and protocols (HTTP, SOAP) to ensure that different parts of GlobeMash can also be used in other contexts than GLOBE. The visualization of the history can for instance be used in other tools that collect attention metadata. GlobeMash is created with the help of the Dojo toolkit which is an Open Source DHTML toolkit written in JavaScript. It allows developers to easily build dynamic capabilities into web pages [DOJO, 06].

The federated search layer of GLOBE is accessed when a user issues a keyword query. This works by sending an AJAX request from the client to a java servlet at our server that uses the CEN/ISSS Simple Query Interface (SQI) [van Assche, et al., 06] for querying the repositories within GLOBE. The results that are returned to the servlet are in the IEEE LTSC Learning Object Metadata (LOM) format [LOM, 04]. These are returned in JSON format [JSON, 06] to the client where we parse the JSON data and add the results to the map and the list presentation. The map is rendered using the Google Maps API [GoogleMaps, 07]. The geographical locations of the results are found by performing three look-up requests in GeoIP [GeoIP, 07], NetGeo [Moore, et al., 00] and Hostip [Hostip, 07]. By performing multiple requests, it is more likely to get a valid location as the three geolocation databases together have a better coverage of the world.

The history visualization consists of two parts. The upper part is an SVG (Scalable Vector Graphics) drawing that is rendered by the Dojo framework in the client browser. The timeline visualization part is created using the SIMILE project and is a DHTML-based AJAX widget for visualizing time-based events [SIMILE, 03]. We synchronized the SVG visualization and the timeline widget by using the javascript event system.

All events in the web application are logged in the Contextualized Attention Metadata (CAM) specification and stored in a PostgreSQL database. The events are transferred to the client browser by mapping the attention metadata to JSON after which the history visualization gets created by executing a number of our javascript functions in the client browser.

4 Conclusion

This paper is part of our continuing research on accessing learning object repositories in an efficient and flexible way. In [Klerkx, et al., 05], we created an open and extensible information visualization framework, which we used to access the Ariadne Knowledge Pool System. In this paper we broadened our scope by moving from one LOR towards opening up a collection of repositories, namely GLOBE. Therefore, we created GlobeMash: a mashup application that distinguishes itself from existing interfaces by displaying the repositories on a geographical map, by adding a visualization of the history of the keyword queries of the users and by adding functionalities that improve the usability of search by e.g. enabling users to merge results from multiple repositories.

Logfile analysis has been the starting point in order to begin to know what happens inside GlobeMash. We know the user attention from looking at the attention metadata that is collected and from using the history visualization. But this approach is insufficient as an evaluation of the usability of the mashup from an end user point-of-view. We want to know what users like and what they dislike, what they find useful, what they find hard to understand, etc. Therefore, we are planning to perform a thorough usability test. We will evaluate GlobeMash against a number of usability criteria that were defined in [Dorner, et al., 03]. The results will be used to add new features or leave out the features that are not useful in the web application.

Acknowledgements

We gratefully acknowledge the financial support of the K.U.Leuven research council through the BALO project, the Interdisciplinary Institute for Broadband Technology (IBBT) through the Acknowledge project, and the European Commission through the ProLearn Network of Excellence on Professional Learning.

References

- [van Assche, et al., 06] F. van Assche, E. Duval, D. Massart, D. Olmedilla, B. Simon, S. Sobernig, S. Ternier, and F. Wild, "Spinning Interoperable Applications for Teaching & Learning using the Simple Query Interface", *Educational Technology & Society* 9 (2) (2006), 51-67.
- [DOJO, 06] The Dojo Toolkit: <http://dojotoolkit.org>
- [Dorner, et al., 03] Dorner, D.G., Curtis, A. (2003), "A comparative review of common user interface software products for libraries", National Library of New Zealand
- [GeoIP, 07] GeoIp Technology, <http://www.maxmind.com/app/ip-locate>
- [GLOBE, 04] Global Learning Objects Brokered Exchange: www.globe-info.net/
- [GoogleMaps, 07] Google Maps API: www.google.com/apis/maps/
- [Hassan-Montero, et al., 06] Y. Hassan-Montero and V. Herrero-Solana. "Improving Tag-Clouds as Visual Information Retrieval Interfaces, Proc. InSciT 2006.
- [HostIp, 07] hostIP geolocation database: www.hostip.info
- [JSON, 06] RFC4627: JavaScript Object Notation (JSON)
- [Klerkx, et al., 05] J. Klerkx, M. Meire, S. Ternier, K. Verbert, and E. Duval, "Information Visualization: Towards an Extensible Framework for Accessing Learning Object Repositories"; Proc. ED-Media'05, AACE, Montreal, Canada (2005), 4281-4287
- [LOM, 04] 1484.12.1: IEEE Standard for Learning Object Metadata
- [Moore, et al., 00] D. Moore, R. Periakaruppan, J. Donohoe, K. Claffy, "Where in the World is netgeo.caida.org?" CAIDA Publications, (2000)
- [Najjar, et al., 06] J. Najjar, M. Wolpers and E. Duval, "Attention Metadata: Collection and Management", WWW2006 workshop on Logging Traces of Web Activity: The Mechanics of Data Collection, Edinburgh, Scotland, May 23, 2006
- [SIMILE, 03] The SIMILE project: <http://simile.mit.edu/wiki/SIMILE>About>
- [SVG, 03] W3C Scalable Vector Graphics (SVG) format
- [Wolpers, et al., 06] M. Wolpers, G. Martin, J. Najjar and E. Duval, Attention Metadata in Knowledge and Learning Management, Proc. IKNOW'06, Graz, Austria
- [XML, 06] W3C Extensible Markup Language (XML)

Sky-Metaphor Visualisation for Self-Organising Maps

Khalid Latif and Rudolf Mayer

Institute of Software Technology & Interactive Systems
Vienna University of Technology
Favoritenstrasse 9-11/188, A-1040 Vienna, Austria
{klatif, mayer}@ifs.tuwien.ac.at

Abstract: Self-Organising Maps are utilised in many data mining and knowledge management applications. Although various visualisations have been proposed for SOM, these techniques lack in distinguishing between the items mapped to the same unit. Here we present a novel technique for the visualisation of Self-Organising Maps that displays inputs not in the centre of the map units, but shifts them towards the closest neighbours, the degree of the movement depending on the similarity to the neighbours. The night-sky visualisation facilitates better understanding of the underlying data. We report results from applying our method on two synthetic and a real-life data set.

Key Words: Self-organising Map, Visualisation, Night-sky

Category: H.3.3, H.5.2, I.2.6

1 Introduction

The Self-Organising Map (SOM) is a prominent tool for data mining and knowledge management. Part of its popularity can be attributed to the various visualisation methods which summarise the characteristics of the data set and help the user in understanding and analysing the underlying structure in the input data. The location of the input objects on the map allows the user to quickly identify similar and different objects.

However, the mapping of an input onto single map units is coarse and inaccurate to some extent. Depending on the resolution of the map, i.e. the number of units, inputs mapped onto the same unit might bear significant differences, which are not easy to transmit or visualise. Therefore, we propose a novel visualisation technique that takes into account not only the best-matching unit of an input object, but also the input's distances to the neighbouring units. As a result, the objects will not be placed at the centre of the map unit, but drift towards some of the neighbouring units. This helps the user on one hand to more easily distinguish between the items in the same unit, and on the other hand to grasp the similarities between data objects across unit boundaries.

This paper is organised as follows: Section 2 gives a brief introduction to the SOM algorithm, a survey of the most relevant SOM visualisations, and other systems using a night-sky metaphor for visualisation. Section 3 introduces our new visualisation technique, and Section 4 demonstrates its applicability. In Section 5 we conclude our findings and provide an outlook on future work.

2 Related Work

In this section we give an overview of the Self-Organising Map (SOM), its various visualisations, and other systems using the sky metaphor for data visualisation.

2.1 Self-Organising Map

The SOM [3] is a unsupervised neural network model that provides a mapping from a high-dimensional input space to a lower, often two-dimensional, output space. An important property of this mapping is that it is topology preserving – elements which are located close to each other in the input space will also be closely located in the output space. The generated map can help the user in getting a quick overview of the patterns in the input space.

The input space consists of any kind of data collection that can be represented in the numerical form - e.g. a vector space bag-of-words representation of text documents, features extracted from audio or images, or any other kind of numerical data. The output space is in many applications organised as a rectangular grid of units, a representation that is easily understandable for users due to its analogy to 2-D maps. Each of the units on the map is assigned a *weight vector* \mathbf{m}_i , which is of the same dimensionality as the vectors \mathbf{x}_i in the input space. During the training process, the vectors \mathbf{x}_i are presented to the Self-Organising Map, and the unit with the most similar weight vector to this input vector, the *best-matching unit*, is determined. The weight vector of this unit, and, to a lesser extent, of the neighbouring units, are adapted towards the input vector, i.e. their distance in the input space is reduced – the output space ‘folds’ as closely as possible into the input space. After the training is finished, the inputs are mapped onto their ultimate best-matching unit. Some units might accumulate a lot of inputs, while others, probably located between clusters, may be left empty. Further details on the SOM training process can be found in [3].

The SOM provides clustering of the data without explicitly assigning data items to the clusters or identifying cluster boundaries as opposed to, for example, the *k*-Means method. To allow an easier interpretation of the cluster structures and correlations in the content, visualisation techniques highlighting cluster boundaries and cumulations in the map are needed.

2.2 Self-Organising Map Visualisations

SOM visualisations can utilise the output space as a platform [13], where quantitative information is most commonly depicted as colour values or markers of different sizes. More advanced approaches use e.g. the analogy to geography [9].

Weight-vector based techniques, rely solely on the weight vectors. Among them, Component Planes are projections of single dimensions of the weight vectors \mathbf{m}_i . By plotting all dimensions, all information about the weight-vectors

is revealed. With increasing dimensionality, however, it becomes more difficult to perceive important information such as clustering structure and underlying dependencies. The unified distance matrix (U-Matrix [12]) is a visualisation technique that shows the local cluster boundaries by depicting pair-wise distances of neighbouring weight vectors. It is the most common method associated with Self-Organising Maps and has been extended in numerous ways. The Gradient Field [6] has some similarities with the U-Matrix, but applies smoothing over a broader neighbourhood and uses a different style of representation. It plots a vector field on top of the lattice where each arrow points to its closest cluster centre. This can be used to contrast different groups of Component Planes [7].

A second category of visualisation techniques take into account the **data distribution**. The most simple ones are hit histograms, which show how many data samples are mapped to a unit, and labelling techniques, which plot the names and categories, provided they are available, of data samples onto the map lattice. More sophisticated methods include Smoothed Data Histograms [5], which show the clustering structure by mapping each data sample to a number of map units, or graph-based methods [8], showing connections for units that are close to each other in the feature space. The P-Matrix [11] depicts the number of samples that lie within a sphere of a certain radius, namely a quartile of the pair-wise distances of the data vectors, around the weight vectors. Our newly proposed method falls into this category, and has certain similarities to hit histograms.

Emergent SOMs [10] work with a very high number of map units, i.e. provide a very high resolution. While a higher resolution can achieve to some extent similar results as our visualisation, namely distinguishing on a more detailed level between similar input objects, it does not completely solve the issue – input objects might still be mapped onto the same unit, and the map does not give hints on the similarity to inputs mapped on neighbouring units. Moreover, the ideal size of the SOM would need to be known, determined through experimental trials. Increasing the map size also has implications on performance.

2.3 Sky-metaphor Data Visualisation Techniques

The IN-SPIRE [14] tool builds on galaxy visualisation by making use of the metaphor of stars in the night sky. Each star represents an individual document, and clusters around centre points represent themes. The galaxy metaphor is also investigated in a prior work [2] to visualise document similarity. The night sky metaphor is also used in InfoSky [1] with a different approach. InfoSky is contingent on the assumption that documents are already organised in a hierarchy of collections. The collections are rendered as Voronoi cells, and hierarchically related collections are placed alongside each other. In contrast, the SOM is used for organizing the collections, and sky metaphor as a novel visualization of the SOM. documents

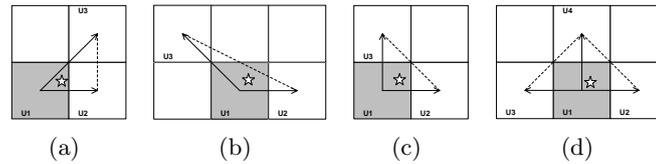


Figure 1: Neighbouring forces on an element

3 Sky-metaphor Visualisation for Self-Organising Maps

In this section, we present our Sky-metaphor visualisation technique. The map uses a black background to resemble the night sky. Individual objects from the input space are represented as stars, which together with other similar objects may form star clusters. This effect can be enhanced by using a Smoothed Data Histograms [5] visualisation on top of the background, resembling galaxies. Units that do not contain any inputs remain black and will resemble dark nebulae.

Different interaction strategies such as zooming & panning, individual document and area selection are not specific to the sky metaphor, but are supported by our SOM toolkit [4].

Star Clusters

Traditionally, the SOM algorithm assigns input objects only to a discrete map unit. We however want to reveal more details about the relations between the objects that are mapped onto the same unit, and also the similarities of the objects to other objects in neighbouring units. Therefore, we propose to place the input objects not in the centre of a unit cell, but spread them across the cell.

Neighbourhood Forces

We calculate the exact location of an input \mathbf{x} which is mapped onto its best-matching unit U . Our assumption is that the location of the input \mathbf{x} in unit U is driven by the position of the next closest units, with the distance of \mathbf{x} to these units acting as a pull force to the input. The pull force (\mathbf{F}) of a unit is inversely proportional to the distance of the input from the unit and is relative to the distance of the input to its best matching unit:

$$\mathbf{F}_i \propto \frac{d(x, U_1)}{d(x, U_i)} \quad \text{for } i > 1 \quad (1)$$

where d denotes the metric measuring the distance from the input to the weight vector of a unit.

As the second-best matching unit is nearer to the input than third-best-matching unit, its pull force is higher in magnitude. For this reason, the displacement effect is insignificant for farther units. In most of the cases the second

and third closest units, denoted as U_2 and U_3 , are sufficient for calculating the displacement of the input \mathbf{x}_i from the centre of the unit U . Their pull forces make up a virtual triangle, as illustrated in Figure 1. There is one rare exception to this assumption, namely in cases where both the second and third best matching unit are found to be on one axis with U . This implies that the input \mathbf{x}_i would drift along only one dimension as a triangle effect can not be realized. In those cases, the fourth closest unit U_4 is taken into account (c.f. Figure 1(d)).

The x and y coordinates of the exact position \mathbf{p} of input \mathbf{x} on unit U can then be defined as:

$$\mathbf{p}_{\langle x,y \rangle} = \left\langle \lambda * \sum_{i=2}^k \mathbf{F}_i * \frac{1}{U_{i\langle x \rangle} - U_{1\langle x \rangle}}, \lambda * \sum_{i=2}^k \mathbf{F}_i * \frac{1}{U_{i\langle y \rangle} - U_{1\langle y \rangle}} \right\rangle$$

where k is an index over the two or three nearer units U_2 , U_3 and U_4 respectively, i.e. $k = 3$ or $k = 4$. A grid-constant λ is used to reconcile the displacement according to the display co-ordinates and is initially set to approximately a quarter of the unit's pixel size. In some cases two or more inputs may overlap each other too much due to very high similarity. In such a situation we marginally shift the inputs apart by applying a force of repulsion, where overlapping units push each other in opposite direction.

Constellations as Interconnection Trails

In the physical world, entities are usually interconnected either by physical or by semantic means. In the proposed night-sky visualization, the interconnections are realized by exploiting the notion of constellations. Closely related stars form a pattern and highlights the relationship between the inputs, which may otherwise be mapped to different units. We allow both user defined and automatic trails (such as based on meta-data) to illustrate usefulness of constellations by drawing connection lines between the stars.

4 Experiments and Results

For the experiments described in this section, we used two synthetic data sets to demonstrate the visualisation, and one text data set to test its applicability to a large real-life corpus.

Figure 2 shows experiments with the 'chain-link' data set, i.e. two intertwining rings in three-dimensional space. This data-set cannot be projected to two dimensions while preserving the ring-structures, the normal behaviour is for the rings to 'break'. The visualisation resembles the structure of the two rings well, with the points stretching over the cell space in such a way that an almost continuous line is formed. This is very similar to the original data, which also

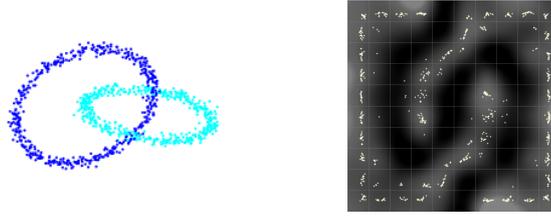


Figure 2: Chain-link data set and a trained map

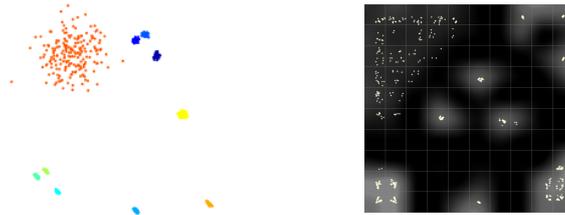


Figure 3: A data set of several different Gaussian clusters and a trained map

doesn't form the ring as a continuous data chain, but rather as several small clusters of data points.

Figure 3 depicts a plot of the two principal components of a ten-dimensional data set, generated using several Gaussian distributions with different centres and kernel widths. By not placing the data items in the centre of the units, the Sky-metaphor visualisation shows the concentration of inputs more effectively and also provides clear cluster boundaries.

The text corpus we used for our last experiment is the *20 newsgroups* data set. It consists of 1000 newsgroup postings for each of its 20 different newsgroups, such as *alt.atheism* and *comp.sys.mac.hardware*. We considered only the subject and the message body, but omitted other header lines. A standard *bag-of-words* indexing approach was used, applying a manually created stop-word list and document frequency threshold to reduce the dimensionality. A $tf \times idf$ weighting scheme was employed to obtain the vector values for the 2896 remaining terms. Finally, we trained a SOM of the size of 50×40 units.

Figure 4 depicts the overview of the trained map. Due to spreading the inputs over the SOM cells and the tendency to the inputs being moved towards the cluster centres, these become more compact and dense, while the areas between two clusters become larger – it becomes easier to identify groups of similar inputs.

Figure 5 depicts two sections of the map. The left image in the figure illustrates the concept of constellations: postings that are in relation to each other,

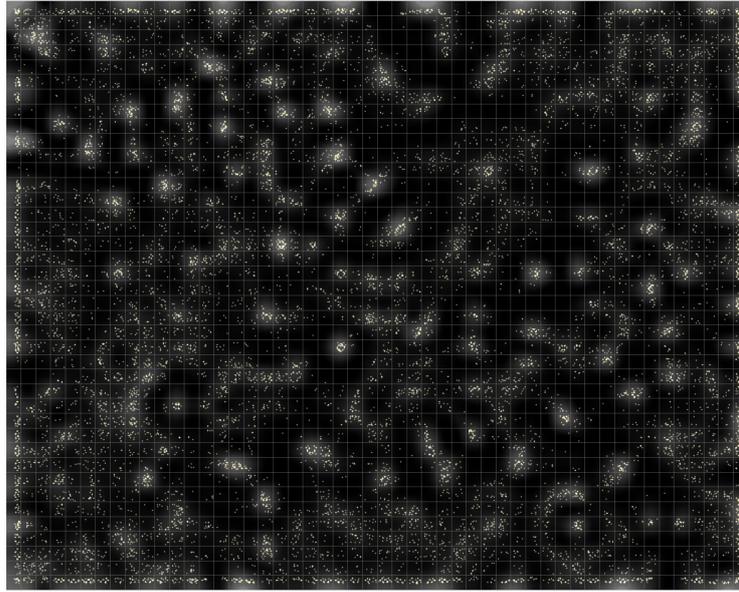


Figure 4: Sky Visualisation of 20 Newsgroups maps – overview

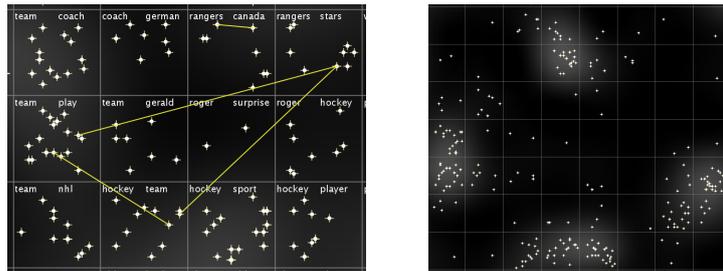


Figure 5: Detailed view of the 20 Newsgroups map

here direct replies to other postings, are linked. Such associative referencing allows instantly recalling other linked items in the data set. The right image shows a detailed view of cluster boundaries between two *sci.med* clusters in the upper-left and upper-middle area, and two *rec.motorcycles* and *rec.autos* clusters located in the lower-middle and right-middle area. Even though there are only few or no empty units between the cluster centres, the inputs on the units between those centres have been placed closer towards the centres, and therefore the cluster boundaries become easily visible.

5 Conclusions

In this paper we presented a novel method for visualising Self-Organising Maps. The night-sky metaphor is used to represent and interactively explore the underlying data set. The relationship of similarity between the inputs was depicted through star clusters and other complex interconnections by constellations. Our experiments with different data sets show that even a large stockpile of data could be turned into very useful knowledge map with effective visualisations.

References

1. Keith Andrews, Wolfgang Kienreich, Vedran Sabol, Jutta Becker, Georg Droschl, Frank Kappe, Michael Granitzer, Peter Auer, and Klaus Tochtermann. The InfoSky visual explorer: Exploiting hierarchical structure and document similarities. *Information Visualization*, 1(3/4):166–181, December 2002.
2. Beth Hetzler, Michelle Harris, Susan Havre, and Paul Whitney. Visualizing the full spectrum of document relationships. In *Proc. of 5th Intl. Conf. on Structures and Relations in Knowledge Organization*, pages 168–175, Lille, France, August 1998.
3. Teuvo Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, 1995.
4. Robert Neumayer, Michael Dittenbach, and Andreas Rauber. PlaySOM and PocketSOMPlayer: Alternative interfaces to large music collections. In *Proc. of the Sixth Intl. Conf. on Music Information Retrieval*, pages 618–623, London, UK, September 11-15 2005.
5. Elias Pampalk, Andreas Rauber, and Dieter Merkl. Using Smoothed Data Histograms for Cluster Visualization in Self-Organizing Maps. In *Proc. of the Intl. Conf. on Artificial Neural Networks (ICANN'02)*, pages 871–876, Madrid, Spain, August 27-30 2002. Springer.
6. Georg Pözlbauer, Michael Dittenbach, and Andreas Rauber. A visualization technique for Self-Organizing Maps with vector fields to obtain the cluster structure at desired levels of detail. In *Proc. of the Intl. Joint Conf. on Neural Networks*, pages 1558–1563, Montreal, Canada, July 31 - August 5 2005.
7. Georg Pözlbauer, Michael Dittenbach, and Andreas Rauber. Advanced visualization of Self-Organizing Maps with vector fields. *Neural Networks*, 19(6-7):911–922, July-August 2006.
8. Georg Pözlbauer, Andreas Rauber, and Michael Dittenbach. Advanced visualization techniques for Self-Organizing Maps with Graph-Based Methods, May 30 - June 1 2005.
9. André Skupin. A picture from a thousand words. *Computing in Science and Engineering*, 6(5):84–88, Sept.-Oct. 2004.
10. Alfred Ultsch. *Data mining and knowledge discovery with emergent self-organizing feature maps for multivariate time series*, pages 33–45. Elsevier, 1999.
11. Alfred Ultsch. Maps for the visualization of high-dimensional data spaces, 2003.
12. Alfred Ultsch and Hans Peter Siemon. Kohonen's self-organizing feature maps for exploratory data analysis, July 1990.
13. Juha Vesanto. SOM-based data visualization methods. *Intelligent Data Analysis*, 3(2):111–126, 1999.
14. Pak Chung Wong, Beth Hetzler, Christian Posse, Mark Whiting, Susan Havre, Nick Cramer, Anuj Shah, Mudita Singhal, Alan Turner, and Jim Thomas. IN-SPIRE InfoVis 2004 Contest Entry. In *Proc. of the IEEE Symposium on Information Visualization*, pages 216–217, Washington, DC, 2004.

On the Generalized Box-Drawing of Trees —Survey and New Technology—

Benno Stein

(Faculty of Media / Media Systems
Bauhaus University Weimar, Germany
benno.stein@medien.uni-weimar.de)

Frank Benteler

(Computer Science Department
University of Paderborn
33095 Paderborn, Germany)

Abstract: This paper is on the aesthetic layout of n -ary trees with nodes of *variable size*, also referred to as generalized box-drawing. For this layout problem a few algorithms have been proposed, which differ in their runtime performance and the attained aesthetic criteria.

We introduce a new approach to this layout problem, which is interesting because of both its simplicity and elegance, and which employs the following piggy-back metaphor: The original box-drawing problem, Π , is topologically reformulated as a layout problem with *uniform node size*, resulting in a new drawing problem Π' , which then is handled by the best tree layout algorithm for uniform node size developed so far. The reformulation step can be done in linear time in the number of nodes of the original tree, resulting in overall linear time complexity. Compared to the existing approaches our approach fulfills more aesthetic criteria; experiments have also shown its efficiency in complex layout settings.

Key Words: tree layout, generalized box-drawing, topology reformulation, meta drawing

Category: H.5, I.3.5

1 Introduction

The drawing of rooted, ordered trees of unbounded degree is considered a solved problem in the field of graph drawing—as long as the trees are restricted to having uniform node size. A generalization of this layout problem is the so-called box-drawing, where each node represents a rectangular box of predefined size. In the relevant literature, in particular in [Cruz and Tamassia 1988; DiBattista et al. 1994, 1999], less attention is paid to this layout problem though various applications exist that require a solution of a certain instance of this drawing task:

- Presentation of linked information items and networks [Eklund et al. 2002].
- Browsing and visualizing hierarchical relations of the World Wide Web.
- Visualization of hierarchies in object-oriented systems [Girba and Lanza 2004].
- Drawing of structured flowcharts for diagram generation [Ogura et al. 1992].
- Drawing of proof trees for the analysis of logical formulas [Bajaj et al. 2003].
- VLSI layout, in particular the placement for modules [Lengauer 1990].

Before we delve into details of the layout problem, Definition 1 will recapitulate the terminology used in the paper.

Definition 1 (Terminology). A rooted tree, $T = \langle V, E \rangle$, is a connected, acyclic graph with node set V and edge set E , having a designated node $r \in V$, called root. Each node v has a unique path $p_r(v)$ to the root; the path length, $|p_r(v)|$, is called the level of v . Given an edge $\{v, w\} \in E$, v is called parent

node of w and w is called child node of v if $|p_r(v)| < |p_r(w)|$. $\beta : V \rightarrow \mathbf{N} \times \mathbf{N}$ is a mapping that assigns each node v a bounding box of predefined width and height.

A drawing is a mapping $\delta : V \rightarrow \mathbf{N} \times \mathbf{N}$ that assigns each node v a position; a boolean layout constraint is a two-valued mapping, $\gamma : T \times \beta \times \delta \mapsto x \in \{0, 1\}$, indicating whether or not the constraint is fulfilled [see 1]. A tree T along with an assignment of boxes, β , and a set of layout constraints, Γ , forms a generalized box-drawing problem II . A drawing δ is called a valid layout for II if all layout constraints $\gamma \in \Gamma$ are fulfilled.

Given a generalized box-drawing problem II , let bb_{\max} be a bounding box whose width (height) is the supremum of the widths (heights) of all bounding boxes β . A common ad-hoc solution to II is the application of a standard tree layout algorithm that presumes uniform node size—either with or without scaling the uniform node size in order to provide sufficient room for bb_{\max} . Another technique is slicing, which organizes the bounding boxes in vertical or horizontal lanes. Such approaches are easy to implement but aesthetically unsatisfactorily, as Figure 1 shows.

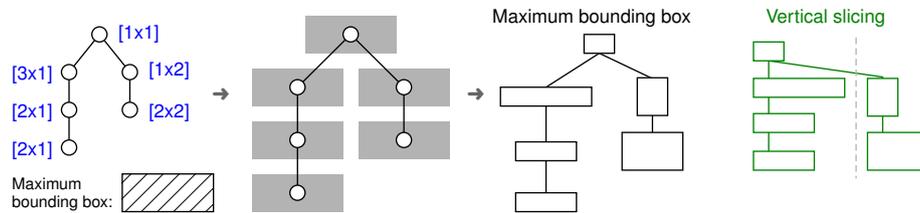


Figure 1: Ad-hoc solutions to II . The use of standard tree layout algorithms for the generalized box-drawing problem yields unacceptable drawings in most cases.

The complexity of an algorithm for drawing a tree depends on the aesthetic criteria to be fulfilled. In fact, the problem of computing a drawing of a binary tree on an integer grid of minimum area becomes NP-hard under reasonable layout constraints [Supowit and Reingold 1983]. However, given a different set of aesthetic criteria, an $O(n\sqrt{n})$ runtime algorithm for finding a minimum area drawing can be stated [Eades 1992]. The famous algorithm of Reingold and Tilford runs in linear time, and, though the resulting drawing does not give minimum compactness, the entirety of the fulfilled layout constraints results in an aesthetic drawing. These considerations show that there is considerable room to develop new variants of algorithms for the generalized box-drawing of trees, each fulfilling a different set of aesthetic constraints or being bound by a different runtime complexity.

Both the development of a special purpose algorithm and the ad-hoc application of a standard layout algorithm characterize extreme points. In this paper we propagate a third possibility: The reformulation of the original box-drawing problem II as a layout problem of uniform node size, which then is handled by a standard tree layout algorithm for uniform node size. This possibility exploits the encoded algorithmic know-how as well as the accepted practice of aesthetic criteria.

[1] Section 2 organizes the well-established boolean layout constraints.

	Approach	Runtime	Space	Layout constraints
Uniform-sized nodes	Wetherell and Shannon 1979	$O(n)$		1, 2, 5, 7
	Reingold and Tilford 1981	$O(n)$		1, 2, 3, 4, 5, 6, 7, 8
	Walker 1990	$O(n^2)$		1, 3, 4, 5, 6, 7, 8, 9, 10
	Buchheim et al. 2002	$O(n)$		1, 3, 4, 5, 6, 7, 8, 9, 10
Variable-sized boxes	Vaucher 1980	$O(nh)$	$O(n)$	1, 5, 6, 7, 8
	Bloesch 1993	$O(nh)$	$\Theta(h)$	3, 5, 6, 7, 8, 9, 11, 13, 14
		$O(nh)$	$O(nh)$	3, 4, 5, 6, 7, 8, 9, 11, 13, 14
	Miyadera et al. 1998	$O(n^2)$		5, 6, 7, 11, 13, 15
	Hasan et al. 2002	$O(n)$		5, 6, 7, 11, 13, 15

Table 1: Well-known algorithms for tree layout along with their runtime and the fulfilled layout constraints. n designates the number of nodes; h designates the drawing height of the tree, which is measured in unit bounding boxes for example (cf. Section 3).

The paper is organized as follows: Section 2 provides a unified view onto the existing work for tree layout; Section 3 introduces our reformulation approach, presents new analysis results, and discusses implications.

2 Classification of Existing Work

The development of algorithms for the drawing of trees has a long tradition. In 1979 Wetherell and Shannon presented a linear time algorithm for the drawing of binary trees of uniform-sized nodes. An improvement of this algorithm was given by Reingold and Tilford in 1981, whose algorithm draws isomorphic subtrees identically up to translation. 1990 Walker introduced an algorithm for the aesthetic layout of n -ary trees, allegedly running in linear time. 2002 Buchheim et al. demonstrated a runtime complexity of $O(n^2)$ for Walker's algorithm, while improving it towards $O(n)$ at the same time. Other and little known approaches can be found in [Brüggemann-Klein and Wood 1989; Eades 1992; Eades et al. 1992; Gibbons 1996; Kennedy 1996; Radack 1988].

The generalized box-drawing problem has achieved less attention. Vaucher gave a relevant algorithm with a runtime complexity of $O(nh)$, where h denotes the tree's drawing height. His work served as starting point for Bloesch, who developed both a variant with improved space requirements, $O(h)$, and a version that fulfilled all constraints claimed by Reingold and Tilford [Bloesch 1993; Vaucher 1980]. Miyadera et al. presented an $O(n^2)$ -runtime algorithm, and Hasan et al. presented the first linear-time algorithm for an acceptable number of layout constraints.

In the following we have organized the most important and commonly referred boolean layout constraints for the drawing of trees. The compilation considers all of the mentioned papers and can serve as a "unification" of the language usage; moreover, it classifies the constraints with respect to a layout problem class. Table 1 relates the constraints to algorithms while Figure 2 illustrates selected constraints.

Boolean layout constraints for binary trees with uniform-sized nodes:

- (1) The vertical coordinate of a node corresponds to its level in the tree [see 2].
- (2) A left child is positioned to the left of its parent and a right child to the right.

[2] W.l.o.g. it is assumed that the layout direction from root node to leaf nodes is top down. The constraints in [Hasan et al. 2002] presume a layout direction from left (root) to right (leaves).

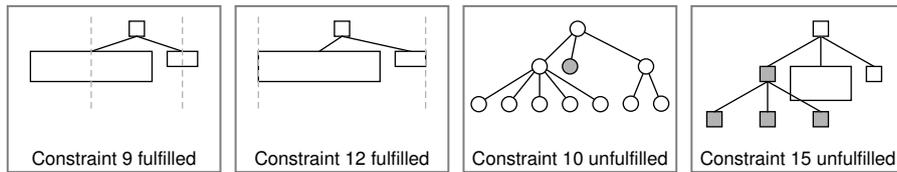


Figure 2: Illustration of selected layout constraints from our compilation.

- (3) A parent node is centered over its child nodes.
- (4) Nodes at the same level have a minimum horizontal distance.
- (5) Edges do not intersect.
- (6) Isomorphic subtrees are layouted identically up to translation.
- (7) The order of the children of a node is retained in the drawing.
- (8) If for each node the order of its children is reversed, the original and the resulting tree produce drawings that are reflections of each other.

Additional boolean layout constraints for n -ary trees with uniform-sized nodes:

- (9) A parent node is centered over the center of its leftmost and rightmost children.
- (10) Horizontal distances between nodes at the same level have minimum variance while obeying Constraints (8) and (9).

Additional boolean layout constraints for n -ary trees with variable-sized boxes:

- (11) The box of each node is separated vertically from its parent by a minimum predefined distance.
- (12) A parent node is centered between the left border of its leftmost child and the right border of its rightmost child.
- (13) The top border of the boxes corresponding to a parent's children align horizontally.
- (14) Each edge connects the center of the bottom of some parent node with the center of the top of some child node.
- (15) Edges and nodes do not intersect.

Remarks. (a) The algorithm of Reingold and Tilford is for binary trees only. (b) Constraint (1) is not reasonable for the generalized box-drawing-problem, and Constraint (11) will serve as an equivalent generalization. (c) Constraint (10) is an improvement introduced by Walker: it makes no sense for the binary tree layout problem and tightens Constraint (4), which states no restriction for the horizontal position of inner subtrees. (d) Together Constraint (4) and (11) guarantee that no two boxes overlap. (e) Constraint (9) implies Constraint (3), and (f) Constraint (15) implies Constraint (5).

3 A Meta Algorithm for the Generalized Box-Drawing Problem

Algorithms for the drawing of trees with uniform-sized nodes have been constantly improved, they are technically mature and efficient. Prevalent principle is to draw a tree recursively in a bottom up sweep, where subtrees emanating from a parent's children are drawn independently and then shifted to the right of the left subtree. Bloesch argues that this principle requires to partition the drawing into horizontal slices, and, that it is difficult to extend the principle toward trees with variable-sized nodes [Bloesch 1993].

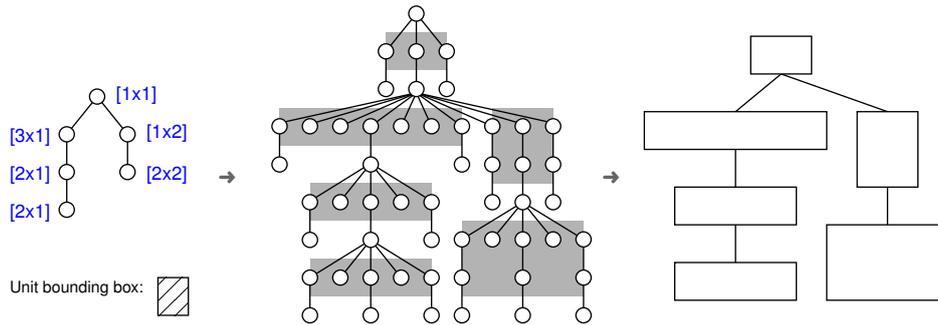


Figure 3: The algorithm illustrated at a small graph and a non-quadratic unit bounding box bb_1 .

However, things look different if the information about the bounding boxes, β , are encoded as subtree layout problems with uniform-sized nodes. Then, a standard drawing algorithm can be applied directly.

Let $\Pi = \langle T, \beta, \Gamma \rangle$ and $\Pi' = \langle T', \Gamma' \rangle$ designate the original and the reformulated drawing problem respectively; the transformation $\Pi \rightarrow \Pi'$ is accomplished with the following meta algorithm (see Figure 3 for a pictorial illustration):

- (1) *Definition of Unit Bounding Box.* The nodes in T' are of uniform size, represented by the unit bounding box bb_1 . Its actual size is subject to the content of the original boxes and the proportion of β 's extreme values. E. g., if the original bounding boxes contain text, bb_1 may not be smaller than the outline of a capital letter.
- (2) *Topology Reformulation.* For each bounding box $\beta(v)$ of a node v in T a subtree T_v is constructed. T_v has l_v leaves, from which $l_v - 3$ leaves are at level one and three leaves are at level h_v to model the outline of $\beta(v)$ and a centered bottom node. In particular, $l_v = \lceil \text{width}(\beta(v)) / \text{width}(bb_1) \rceil + 1$ and $h_v = \lceil \text{height}(\beta(v)) / \text{height}(bb_1) \rceil + 1$. T' follows the topology of T : If a node v in T has children v_1, \dots, v_k , the root nodes of the respective subtrees T_{v_i} are unified with the bottom node of T_v .
- (3) *Layout.* Application of improved Walker algorithm [Buchheim et al. 2002] to T' .
- (4) *Topology Reverse Mapping.* Substitution of the bounding boxes, β , for the subtrees introduced in Step (2). The top of a bounding box $\beta(v)$, v in T , is centered and aligned with respect to the root node of T_v .
- (5) *Rendering.* Rendering of the bounding boxes' contents with a graphics API.

Remarks. (a) Since both the topology reformulation step and the reverse mapping step require constant time for each node, the overall runtime is in $O(n)$. However, the size of bb_1 defines the horizontal and the vertical resolution of the approach, this way determining a multiplicative constant in terms of the runtime complexity. (b) Obviously all layout constraints of Reingold and Tilford as well as of Walker are fulfilled. Moreover, due to construction also all additional layout constraints for variable-sized boxes are fulfilled; this pertains in particular to the constraints (10), (12), and (15), which are not fulfilled by the special purpose algorithms from Bloesch, Miyadera et al., or Hasan et al. (c) Note that the minimum enclosing bounding box of a subtree T_v may be significantly larger than the bounding box $\beta(v)$ of the node v it represents. This is caused by the aesthetic optimization constraint (10) among others and does neither compromise the reverse mapping step nor the fulfillment of some layout constraint.

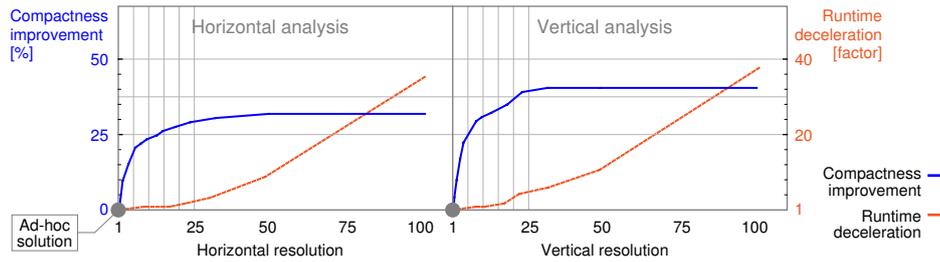


Figure 4: Trade-off between layout compactness and runtime. Due to the nature of the layout algorithm there is less optimization potential in the horizontal direction (solid curve on the left-hand side) than in the vertical direction (solid curve on the right-hand side).

3.1 Qualitative and Quantitative Analysis Results

This subsection presents analysis results to demonstrate the practical applicability of our approach with respect to both layout quality and runtime performance. The determinants of the analyzed graphs are given in Table 2; they resemble problem properties from the applications mentioned at the outset.

number of nodes in the original graph $T = \langle V, E \rangle$	1000 - 30,000
degree of a node $v \in V$	0-10, uniformly distributed
width of the bounding boxes $\beta(v), v \in V$	1-100 units, uniformly distributed
height of the bounding boxes $\beta(v), v \in V$	1-100 units, uniformly distributed

Table 2: Parameters and their distribution of the constructed random graphs.

The following questions shall be answered by the analysis:

- (1) How is the resolution, say, the size of bb_1 , related to the compactness of the layout?
- (2) What is the best trade-off between runtime effort and layout compactness?
- (3) Does the nature of the algorithm (bottom up sweep, left-right subtree positioning) prefer horizontal compactness over vertical compactness or vice versa?

Figure 4 compiles several results of our experiments. The curves average over different graphs and show the compactness improvement (solid line) against the bb_1 -resolution as well as the runtime deceleration (dashed line) against the bb_1 -resolution. The bb_1 -resolution defines how many unit bounding boxes fit horizontally (left plot) or vertically (right plot) into a tree's maximum bounding box. Take notice of the different scales: The left scale measures the improvement of the layout compactness as [%]-fraction of the saved place in relation to the ad-hoc solution; the right scale measures the runtime slowdown in multiples of the runtime of the ad-hoc solution. It is very interesting to observe that in fact the layout algorithm of Walker (or Reingold and Tilford) is biased: The horizontal compactness of the ad-hoc solution is significantly closer to the optimum than is the vertical compactness. The results show also that for a substantial compactness improvement the additional runtime effort is pretty small. These insights, which can exactly be quantified, provide valuable information to estimate the size of bb_1 within Step (1) of the meta algorithm.

While Figure 4 is mainly concerned with trade-off analyses, it may also be useful to learn about absolute runtime figures. In this connection the algorithms from Reingold and Tilford as well as from Walker (respectively Buchheim et al.) are of major interest

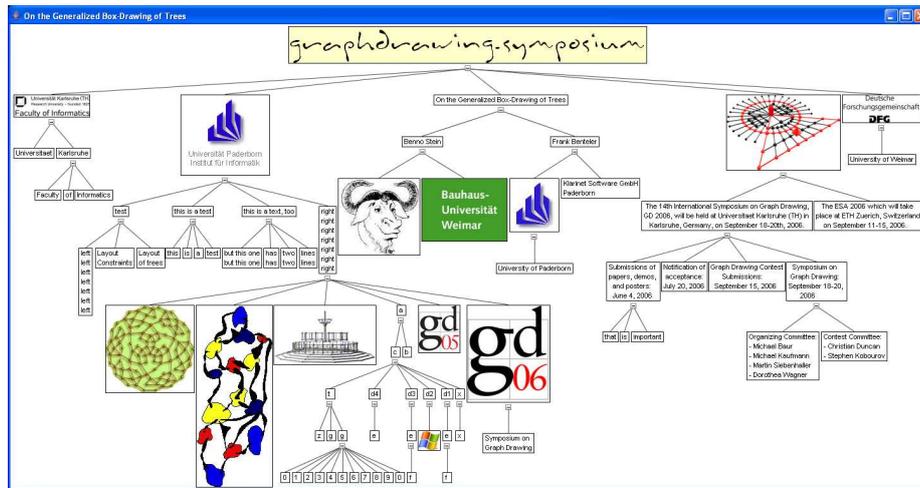


Figure 5: Snapshot of a layout generated with our implementation. Input was a GraphML file that defined a tree with about 100 nodes along with the nodes' contents.

to us. On a standard PC (2Ghz) the layout process of large trees requires approximately 100msec per 10,000 nodes with Reingold and Tilford; Buchheim et al.'s version, which additionally fulfills Constraint (10), is merely less than 1% slower. Underlying is an implementation in Java. As expected, the layout process is not the bottleneck in entire drawing process, which additionally justifies the meta drawing approach.

4 Summary

The paper discussed the generalized box-drawing problem and introduced a scalable solution whose key idea is a topology reformulation in terms of a simplified problem: the layout of trees of uniform-sized nodes. Moreover, we presented a classification of existing tree layout algorithms along with their layout constraints. Our approach fulfills more constraints than the algorithms found in the literature, and qualitative and quantitative analyses show its efficiency in complex layout settings. The advantages of our approach are threefold:

- (1) *Conformance.* All well-established layout constraints are fulfilled by virtue of the respective standard algorithm. Moreover, with the topology reformulation powerful extensions of these constraints emerge that are related to the generalized box-drawing problem. Constraint (12) and (15) may serve as examples; their validity can be inferred from the constraints that are fulfilled by the standard algorithm.
- (2) *Performance.* If the reformulation step is done in constant time c for each node, the layout step determines the overall runtime, which is linear in the original number of nodes. Our analysis shows that an almost optimum compactness is achieved with pretty small values for c .
- (3) *Sustainability.* The approach benefits from improvements in the layout step, be it the fulfillment of additional layout constraints, or a better runtime performance.

The topology reformulation principle is also suited to model new kinds of layout constraints. Examples include the enforcement of bounds for the angle of edges between

parents and children, or bounding-box-dependent distances between a node's parent or siblings. Various tasks require an efficient and high-quality solution of the generalized box-drawing problem. Our application domain is information visualization, and, in particular, the interactive navigation and browsing in large document collections, where performance is of paramount importance.

References

- [Bajaj et al. 2003] C. Bajaj, S. Khandelwal, J. Moore, V. Siddavanahalli. Interactive Symbolic Visualization of Semi-automatic Theorem Proving. *IEEE Symp. on Inf. Visualization*, 2003.
- [Bloesch 1993] A. Bloesch. Aesthetic Layout of Generalized Trees. *Software Practice and Experience*, 23(8):817–827, 1993.
- [Brüggemann-Klein and Wood 1989] A. Brüggemann-Klein and D. Wood. Drawing Trees Nicely with TeX. *Electronic Publishing*, 2(2):101–115, 1989.
- [Buchheim et al. 2002] C. Buchheim, M. Jünger, S. Leipert. Improving Walker's Algorithm to Run in Linear Time. *Graph Drawing 2002, LNCS 2528*, pp. 344–353. Springer, 2002.
- [Cruz and Tamassia 1988] I. F. Cruz and R. Tamassia. Graph Drawing Tutorial, 1988. URL <http://www.cs.brown.edu/people/rt/papers/gd-tutorial/>.
- [DiBattista et al. 1994] G. DiBattista, P. Eades, R. Tamassia, I. Tollis. Algorithms for Drawing Graphs: An Annotated Bibliography. *Computational Geometry*, 4, 1994.
- [DiBattista et al. 1999] G. DiBattista, P. Eades, R. Tamassia, I. Tollis. *Graph Drawing – Algorithms for the Visualization of Graphs*. Prentice Hall, 1999.
- [Eades 1992] P. Eades. Drawing Free Trees. *Bulletin of the Institute for Combinatorics and its Application*, 5:10–36, 1992.
- [Eades et al. 1992] P. Eades, T. Lin, Xuemin. Minimum Size h-v Drawing. *Proc. Advanced Visual Interfaces, AVI'92*, pp. 386–394. World Scientific, 1992.
- [Eklund et al. 2002] P. W. Eklund, N. Roberts, S. Green. OntoRama: Browsing RDF Ontologies Using a Hyperbolic-Style Browser. *Proc. 1st Int. Symp. on Cyber Worlds*, pp. 405–411, 2002.
- [Gibbons 1996] J. Gibbons. Deriving Tidy Drawings of Trees. *Functional Programming*, 6(3): 535–562, 1996.
- [Girba and Lanza 2004] T. Girba and M. Lanza. Visualizing and Characterizing the Evolution of Class Hierarchies. *Proc. 5th ECOOP Workshop on Object-Oriented Reengineering*, 2004.
- [Hasan et al. 2002] M. Hasan, M. S. Rahman, T. Nishizeki. A Linear Algorithm For Compact Box-Drawings Of Trees. *CCCG*, pp. 154–157, 2002.
- [Kennedy 1996] A. J. Kennedy. Drawing Trees. *Functional Programming*, 6(3):527–534, 1996.
- [Lengauer 1990] T. Lengauer. *Combinatorial Algorithms for Integrated Circuit Layout*. John Wiley & Sons, New York, 1990.
- [Miyadera et al. 1998] Y. Miyadera, K. Anzai, H. Unno, T. Yaku. Depth-First Layout Algorithm for Trees. *Information Processing Letters*, 66(4):187–194, 1998.
- [Ogura et al. 1992] K. Ogura, N. Go, M. Kishimoto, Y. Miyadera, Okada, Tsuchida, Unno, Yaku. Generation of the Hichart Program Diagrams. *Information Processing*, 15(2):293–300, 1992.
- [Radack 1988] G. M. Radack. Tidy Drawing of M-ary Trees. Report CES-88-24, Department of Computer Engineering and Science, Case Western Reserve University, Cleveland, OH, 1988.
- [Reingold and Tilford 1981] E. M. Reingold and J. S. Tilford. Tidier drawing of trees. *IEEE Transactions on Software Engineering*, 7(2):223–228, 1981.
- [Supowit and Reingold 1983] K. J. Supowit and E. M. Reingold. The Complexity of Drawing Trees Nicely. *Acta Informatica*, 18:377–392, 1983.
- [Vaucher 1980] J. Vaucher. Pretty-Printing of Trees. *Software-Practice and Experience*, 10(7): 553–561, 1980.
- [Walker 1990] J. Q. Walker. A Node-positioning Algorithm for General Trees. *Software Practice and Experience*, 20(7):685–705, 1990.
- [Wetherell and Shannon 1979] C. Wetherell and A. Shannon. Tidy Drawings of Trees. *IEEE Transactions of Software Engineering*, 5(5):514–520, 1979.

Usability of Diagrams for Group Knowledge Work: Toward an Analytic Description

Sabrina Bresciani

(University of Lugano (USI), Switzerland
sabrina.bresciani@lu.unisi.ch)

Martin J. Eppler

(University of Lugano (USI), Switzerland
martin.eppler@lu.unisi.ch)

Abstract: Diagrams are often conceived as static representations. In this paper, by contrast, we explore how conceptual graphic representations can be used as communicative devices for knowledge-intensive activities in groups. We do so by reviewing and extending existing cognitive and communicative dimensions and criteria. The aim of our research is to formalize the description of diagrams used in management discussions, analyzing their communicative and collaborative functions in order to better use them in knowledge creation and sharing in teams. We start with reviewing and merging the literature on cognitive and communicative dimensions and the literature on boundary objects, as well as information visualization, in order to compile the most relevant dimensions for knowledge creation and sharing in team. Thus, we cluster those dimensions obtaining eight macro-dimensions. We propose a formal analysis of diagrams which can be used as a tool for selecting and modifying appropriate visualizations for different knowledge-intensive activities in teams. An application example illustrates this approach.

Keywords: diagram, communicative dimensions, cognitive dimensions, visualization, knowledge visualization, knowledge work.

Categories: H.4.3, H.5.3, J.5

1 Introduction: Toward an Analytic Description of Diagrams

In recent publications, the potential of interactive, dynamic visual languages for knowledge creation and particularly sharing have been highlighted ([Maurer et al. 03], [Stubenrauch et al. 03] and [Ewenstein and Whyte 07]). Our on-going research effort that is described in this paper is part of this emerging *visual turn in knowledge management* and tries to identify the factors that make interactive visualizations effective catalysts for knowledge sharing and creation. Specifically, we aim to isolate the dimensions that increase the communication and collaboration function of diagrams in knowledge-intensive interactions. The term *dimension*, as used by Green [Green and Petre 96] and others, indicates a desired feature of an interactive diagram that supports communication and can be modified and implemented in a software application. Our research draws on three main research fields: (1) the cognitive and communicative dimensions framework literature (rooted in diagrams studies), (2) the

literature on boundary objects (rooted in management research) and (3) the information and knowledge visualization literature. The first area is based on the cognitive dimensions framework developed by Green and Blackwell [Green and Petre 96], [Blackwell et al. 01], and the communicative dimensions framework by Hundhausen [Hundhausen 04]. Our objective is to analyze, adapt, and extend those frameworks, developed originally in the context of notation and visual programming languages, to the field of team knowledge creation and sharing. The second area of research is the discourse on boundary objects [Star and Griesemer 89] as knowledge transfer and integration devices. The dimensions proposed in that stream of literature, however, have so far not been applied to interactive diagramming. A third field where we find elements for the formal description of diagrams is information and knowledge visualization, as in the works of [Shneiderman 96], [Rollett et al. 01], [Eppler 04], [Bugajska 05] and [Karabeg 06]. Based on this body of literature, we can distinguish between *basic* communicative dimensions (i.e., low level functionalities based on pre-attentive stimuli such as changing size, color, position of diagram elements) and more sophisticated, *high-level* dimensions that support sense making. In order to provide a conceptual overview and develop a theoretical basis, we focus on the latter in this paper. The scope of our research is thus to provide a framework for the analytic description of diagrams used as catalysts for knowledge processes. This formalization will help to match the usability dimensions of diagrams with communicative and collaborative requirements of various knowledge-intensive interaction types as, for example, idea creation, knowledge sharing, problem solving, assessment and deliberation. This matching process will allow us to provide a formal tool for choosing or adapting an appropriate diagram for a specific collaborative situation. In the next section, we will summarize those dimensions. In the third section we categorize them and provide a new conceptual framework. In the fourth section, we apply it to an example. Finally, in the last section, we provide directions for future research.

2 Review: Diagram Dimensions

In the table below, we present a first description of dimensions derived from the literature that highlights the cognitive, communicative and collaborative functions of diagrams [Tab. 1]. We anticipate that some of these attributes are always beneficial to communication, whereas others depend on the interaction context.

Author	Dimension	Description
Star and Griesemer 89	Plastic-robust	Plastic graphic structures to be adapted to local needs; robust structures to maintain a common identity
	Weakly-strongly structured	Weakly structured in common use; may become strongly structured in individualist use
	Abstract-concrete	Abstract refers to conceptual representation; concrete refers to specific depictions
	Different meaning - common structure	Different meaning in different social worlds, but their structure is common enough to be recognizable
	General-specific	Specific for an application; general to be understandable

	Conventionalized-customized	Conventional to be recognizable; customized to be more meaningful
Green and Petre 96	Abstraction	Types and availability of abstraction mechanisms
	Hidden dependencies	Important links between entities are not visible
	Premature commitment	Constraints on the order of doing things
	Secondary notation	Extra information in means other than formal syntax
	Viscosity	Resistance to change
	Visibility	Ability to view components easily
	Closeness of mapping	Closeness of representation to domain
	Consistency	Similar semantics are expressed in similar syntactic forms
	Diffuseness	Verbosity of language
	Error-proneness	Notation invites mistakes
	Hard mental operations	High demand on cognitive resources
	Progressive evaluation	Work-to-date can be checked at any time
	Provisionality	Degree of commitment to actions or marks
Role-expressiveness	The purpose of a component is readily inferred	
Blackwell 01	Creative Ambiguity	A notation encourages the user to see something different
	Specificity	Elements have a limited number of potential meanings
	Detail in context	How elements relate to others in the same notational layer
	Indexing	Includes elements to help the user find specific parts
	Synopsis	Understanding of the whole, "stand back and look"
	Free rides	New information is generated as a result of following the notational rules
	Useful awkwardness	Awkward interfaces force the user to reflect on the task
	Unevenness	The system pushes your ideas in a certain direction
	Lability	The notation changes shape easily
	Permissiveness	The notation allows several different ways of doing
Hundhausen 04	Programming Salience	Whatever a user focuses on during the construction of a visualization tends to become the focus of discussions
	Provisionality	Extent to which the visualization resembles a final product
	Story Content	Portrays domain concepts in terms of an underlying story
	Modifiability	Degree to which the visualization can be dynamically altered in response to the dynamics of a discussion
	Controllability	Enable a presenter to dynamically respond to the audience
	Referencability	Participants can refer to elements of the visualization
Eppler 04	Focus	Draw attention on the issue
	Coordination	Step-by-step structure to organize the interaction
	Documentation	The achieved results are documented
	Consistency	Participants can make more consistent contributions when they can see what has already been shared
	Accountability	Participant's contribution are captured and documented
	Traceability	Re-construction of the interaction and flow of ideas
Karabeg 06	Visual immediacy	The first impression; characteristic that enables the viewer to perceive and recognize "at a glance"
	Visual impetus	How attractive and inviting to action and further exploration
	Visual impedance	The hindrance; the often unintended negative implicature that causes the receiver to be less receptive

Table 1: Diagram dimension from the literature

From the table we can observe that the existing dimensions *differ* in many ways and form a heterogeneous list: they are on different levels of abstraction (such as synopsis and salience); some are binary (indexing), while others are on a spectrum (controllability); some dimensions are close to being functionalities (such as documentation), while others are more on a conceptual level (such as unevenness). Some dimensions are mainly communicative (story content), while others are cognitive (premature commitment) or collaborative (coordination). A few are always desirable (consistency), whereas some are more or less desirable (such as provisionality or hidden dependencies), depending on the task. Finally, various dimensions are strongly oriented towards electronic interaction (such as indexing) or specific contexts (such as programming salience). Given this great variety of dimensions, we attempt to reduce and structure the most relevant dimensions as explained in the next section. Then, we apply them to a real-life example in section four. This will enable us to see whether they can actually lead to improvements in the communicative and collaborative function of a diagram.

3 Reduction and classification of diagram usability dimensions

From the literature review we have obtained a long list of dimensions that differ in their level of abstraction, purpose and context, and that cannot be used for practical purposes such as designing visual knowledge management systems. We have considered classification a viable method for organizing and categorizing those dimensions, in order to be able to provide a tool for the analytic usability description of diagrams. As a first step, we have focused on high-level dimensions, eliminating the most functional dimensions, that are actually attributes and not dimensions (such as indexing, synopsis, detail in context, and most of the elements found in information visualization literature). Then, we have constructed a schema with three axes: cognitive, communicative and collaborative dimensions, on which we placed all the high-level dimensions from the analyzed literature. Based on the conceptual proximity of the dimensions emerging from the schema, we have classified them into as few groups as possible. As a rule of thumb, we have decided to consider the cognitive limit of information processing [Miller, 56] of 7 ± 2 items, in order to provide a number of clusters that is suitable for practical use. The final grouping we have obtained through this process has led us to eight groups that we named after their overall characteristics: cooperation, interaction control, abstraction, focus, modifiability, provisionality, visibility and visual attractiveness [Fig. 1]. Each of these eight *macro-dimensions* contains a number of the original dimensions found in the literature that contribute to their definition. We consider the here presented framework provisional, as we anticipate that some dimension could still be discovered, which then need to be defined, tested, and consolidated. The process through which we plan to reach this goal is explained in the last section of the paper. We now consider an example of application of the proposed framework.

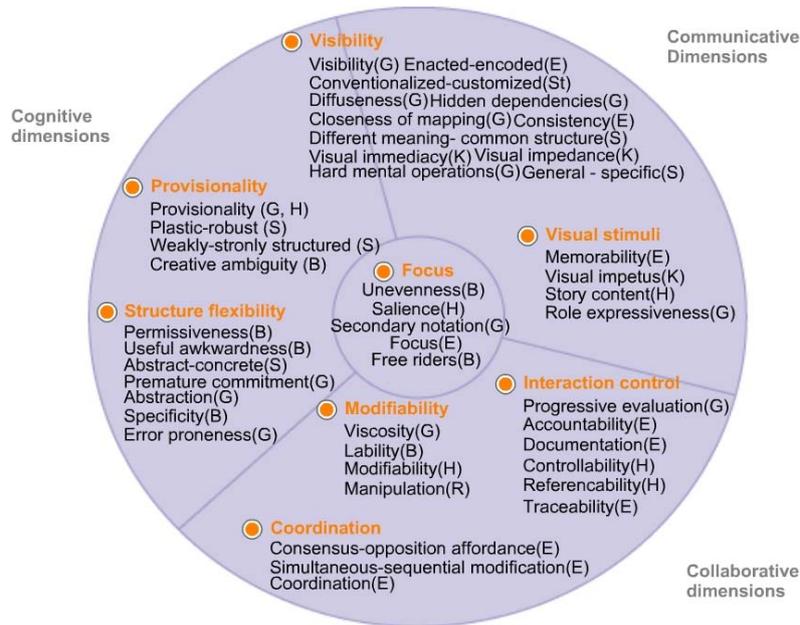


Figure 1: Classification of usability dimensions of diagrams

4 Exemplary application of the analytical usability analysis

To clarify the relevance of the previously described dimensions of diagrams, we provide an example of a formal diagram usability analysis. We consider the Gartner Hype Cycle diagram [Fig. 2] and we describe it analytically through our proposed framework in [Fig. 4]. The rating is on a five-point scale, ranging for very low (centre) to very high (external) values of the dimensions. Then, we modify the dimension values and provide a modified version with the explanation of the modifications that we have created [Fig. 3].

A Hype Cycle is a graphic representation of the maturity, adoption and business application of specific technologies. Since 1995, Gartner has used Hype Cycles to characterize the over-enthusiasm or "hype" and subsequent disappointment that typically happens with the introduction of new technologies [Gartner 07]. The diagram is used in organizations for assessment activities, as for example when making a strategic decision on a new investment in a specific technology. It supports convergent thinking, as opposed to other kinds of diagrams, such as mind maps, where divergent thinking is the goal. In its traditional format, the Hype Cycle is not a highly collaborative diagram, as the low scores on modifiability, coordination and interaction control show (and our select interviews with CIOs who use them confirmed). In our example [Fig. 3] we provide an alternative and more collaborative version of the same diagram, based on the proposed usability dimensions framework, and implemented with a software package [lets-focus 07].

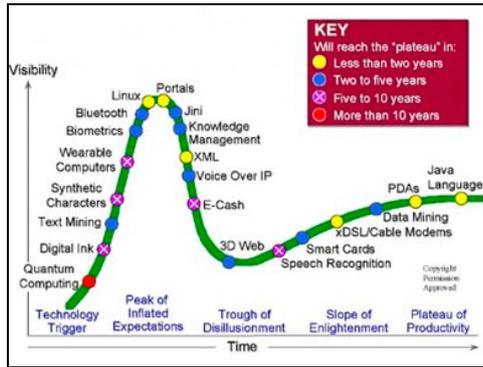


Figure 2: Gartner Hype Cycle

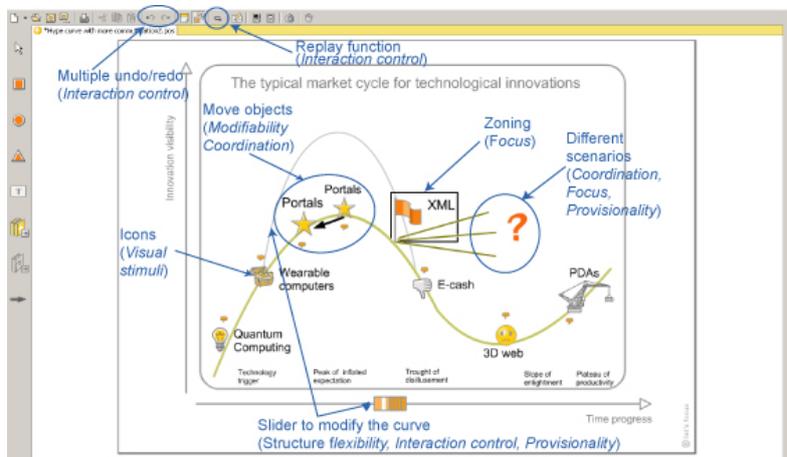


Figure 3: Gartner Hype Cycle diagram modified with desired dimensions

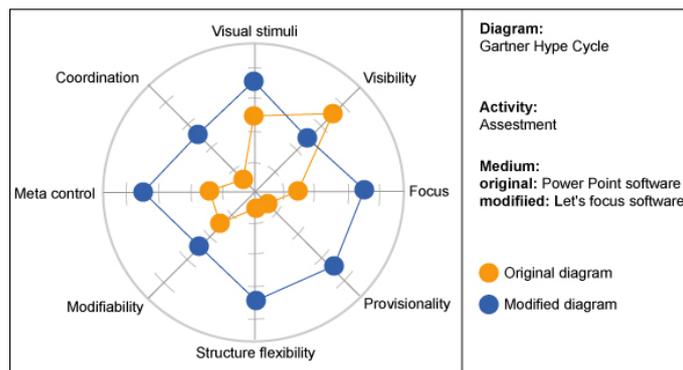


Figure 4: Original and modified Gartner Hype Cycle dimensions' rating

In detail, we have increased *visual stimuli* by providing specific icons that attract attention and are more memorable than the generic symbols used in the original version. *Visibility* - that is the ease with which the diagram is perceived at a glance- is decreased because much more items and options are present in the figure, and thus make it more complete, but also more complex to understand. Zoning, through the black square, indicates that the current debate is on XML technologies (*focus*). *Provisionality* is higher because the diagram is perceived less as a finished polished product. Higher *modifiability* is achieved through the use of a slider to modify the slope of the curve and by the possibility to move the objects in the diagram. To provide higher *structure flexibility*, the slope of the curve can be changed and different scenarios are provided. *Interaction control* is increased by allowing multiple undo/redo and the replay function, while *coordination* is supported by the multiple scenarios. From this pilot application, we have seen that simple changes in the diagram application can have significant effects on their communicative and collaborative functions and that these can be captured in the dimensions that we have proposed.

5 Outlook and conclusion

Further research is needed to refine the classification and definitions of the dimensions, to consolidate the proposed usability framework, and to match the dimensions with knowledge work types in specific organizational, collaborative settings. Appropriate and complementary research methods addressing those research purposes seem to be context-rich, but inaccurate case study analysis [Yin (03)] and context-free, but accurate in-class or natural experiments [Keppel and Wickens (04)]. Our research process started with a literature review, considering existing communicative dimensions and proposing additional ones. Then we have categorized them providing eight macro-dimensions. We now propose to conduct corporate case studies so that the dimensions can be analyzed in the context in which they become valuable. In addition, an experiment could be conducted to test and compare different communicative dimensions of diagrams and their impact on group performance. For instance a situation of knowledge asymmetry could be created in which one activity type is chosen and only one communicative dimension is manipulated or modified among groups. Post-test surveys would be given to participants to investigate their satisfaction (see [Mengis and Eppler 06] for a similar experiment). Alternatively, a focus group session could be conducted during which experienced designers discuss the role of different communicative dimensions. We also intend to investigate the tradeoffs between dimensions, to better understand the consequences of a possible design change, as described by [Green and Petre 96]. Theoretically, the contribution aims to enlarge the scope of the cognitive and communicative frameworks and adapt them to a collaborative organizational context as well as bridging the two fields of diagram research and knowledge boundary object literature. The results of this research are expected to contribute to the understanding of the function of diagrams in a managerial context as a tool for joint knowledge creation and sharing. For this purpose we need to further consolidate the dimensions, try them out on real-life applications, and improve their wording for general comprehensibility.

References

- [Blackwell et al. 01] A.F. Blackwell, C. Britton, A. Cox, T.R.G. Green, C.A. Gurr, G.F. Kadoda, M. Kutar, M. Loomes, C.L. Nehaniv, M. Petre, C. Roast, C. Roes, and A. Wong, R.M. Young, "Cognitive dimensions of notations: design tools for cognitive technology", in: M. Beynon, C.L. Nehaniv, K. Dautenhahn (Eds.), *Cognitive Technology 2001 (LNAI 2117)*, Springer, Berlin, 2001, pp. 325–341.
- [Bugajska 05] M. Bugajska, "Framework for Spatial Visual Design of Abstract Information", *Proc. Ninth International Conference on Information Visualisation (IV'05)*, (2005).
- [Eppler 04] M.J. Eppler, "Facilitating Knowledge Communication through Joint Interactive Visualization", *Proc. I-KNOW '04*, Graz, Austria, (2004).
- [Ewenstein and Whyte 07] B. Ewenstein and K. Whyte, Visual representations as 'artefacts of knowing', *Building Research & Information*, 35, 1 (2007), 81-89.
- [Gartner 07] Gartner Hype Cycles, <http://www.gartner.com/it/products/hc/hc.jsp>
- [Green Petre 96] T.R.G. Green and M. Petre, Usability Analysis of Visual Programming Environments: A 'Cognitive Dimensions' Framework, *Journal of Visual Languages and Computing*, 7 (1999), 131 – 174.
- [Hundhausen 04] C.D. Hundhausen, Using End User Visualization Environments to Mediate Conversations: A 'Communicative Dimensions' Framework, *Journal of Visual Languages & Computing*, 16, 3 (2004), 153-185.
- [Karabeg 06] A. Karabeg, "Value Creation and its Visualization in E-business", *Proc. Ninth International Conference on Information Visualisation (IV'06)*, (2006).
- [Keppel and Wickens (04)] G. Keppel and T.D. Wickens, *Design and analysis. A researcher's handbook* (4th ed.), Englewood Cliffs (NJ), Prentice Hall, 2004.
- [lets-focus 07] let's focus drawing attention, <http://en.lets-focus.com/index.php>
- [Maurer et al. 03] H. Maurer, R. Stubenrauch, D. G. Camhy, Foundations of MIRACLE: Multimedia Information Repository, A Computer-supported Language Effort, *Journal of Universal Computer Science*, Vol. 9, 4 (2003), 309-348.
- [Mengis and Eppler 06] J. Mengis and M. Eppler, Seeing versus Arguing. The Moderating Role of Collaborative Visualization in Team Knowledge Integration, *Journal of Universal Knowledge Management*, 1, 3 (2006), 151-162.
- [Shneiderman 96] B. Shneiderman, *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization*, Technical Report 96-66, Institute for Systems Research, University of Maryland, 1996.
- [Stubenrauch et al. 03] R. Stubenrauch, D. G. Camhy, J. A. Lennon, H. Maurer, Application of MIRACLE: Working with Dynamic Visual Information, *Journal of Universal Computer Science*, 9, 4 (2003), 349-367.
- [Star and Griesemer 89] S.L. Star and J.R. Griesemer, Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39, *Social Studies of Science*, 19, 3 (1989), 387-420.
- [Yin 03] R.K. Yin, *Case Study Research: Design and Methods* (3rd ed), Sage Publications, Inc (2003).

Text Mining for Indication of Changes in Long-Term Market Trends

Petr Kroha

(TU Chemnitz, Germany
kroha@informatik.tu-chemnitz.de)

Thomas Reichel

(TU Chemnitz, Germany
thomas.reichel@s2000.tu-chemnitz.de)

Björn Krellner

(TU Chemnitz, Germany
bjk@informatik.tu-chemnitz.de)

Abstract: For investment decisions the development of market trends is very important. In this contribution we present our results concerning the influence of news on market trends. We processed the stock news delivered by the Wall Street Journal with two methods of text mining – Bayes classification and grammar-driven classification. We found some potentialities of Dow Jones trend prediction and present promising results.

Key Words: text mining, classification, parsing, market forecasting

Category: H.3.3

1 Introduction

News about economic and political events that concern market and stock exchange situation are a very well-selling article. Most investors think (or believe) that markets are driven by news so that they need to read them and use them for their investment decision. There are profitable agencies selling news subscriptions of any possible kind.

Theoretically, this concept is controversial or at least disputed. The currently accepted theory of efficient markets [Fama1970] states that the information contained in any public news story is already contained also in stock prices so that it cannot bring a profit anymore. However, there is also the theory of inefficient markets [Shleifer2000] that has interesting arguments, too.

Practically, investors hope that markets are inefficient and use various strategies how to exploit the news obtained. Some of them are buying on good news, some of them are selling on good news. If the majority of investors wants to buy, the up-trend will be generated. If the majority of investors wants to sell, the down-trend will be generated. The process is complex because many investors do not have a stable behavior and they do not use fixed rules but often only an intuition.

Our motivation was to investigate the influence of news on long-term market trends by using different techniques of text mining. In this paper we compare methods that we have developed before and presented in [Kroha2005], [Kroha2006], [Kroha2007]. The first method [Kroha2005], [Kroha2006] uses Bayes classification in two modifications. Its disadvantage is that it is based on frequency of words and does not respect the structure of sentences. The next method [Kroha2007] is grammar-driven, i.e. it uses a grammar for description of classification so that it can mine more from sentences, so far from composed sentences. This method can be seen as a refinement of the first method.

The rest of the paper is organized as follows. Related work is recalled in section 2. Section 3 shortly mentions the statistical method used for text mining and Section 4 briefly describes the grammar-driven method of text mining. In Section 5 we describe the data used for experiments. In Section 6 we discuss the results achieved and we conclude in Section 7.

2 Related work

In related papers, the approach to classification of market news is similar to the approach to document relevance. Experts construct a set of keywords which they think are important for moving markets. The occurrences of such a fixed set of several hundreds of keywords will be counted in every message. The counts are then transformed into weights. Finally, the weights are the input into a prediction engine, which forecasts which class the analyzed message should be assigned to.

In [Nahm2002] a small number of documents was manually annotated (we can say indexed) and the obtained index, i.e. a set of keywords, will be induced to a large body of text to construct a large structured database for data mining. The authors work with documents containing job posting templates. A similar procedure can be found in [Macskassy2001]. The key to his approach is the user's specification to label historical documents. These data then form a training corpus to which inductive algorithms will be applied to build a text classifier.

In [Lavrenko2000] a set of news is correlated with each trend. The goal is to learn a language model correlated with the trend and use it for prediction. A language model determines the statistics of word usage patterns among the news in the training set. Once a language model has been learned for every trend, a stream of incoming news can be monitored and it can be estimated which of the known trend models is most likely to generate the story. Compared to our investigation there are two differences. One difference is that Lavrenko uses his models of trends and corresponding news only for day trading. The weak point of this approach is that it is not clear how quickly the market responds to news releases. The next difference is that our grammar-driven method respects the structure of a sentence that can have a fundamental influence on the meaning of the sentence.

In our previous work [Kroha2005] we have been experimenting with statistical methods of text classification that are based on frequency of terms to distinguish between positive news and negative news in terms of long-term market trends. In [Kroha2007] we presented a grammar-driven text mining method, i.e. we have built a grammar that describes templates typical for specific groups of news stories. Each sentence in a news story is analyzed by a parser that determines the template to which the sentence belongs. Sentences and news are classified according to these assignments.

3 Classification by Word Frequency

Our first approach to the classification of news uses methods based on word frequency, i.e. we supposed that positive and negative news each have a typical word frequency vector. We tested some of these methods and we found in [Kroha2005] that the best results in our case can be obtained by using the naive Bayes method. In our project we used about 400.000 news in German language concerning the German market represented by the index DAX. We identified positive and negative news and investigated how their ratio changes in the time. In Fig. 1 we can see changes in market and changes in news indicator, i.e. ratio $\text{positive news} / (\text{positive news} + \text{negative news})$. The next Fig. 2 shows results of the same procedure applied to news in English concerning Dow Jones Index.

In Fig. 2 there is only one crossing point between the curve of news indicator and the line representing the same amount of positive and negative news. This crossing point came 5 months before the market changed its direction. This means that starting from this date positive news were in the majority. All details are described in [Kroha2005]. It remains the problem of two news having the same word frequency but a completely different meaning:

Example 1:

News story 1: "XY company closed with a loss last year but this year will be closed with a profit".

News story 2: "XY company closed with a profit last year but this year will be closed with a loss.

(End of example)

4 Classification by Grammar-Driven Analysis of Sentences

Concerning the disadvantages of statistical methods based on word frequency illustrated in the example above we developed a grammar-driven method of text mining [Kroha2007]. This method distinguishes positive and negative news using a grammar for parsing sentences of news. Rules of the grammar describe templates that usually describe typically positive or typically negative news. Thus, the classification of each sentence can be easily done.

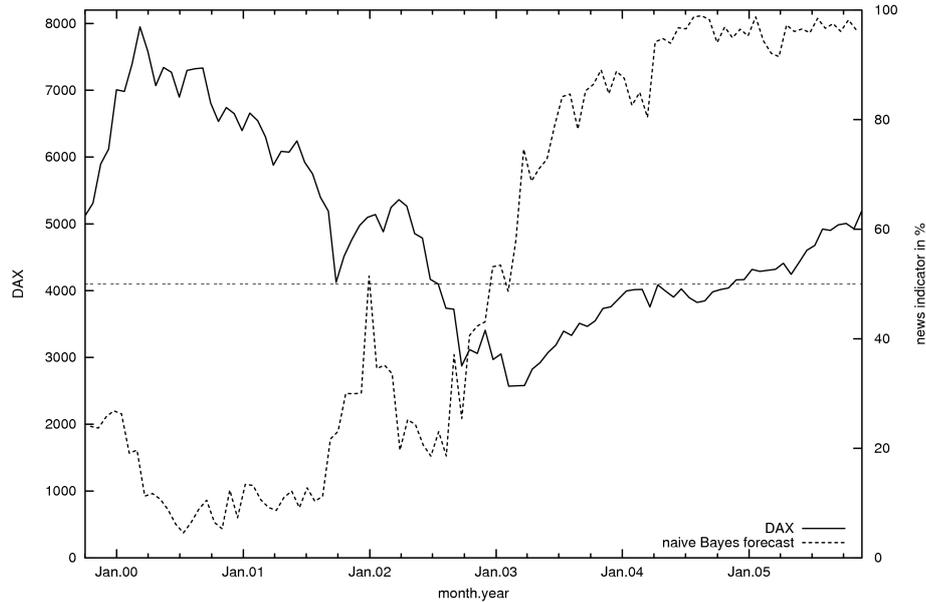


Figure 1: Results of statistical text mining of news in German and DAX. The trend of news indicator changed 2.5 months before the market trend changed.

The grammar contains a filter template that identifies sentences that cannot be classified using our grammar. There are, for example, sentences consisting of less than two words or sentences that do not contain a company name. Another filter template identifies sentences that in our opinion have no direct influence on markets. We also wrote filter templates to identify sentences that contain data we obtain from other sources, e.g. how many points Dow Jones declined yesterday. Having used these filters templates about 50 % of the messages were eliminated. The next part of the grammar contains 46 rules that represent classes of messages. They were obtained from messages of 2004 that we classified manually and used as a training set. We validated them using messages of 2002 and achieved a precision in average of 92 %. The recall has not been measured because for that it would be necessary to classify all 5000 messages of 2002 manually, which would be too time consuming. Using this part of the grammar, about 45 % of the messages were assigned to known classes, the rest was assigned to the class NotClassified. The results of the grammar-driven classification (smoothed) are shown in Fig. 3.

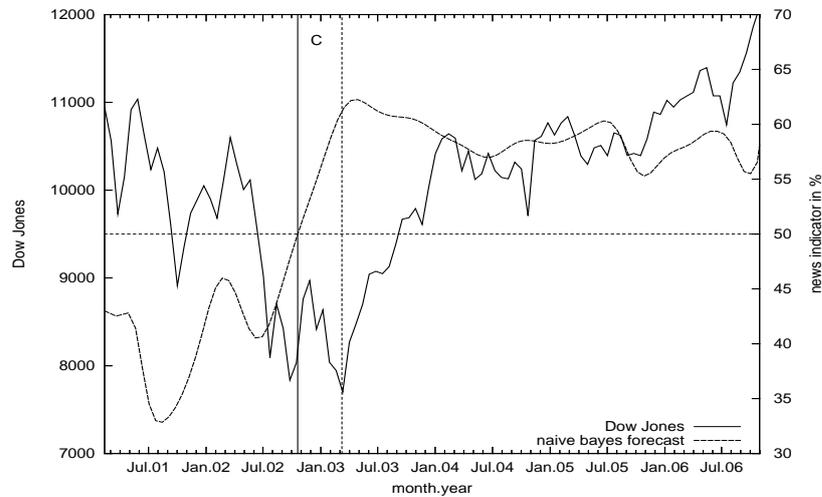


Figure 2: Results of statistical text mining of news in English and Dow Jones. The trend of news indicator changed 5 months before the market trend changed.

5 Data for Experiments

In our first attempt to statistical approach to text mining of news [Kroha2006] we used 400.000 news in German language. We achieved the results presented in Fig. 1. For experimenting with the grammar-driven text mining we used news in English of Wall Street Journal Electronic Edition. The reason was that writing a grammar for parsing positive and negative news was easier for English than for German because of the grammatical complexity of declination and conjugation in German. Of course, we wanted to compare results of both methods which is why we used the statistical method once more but for the same data as the grammar-driven method. We achieved the results presented in Fig. 2.

Altogether we had 27.720 news available in English from the time interval February, 7, 2001 to November, 17, 2006 (exactly: year 2001: 4176 news, year 2002: 5003 news, year 2003: 4937 news, year 2004: 4622 news, year 2005: 4690 news, year 2006: 4292 news).

6 Comparison and Results

We used both methods described above to state the news indicator as:

$$newsIndicator = \frac{numberOfPositiveNews}{numberOfPositiveNews + numberOfNegativeNews}$$

The results can be seen in Fig. 2, 3. The goal of the investigation was to find whether the differences between the curve of news indicator and the curve of index quotation have some features that can be used for forecasting.

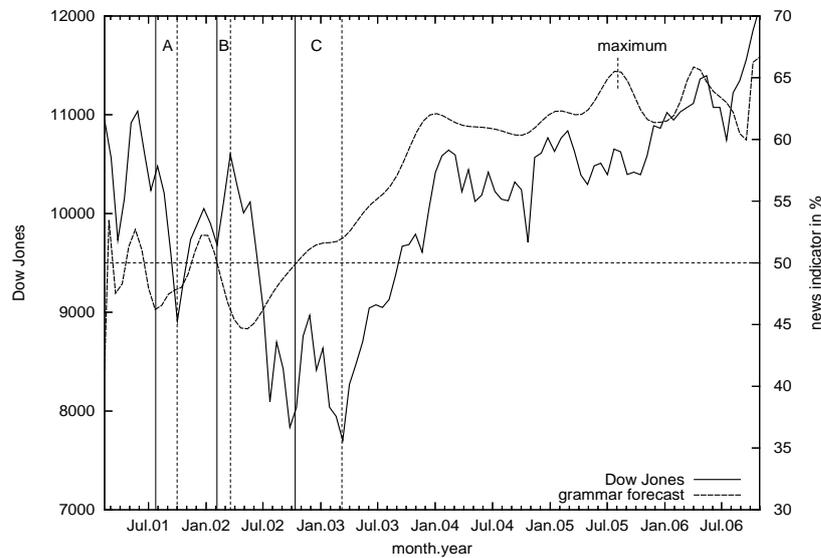


Figure 3: Results of grammar-driven text mining of news in English and Dow Jones. The changes of the news indicator precede changes of the market.

The results of the statistical method of text mining based on Bayes classification in Fig. 2 show that this method is not as sensitive as the grammar-driven method in Fig. 3. The statistical method delivered only one crossing point between the curve of the news indicator and the line representing news indicator = 50 %. This crossing point came 5 months before the market changed its direction. As we can see in Fig. 3, there are more crossing points between the curve of the

news indicator and the line representing news indicator = 50 %. The grammar-driven method is more sensitive because it uses fine semantic differences given by the sentence structure.

The region A (July, 25, 2001 - October, 3, 2001) in Fig. 3 represents a situation with local minimum of forecasting at July, 25, 2001 and local minimum of market at October, 3, 2001. However, we can see the growing region A that started in the news in July, 25, 2001 and in markets only after the September, 11 attack.

The region B (February, 2, 2002 - March, 3, 2002) in Fig. 3 represents the situation when markets are growing but the news indicator is worsening. We can see that from February, 7, 2002 Dow Jones is going up until March, 19, 2002 and then falling down. In the same time interval the news indicator is only falling. In this case the difference between forecasting and the event is about 6 weeks.

In the region C (October, 12, 2002 - March, 8, 2003) in Fig. 3 corresponds to the only crossing point from Fig. 2 but the margin representing the time difference between the forecasting and the event is about 5 months, i.e. much greater.

In the region on the right hand side from C in Fig. 3, we can see a local maximum of positive news at about the August, 2005 and a start of a growing trend of Dow Jones at about October, 2005, i.e. a time difference is here 2 months.

7 Conclusions

As we mentioned above, the news indicator processed by automatic text mining methods seems to have interesting features that could be used for forecasting because the indicator changes its trend some months before the market changed its trend.

Unfortunately, we do not have enough news in electronic form to prove this hypothesis by using market data of the last one hundred years. Points at which long-term trends change their direction are relatively rare. So, we cannot say that we can forecast a change in long-term market trend with a given probability because we cannot calculate it.

Processes running behind markets and building trends are very complex. They do not only depend on economic parameters but also on psychological features of investors and their current mood.

Investors have to differ in their interpretations of the market situation. The buying investor thinks on growing market, whereas the selling investor thinks on falling market. Investors differs also in interpretation of one news story. The same news story can generate a buy signal for one investor but a sell signal for another investor. Investors must have a controversial interpretation of the

market situation an news to trade with each other. This psychology component makes the processes even more chaotic.

Of course, there is the theoretical question whether the processes running behind markets are deterministic chaotic or stochastic chaotic and whether they are theoretically predictable or not. As we stated above there are competing hypothesis. We presume that processes running behind markets are deterministic chaotic and we suppose that there is a possibility to construct a short-time prediction of a change of a long-term trend that could have a statistically relevant importance.

However, we argue that text mining of business news can be one of the components used for market behavior prediction.

References

- [Fama1970] Fama, E.: "Efficient capital markets: A review of theory and empirical work"; *Journal of Finance*, 25, pp. 383-417, (1970).
- [Kroha2005] Kroha, P., Baeza-Yates, R.: "A Case Study: News Classification Based on Term Frequency"; *Proceedings of 16th International Conference DEXA'2005, Workshop on Theory and Applications of Knowledge Management TAKMA'2005*, pp. 428-432, IEEE Computer Society, ISBN 0-7695-2424-9, Copenhagen, (August 2005).
- [Kroha2006] Kroha, P., Baeza-Yates, R., Krellner, B.: "Text Mining of Business News for Forecasting"; *Proceedings of 17th International Conference DEXA'2006, Workshop on Theory and Applications of Knowledge Management TAKMA'2006*, pp. 171-175, IEEE Computer Society, ISBN 0-7695-2641-1, Krakow, (September 2006).
- [Kroha2007] Kroha, P., Reichel, T.: Using Grammars for Text Classification. In: Cardoso, J., Cordeiro, J., Filipe, J.(Eds.): *Proceedings of the 9th International Conference on Enterprise Information Systems ICEIS'2007, Volume Artificial Intelligence and Decision Support Systems, INSTICC with ACM SIGMIS and AAAI*, pp. 259-264, ISBN-978-972-8865-89-4, Madeira, June, 2007.
- [Lavrenko2000] Lavrenko, V., Schmill, M., Lawrie, D., Ogilvie, P., Jensen, D., Allan, J.: "Language models for financial news recommendation"; *Proceedings of the Ninth International Conference on Information and Knowledge Management*, pp. 389-396, (2000).
- [Leung1997] Leung, S.: "Automatic Stock Market Prediction from World Wide Web Data"; MPhil thesis, The Hong Kong University of Science and Technology, (1997).
- [Macskassy2001] Macskassy, S.A., Provost,F.: "Intelligent Information Triage"; *Proceedings of SIGIR'01, New Orleans, USA*, (September 2001).
- [Nahm2001] Nahm, U. Y., Mooney, R.J.: "Mining Soft-Matching Rules from Textual Data"; *Proceedings of the 17th International Joint Conference on Artificial Intelligence IJCAI-01*, (2001).
- [Nahm2002] Nahm, U. Y., Mooney, R.J.: "Text Mining with Information Extraction"; *AAAI 2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases*, Stanford, (2002).
- [Shleifer2000] Shleifer, A.: "Inefficient Markets – An Introduction to Behavioral Finance"; Oxford University Press, (2000).

Service Oriented Information Supply Model for Knowledge Workers

Bilal Zaka

(Institute for Information Systems and Computer Media Graz University of Technology Austria
bzaka@iicm.edu)

Hermann Maurer

(Institute for Information Systems and Computer Media Graz University of Technology Austria
hmaurer@iicm.edu)

Abstract: This paper suggests a powerful yet so far not used way to assist knowledge workers: while they are working on a problem, a system in the background is continuously checking to determine if similar or helpful material has not been published before, elsewhere. The technique described aims to reduce effort and time required to search relevant data on the World Wide Web by moving from a “pull” paradigm, where the user has to become active, to a “push” paradigm, where the user is notified if something relevant is found. We claim that the approach facilitates work by providing context aware passive web search, result analysis, extraction and organization of information according to the tasks at hand. Instead of conventional information retrieval enhancements we suggest a model where relevant information automatically moves into the attention field of the knowledge worker.

Keywords: Internet search, information retrieval, information supply, similarity detection, knowledge management

Categories: K.0, K.3, K.4, K.5, I.7, H.3

1 Introduction

Information search- and retrieval- processes play a vital role in the productivity of a knowledge worker. Every knowledge worker has to do extensive searches at some point in time to find information that may help, or show that certain aspects have already been covered before. Search engines provide the basic means of interaction with the massive knowledge base available on the World Wide Web.

Conventional search technology uses a pull model: i.e. search engines require an input from the user in form of a query consisting of keywords. This active search paradigm has a number of downsides: knowledge workers normally are not trained for really comprehensive searching. They may not know all the tricks required to locate the right sources of information. They may not know how to formulate a query describing all that they want to find.

The formulation of search queries is often difficult due to special terminology, or just the difference of terminology used by authors in various sources. Another constraining factor of typical search engines is the fact that they only cover shallow web contents, ignoring the almost 500 times larger, “deep” or invisible web information base [Bergman, 01]. There are special search interfaces for domain

specific searches but not all are commonly known to general users. Thus, any organization or individual pays a high price due to ineffective information discovery. According to an IDC white paper by Susan Feldman, knowledge workers spend 15% – 35% of their time searching for information. The success rate of searches is estimated at 50% or less. The study further states that knowledge workers spend more time recreating information that already exist, simply because it was not found when needed. These factors contribute to a waste of time, costing individuals and organizations a substantial amount of money [Feldman, 06].

2 Searching the web for knowledge acquisition

The primary step in knowledge work is the acquisition of information. Access to first hand information comes by means of reading literature, by meeting subject specialists, by learning technologies and finally but most importantly making use of the immense information space of the internet.

Most knowledge workers consider online technologies to be the most efficient way to access information. Web search engines no doubt are the gateway to this information base. A research study [Hölscher & Strube, 00] describing behaviour of web search by both experts and newcomers shows that when presented with information seeking tasks, the majority resorts to search engines instead of browsing direct sources such as collections of journals made available by some publishing company. Further experiments show that searchers quite frequently switch back and fourth between browsing and querying. The switching involves reformulation, reformatting of queries and change of search engines, based on previous result sets. The study also states that web search experts make more use of advanced search features such as specialized search engines, Boolean operators, search modifiers, phrase search, and proximity search options. Domain experts tend to show more creativity as far as terminology is concerned and form better and longer queries. The overall information seeking process as described by the experiments in the study shows the difficulty and the reduction of productivity in the majority of cases. A knowledge seeker is required to have good searching capabilities as well as good domain knowledge with rich vocabulary to successfully accomplish the goal. However, this is not always the case: there was and still is a big need for the enhancement of search environments for knowledge seeker.

There are attempts to facilitate searches in the form of

- i. Meta search services covering multiple search databases e.g. Jux2, Dogpile, Clusty¹ etc.
- ii. Web based and desktop tools based on search APIs to facilitate advance searching e.g. Ultraseek, WebFerret² and many web search API mashups available at ProgrammableWeb portal³
- iii. Desktop tools to explore local resources e.g. Google desktop, Yahoo desktop, Copernic⁴ etc.

¹ <http://www.jux2.com/>, <http://www.dogpile.com/>, <http://clusty.com/>

² <http://www.ultraseek.com/>, <http://www.ferretsoft.com/>

³ Mashup & web 2.0 API portal: <http://www.programmableweb.com>

⁴ Copernic search engine for desktop: <http://www.copernic.com/>

- iv. Specialized search engines, typically content specific like Google Scholar, Live Academic⁵, CiteSeer, Scirus, IEEEExplore⁶ etc. or media specific like image search, video search etc.
- v. Semantic search e.g. Swoogle, Hakia⁷ etc.

Meta search approaches provide access to multiple search sources but the process of advanced query generation is more complex due to different query formulation options of the search engines. Some times variations in results by meta search requires greater examining time and effort. Tools to facilitate searches based on APIs provide another meta search option with some query optimization facilities (spelling suggestions, synonyms support, easy use of advance search options, etc.). However, they all suffer from another drawback: the limit of allowed queries in terms of quantity and quality. A quantitative analysis [McCown & Nelson, 07] of results produced using conventional WUI (Web User Interface) and APIs, shows significant discrepancies. Results produced by search engine's own interface and APIs are rarely identical. This seems to suggest that API access probably is restricted to a smaller index. Specialized search engines provide platforms to look for information in a broader context. Locating relevant search sources and seeking information in limited indexes individually is again a time consuming task. More recent attempts to add semantic element to search suffers from the limited scope of available ontologies. The semantic web vision based on the Resource Description Framework (RDF) and Web Ontology Language (OWL) is yet to gain popularity among common content providers and developers. Information systems are still a long way from reasonable data annotation in standardized formats.

Finally, all above automation attempts fall under the same query based active search model (referred to as pull model in introductory part) and only provide surface web search. All regular searchers and particularly knowledge workers feel a strong need to go beyond these restrictions.

3 From information retrieval to information supply

In an interview, Yahoo's Vice-President for research and technology describes the next generation search to be a "search without a box". This formulation indicates a move from information retrieval towards information supply, where information comes from multiple sources, in a given context, and without actively searching [Broder, 06]. For an effective information supply, understanding the context is very important. An ideal approach for context aware search would be the use of semantic inferences. However as we have mentioned earlier even after almost eight years --- this is how long the concept of semantic web has already been around---, there is no sign of mass implementation. Billions of web pages and documents still contain no or very few annotations and pieces of meta information. Thus, a mechanism is required to effectively finding the context of information and bridge the gap between conventional and semantic web. Context can be defined as effectively interpreting the meaning of a language unit at issue. An analysis of a large web query logs [Beitzel et

⁵ <http://scholar.google.com>, <http://academic.live.com>

⁶ <http://citeseer.ist.psu.edu/>, <http://www.scirus.com/>, <http://ieeexplore.ieee.org>

⁷ <http://swoogle.umbc.edu/>, <http://www.hakia.com/>

al. 04] shows that average query length is 1.7 terms for popular queries and 2.2 terms averaged over all queries. This seems to indicate that input information is not enough to determine the context of a search.

We propose an information supply model for knowledge workers based on similarity detection. The proposed information supply method is utilized at the level where information seekers have an initial draft of their work available in written form. This input document is used to define the information need. It could be either the abstract, the initial draft of the task at hand or some document similar to the area of work currently carried out. The information supply engine performs the following processes to facilitate the search process:

3.1 Term extraction and normalization of variation

Terms can be seen as elements in a language to describe particular thoughts. They are the designators of concepts in any document. That is why automated processing of term recognition and extraction has been a critical aspect of natural language processing research. Term extraction approaches can be mainly categorized as statistical and linguistic. Statistical techniques identify terms and important phrases using factors such as consistency and structural location. Other approach makes use of linguistic morphology and syntactical analysis to identify terms of high importance. Substantial research in both techniques has transformed cutting edge research into usable applications and products. We can find examples (Yahoo Term Extraction⁸, Topicalizer and TermExtractor⁹) that successfully apply these methods to identify important terms highlighting concepts behind just textual information. Linguistic resources are used to find lexical variations in terms. Lexical variations (Synsets in WordNet¹⁰) are also used for canonical representation of information need. This normalized representation will be used for search result analysis at later stages. Terms can be extracted from knowledge work space on word count basis (fixed text chunks), paragraphs or complete documents. The extracted term batches will form queries to be processed by search services.

3.2 Determine the subject domain with the help of classification systems

In order to enhance the quality of search, additional meta data association can be very useful. There are several standardized classification initiatives in the form of taxonomies and topic maps. A subject domain can not only help to determine additional meta information to enrich search queries, but can also help in the selection of appropriate search services and results. Domain specific selections by adding this semantic element to information supply to search sources and results will reduce the “noise” (i.e. the undesirable information) generated.

3.3 Query expansion

Another approach is to use lexical enhancements. Cognitive synonyms and domain significant co-occurrences found with the help of lexical resources are used to expand

⁸ <http://developer.yahoo.com/search/content/V1/termExtraction.html>

⁹ <http://www.topicalizer.com/>, <http://lcl2.di.uniroma1.it/termextractor/>

¹⁰ WordNet: Lexical database of English, <http://wordnet.princeton.edu/>

queries. The idea behind lexical enhancements is to identify related information even if user defined terms and information terms do not match. This expansion provides an improvement in classical search where matching is based on simple content match, vector space, link analysis ranking etc. A still further step is to move to advanced query formation. Cognitive terms and phrases are organized to form complex Boolean search queries. A query routing agent determines the use of AND, OR, phrase, include, exclude, and proximity operators for a specific search interface.

3.4 Distributed search services

Our proposed information supply engine maintains an up-to-date index of general search APIs and deep web resources. Knowledge workers may configure the information supply based on general web services or include domain specific deep web search. Almost every search engine provides web service interfaces and access to its index through the XML standard. Search services combine results of common web search engines, along with deep web search engines. The processed term batches are sent as queries to search interfaces available in distributed framework of system.

A recent book by Milosevic highlights the importance of distributed information retrieval systems. The suggested framework makes use of intelligent agents to establish coordination and apply filtering strategies [Milosevic 07]. In our information supply architecture we propose the use of distributed indexing and sharing to address the restriction of view issue of web search companies and access to deep web. On the basis of success of peer to peer file sharing applications a similar indexing and searching network is envisioned. The distributed nodes working at institutional or personal level provide open search access to deep web resources. Such distributed indexing approach can have numerous other applications; one example is illustrated in Collaborative Plagiarism Detection Network architecture [Zaka, 07][CPDNet, 07]. The Web presents a huge and continuously growing information base, but “search engines that crawl the surface of the web are picking up only a small fraction of the great content that is out there. Moreover, some of the richest and most interesting content can not even be crawled and indexed by one search engine or navigated by one relevancy algorithm alone” [OpenSearch, 07]. Open and distributed search initiatives provide common means of search result syndication from hundreds of shallow and deep web search engines.

3.5 Result analysis with the help of similarity detection

In conventional active search models the iterative process of obtaining and quickly examining the results provided by search engine consumes a lot of time. Information supply for knowledge worker provides an automated analysis of result using similarity detection techniques. The resulting URIs are fetched by an analysis module. The fetching module uses PHP's cURL library to extract textual contents and removes the formatting information (HTML tags, scripts etc.). Term extraction and lexical data services are used again to obtain the gist of contents. Similarity is calculated between information need and processed result contents with the help of vector space mathematics. The similarity detection service is fed with term batch of knowledge workspace and the terms of search result URIs. The terms/words are mapped as vectors against a compound term vocabulary. The dot product of two

vectors (cosine similarity) determines the relevance. The semantic space for result analysis is built with the help of two additional services, namely POS (Part of Speech) tagger and Synonym [CPDNet 07]. POS tagger returns the sense of each word (i.e. Verb, Noun etc.) and Synonym service based on WordNet 3.0 returns the related synonyms and Synset IDs. The normalisation or canonical lexical representation of terms introduces a semantic relevance element in similarity detection process. A paper on power of normalized word vectors [Williams 06] presents the described concept in detail. Such analysis capability can replace typical result browsing and filtering in information seeking process.

3.6 Result mapping to knowledge space

Key terms and phrases of the knowledge work space (input document) are linked with matching information sources. Result mapping with a rich web user interface provides a clear picture of relevant documents. The links include subject domain, key terms, persistent phrases, and summary of matching source. This function provides far better first hand knowledge about information source than the simple hit-highlighting as is done by a normal search engine. Users will have a higher degree of knowledge whether to further investigate matches suggested or not.

4 Service oriented model

The model shown in Figure 1 is a step towards the realization of a comprehensive information supply system. In order to provide the functions described in the previous section, we emphasize the use of off-the shelf tools in form of web services. Combination of external and internal services provides the depth in search and automation in information supply and analysis. The first level in the information supply environment is to determine the information need with the help of term extraction. The knowledge work space is submitted to term extraction services. A term extraction process in information supply model consists of converting formatted contents into plain text. The plain text is passed to a Part of Speech (POS) tagger in order to determine word sense and to eliminate function words. Key terms (verbs and nouns with higher occurrence frequencies) are further enriched using linguistic resources: WordNet and Wortschatz¹¹ lexical databases are used to get synonym terms and synonym group IDs with similar sense. This data is used for query expansion and canonical lexical representation. The initial level also attempts to relate information requirement to a subject domain. Initially the subject domain is determined with the help of predefined standard taxonomies. One example is the use of ACM Computing Classification System¹²: the keyword index of each category can be checked for similarity with extracted terms. A web service developed to calculate angular measure among two word/term vectors mapped in compound term vocabulary is used for similarity check. [CPDNet 07]

¹¹ Wortschatz: <http://wortschatz.uni-leipzig.de/>

¹² ACM Classification: <http://www.acm.org/class/>

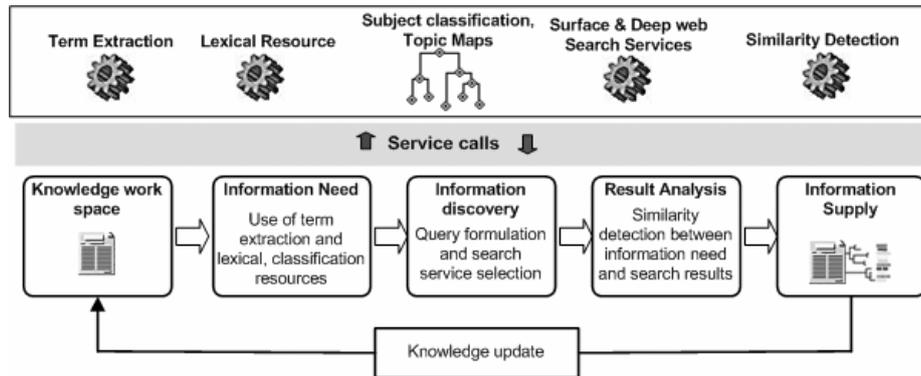


Figure 1: Information supply model for knowledge workers

After formation of the consolidated information need, the system selects available search services. The mashup of search services provides a broader coverage of distributed indexes of the shallow and the deep web. A distributed search test bed [CPDNet 07] is used in the proposed information supply model. The distributed search framework acts as a proxy for search services. It not only includes popular search APIs like Google and Microsoft Live but also support OpenSearch and peer search. The search results are filtered at the next level with the use of already mentioned similarity detection techniques. Cosine similarity measure is determined from term vector of knowledge workspace and term vector of returned results from search services. The term vectors can be additionally presented in normalized form (canonical lexical representation) in order to develop semantic relevance. Filtered results with high angular similarity measure are linked with the knowledge workspace. Knowledge workers can see first hand similar data available on the internet. With an update of the knowledge space by incorporating new relevant documents found users can initiate the same process for an updated information supply.

5 Conclusion and further work

In this paper we are suggesting the systematic use of web services for an information supply model. The model presented is a step forward from classical IR to proactive search systems. We introduce the use of lexical services and similarity detection to (i) find the context of a search and to map the information need to a specific subject domain, and (ii) provide an automated result scanning service, similar to human judgment. Both elements play an essential role in efficient information supply for knowledge workers. The research indicates that additional meta information found via context of search and lexical resources can prove to be very useful in the creation of automatic search queries. The use of mathematical techniques to determine information relevancy can also eliminate a time consuming and iterative process of manually scanning search results. Individual components/services of model available at CPDNet platform [CPDNet 07] and their performance to accurately link relevant

resources provide promising avenue for complete implementation of presented information supply system.

Our experiments of search service mashup indicated very good possibilities of search access across multiple jurisdictions. The research on information retrieval in the internet and role of search engines also pointed out issues of restrictions in general web search APIs. The success of the next generation web or web 2.0 depends not only on the collaborative efforts from users but also on open and honest syndication and standard API provision by enterprises. The discontinuation of application consumable search service by Google, no availability of search syndication by specialized search engines like Google Scholar and Live Academia are examples of undesirable restrictions against which the community should protest before it is too late. There is a strong requirement for a scalable and open search and indexing platform. In order to develop such a platform, use of peer to peer search with user or institute level indexing is worth serious consideration.

References

- [Beitzel et al. 04] S. M. Beitzel, E. C. Jensen, A. Chowdhury, D. Grossman "Hourly Analysis of a Very Large Topically Categorized Web Query Log", In Proc. of 2004 ACM Conf. on Research and Development in Information Retrieval (SIGIR-2004), Sheffield, UK, July 2004.
- [Bergman, 01] M. K. Bergman. The deep web: Surfacing hidden value, <http://www.press.umich.edu/jep/07-01/bergman.html> (Accessed April 07, 2007).
- [Broder, 06] A. Broder. "Search without a Box" (Interview). Yahoo! Search Blog. March 09, 2006; <http://www.ysearchblog.com/archives/000262.html> (Accessed April 2007)
- [CPDNet, 07] Collaborative Plagiarism Detection Network (CPDNet) Web Services: Online at: <http://www.cpdnet.org/nservices.php?> (Accessed April 26, 2007)
- [Feldman, 06] S. Feldman. "The Hidden Costs of Information Work" (IDC #201334, April 2006) <http://www.idc.com/getdoc.jsp?containerId=201334> (Accessed April 2007).
- [Hölscher & Strube, 00] C. Hölscher, G. Strube. "Web search behavior of internet experts and newbies" *Computer Networks*, Vol. 33, Issues 1-6, June 2000, Pages 337-346.
- [Maurer & Tochtermann, 02] H. Maurer, K. Tochtermann. "On a New Powerful Model for Knowledge Management and its Applications" *J.UCS* vol.8., No.1, 85-96.
- [McCown & Nelson, 07] F. McCown, M. L. Nelson. "Agreeing to Disagree: Search Engines and their Public Interfaces" Proc. of Joint Conference on Digital Libraries (JCDL) 2007.
- [Milosevic 07] D. Milosevic "Beyond Centralised Search Engines" An Agent-Based Filtering Framework, VDM Verlag Dr. Müller 2007 ISBN: 978-3-8364-1222-3
- [OpenSearch, 07] Open Search: Documentation, Online at <http://www.opensearch.org/> (Accessed April 26, 2007)
- [Williams 06] R. Williams "The Power of Normalised Word Vectors for Automatically Grading Essays" *The Journal of Issues in Informing Science and Information Technology* Volume 3, 2006 pp. 721-730
- [Zaka, 07] B. Zaka. "Empowering plagiarism detection with web services enabled collaborative network" submitted for publication

An Interactive Visualization Model for Competence Management: an Integrative Approach

Lena Pareto

(University West, Sweden)

(lena.pareto@hv.se)

Ulrika Lundh Snis

(University West, Sweden)

(ulrika.snis@hv.se)

Abstract: This research aims at innovative IT support for competence management (CM) integrating organizational, group and individual perspectives. Our method is action case research resulting in a design model. This paper addresses our overall approach towards an integrative CM system, which is usage-oriented. Based on in-depth analysis of seven business companies' CM activities, we present a new interactive visualization model for CM. The model is suitable for explorative analysis and for communicating competence situations at individual, group or organizational levels. It provides support for constructing competence information incrementally. We view competence descriptions as negotiable estimations with varying verification levels, and use a competence representation which deals with uncertainties of estimations. Our view on competence is novel, and has far reaching consequences. Future work includes a running prototype and planned pilot studies.

Keywords: Competence management, competence representation, interactive visualization, graphical user interface, action case research, ontology-generic

Categories: H.5.2, H.5.3, H.4.3

1 Introduction

Competence management (CM) refers to how competence is managed with respect to organizations, groups and individuals [Berio & Harzallah 2005]. The organizational view has shifted from using the term *qualifications* to the broader term *competence* [Boucher et al. 2007]. Much research emphasises the organisational perspective related to strategic planning, involving identification of competence gaps and strategies to fill these gaps [Lindgren et al. 2004]. They found that for organisational CM activities, skill-based competence descriptions are appropriate and often used. For individual CM activities, job-based descriptions are common. There is a tension between these two views: i) skill-based, as personal descriptions with a present and future focus and; ii) job-based, as positional descriptions with a focus on the past. According to [Lindgren et al. 2004], organizational needs and individual interests should be integrated but there are few information systems designed to support CM at both levels. Integrating these perspectives is one of our aims, as is the initiative in [Berio & Harzallah 2007]. Moreover, visualization and flexible reporting of competence information is a key design challenge [Lindgren et al. 2004].

We base our findings on a research collaboration project conducted together with Networking Companies: a network of consultancy firms, service- and knowledge-based organizations, as well as more traditional production companies.

They all have varying approaches to CM and uses different systems for support. Our challenge is to identify and develop IT support for CM beneficial for individual companies as well as for the network as a whole. One initiative is the visualization model presented here. Due to the companies' diversity of needs, IT support and CM processes used, our solution must work as add-on to existing CM systems (CMS) as well as standalone. It must be flexible with respect to the supported competence information as well as general with respect to competence processes. Our approach is in the spirit of [Ahlberg 1996]: a graphical, explorative information environment, in which data can be examined interactively. Our ambition is similar to [Lindgren et al. 2004], regarding the empowerment of individuals.

The goal of this paper is to develop a flexible interactive visualization model to support competence management from the perspective of organizations, groups, and individuals. In doing so, the following research dimensions are addressed:

- The role and purpose of CMS within organizations.
- The roles of IT support in competence management workflows.
- The view of competence representation in CMS.
- Requirements on competence management systems.

Our research approach resembles action case research - a combination of action research and case study resulting in a design model as described in [Yen et al. 2002]. The paper is organised as follows: Related CMS, Empirical Findings, Our Approach, The Visualization Model, Conclusion and Future Work.

2 Related Competence Management Systems

Most CMS systems are used as storage for competence descriptions of employees, and are mainly used by human resource people and management. Individuals may be responsible for data entry and updating information but there are little incentives for doing so. Consequently, information is often inaccurate and incomplete, and therefore less useful in practice [Lindgren et al. 2004]. It is a challenge to create acceptance for competence definitions, and to keep information dynamic and relevant [Lindgren 2002]. According to [Harzallah et al. 2006], existing tools are inflexible and no ontology supports a common understanding of competence.

There are many attempts to describe competence. [Boucher et al. 2007] use the levels *individual competence*, *collective competence* (a group of persons) and *global competence* (organisational ability). In [Harzallah et al. 2006] *competence required* and *competence acquired* is used. Structures are often fixed and not open for changing contexts, they claim. In [Lindgren et al. 2004] a time-dimension is added: not only the past (Competence-in-the-store) and the present (Competence-in-the-use) but also the future (Competence-in-the-making). In [Berio & Harzallah 2005] four main processes for competence management are identified: *competence identification* (identify required competencies), *competence assessment* (assess acquired competence), *competence acquisition* (how required competencies can be acquired) and *competence usage* (how to use the knowledge of acquired and required competence). In this work we focus on *usage*, but competence assessment and identification are involved as well.

2.1 Requirements of Competence Management Systems from Literature

We have found the following requirements and recommendations from CMS literature, regarding integrative aspects, functionality and competence representation: There should be explicit links between company strategies and required competencies [Harzallah et al. 2006; Lindgren et al. 2004] and individual competence interest *and* organizational needs should be integrated. Standardized and sharable representation of competence is needed [Harzallah et al. 2006]. Flexible reporting as well as possibilities of making distinctions between past, present and future competencies is important [Berio & Harzallah 2005; Lindgren et al. 2004]. Regarding functionality, it is important to see competence development over time for acquired, required and desired competence, from the organizational, group and individual perspective respectively [Lindgren 2002]. Thus, status of acquired and required competence should be represented [Harzallah et al. 2006]. Providing features to query (search) the competence information for required competencies is also needed [Lindgren 2002]. Concerning competence representation, formal and informal indicators in competence descriptions should be balanced; the person behind a chart/diagram should be visible and user control is important for keeping competence data up-to-date. Finally, CMSs should provide flexible visualizations of competences and interests [Lindgren 2002].

3 Empirical Findings

The collaboration with Networking Companies is on-going and started one year ago. The network management and 7 member companies are actively involved in the research project. We have conducted in-depth interviews with the companies, held regular half-day meetings and three 2-day workshops. We have conducted a survey to determine interest in a CMS from the individual, group, organisational and network perspectives. Our empirical findings so far are reported in [Lundh Snis et al 2007] and can be summarized as follows: Awareness of CM is high in all companies, but the level of systemization varies. Common ambitions include to work with CM as a continuous process; to have more efficient tools integrated with current activities; to be more strategic and long sighted; to better integrate individual, group and organizational levels of CM; to get “snapshots” of competence status; to have more explicit competence profiles; to get competence profiles accepted; and to make employees engage in competence development. The amount of work required to produce competence information in most CMSs is considered a barrier to IT use. Most companies have competence profiles and descriptions of positions documented in business software or elsewhere, but need support for competence management.

4 Our Approach Towards an Integrative CMS

We take a *user- and usage-oriented approach* to an integrative CMS, with a skilled- and verification based *view on competence representation*. We have designed an *interactive visualization model* which is *ontology-generic* and allows for *incremental construction* of competence information.

4.1 User- and Usage-Oriented Approach

Competence information *usage* is our focus. The key question is: What is the competence information used for? We argue that the main interest for both management and individuals is the future. Organizational interests include preparation for future challenges and utilizing human resources in accordance with individuals' capacities, ambitions and interests. Individuals' interests include personal career options and stay up-to-date with professional development to be attractive in the job market. Strategic management and individual career planning are complementary, and integration constitutes a basis for empowerment of the individual [Lindgren 2002].

4.2 View on Competence Representation

Competence is a complex notion, but must be represented in some way in IT systems. Our position is in accordance with [Höfferer & Hiermann 2003]: competence of an employee can never be fully reproduced in a system. Facts describing the past are unproblematic, but a future focus involves competencies such as abilities, behaviours, transferability of knowledge in new situations etc. Such competencies are subject to *judgement* and can only be estimations, which are difficult to value. Our approach is to accompany estimations with "arguments", i.e. *verification objects*. These objects reflect the basis for estimation, and can include formal qualifications, experiences, claimed talents or abilities. Someone might argue that being the head of a large household contributes to managerial competence, and in [Lindgren et al. 2004] the information that someone lived in Germany for several years was considered important information regarding language competence. Due to the complexity and uncertainty about what constitutes competence, our approach is to include verification objects as motivations of the estimations, and to make these available to the viewer.

Regarding the problem of correctness and completeness of competence information data, we take the opposite approach to [Berio & Harzallah 2005]: we allow imperfect data. Aligned with our view of competence as more or less verified estimations, we provide the user with a *verification level*: the extent of verifications. The purpose of the verification level is two-folded: to reflect the level of uncertainty of the estimation and to stimulate verification of claimed competencies. The degree of accordance with real conditions is crucial for information usefulness.

4.3 Ontology-generic

A flexible extension of competencies must be possible at all times [Höfferer & Hiermann 2003]. We provide a framework where any hierarchical structure (for competence descriptions) based on any scale (for competence estimations) can be defined and altered as appropriate. It is a three-layer system with competence data as the base level, a data organizer as the middle level, and a visualization model as the top level. This way, competence data can be imported from various systems. In the data organizer competence hierarchies are defined, and data (verification objects) are related to their corresponding competencies. Estimations and verification levels are aggregated to the level above. Several structure definitions can be associated with the same basic data. The visualization model graphically presents the data organized in accordance with the chosen structure definition.

4.4 Incremental Construction of Competence Information

To lower the threshold for data entrance (identified as important in our empirical findings and in [Höfferer & Hiermann 2003]), we searched for solutions providing incremental construction of information. Just as structures can emerge over time, our model allows for competence information to be provided incrementally. This is due to the combination of emergent structures, and the notion of verification level. For example, a first estimation of a person's competence can be done immediately by giving quick estimations in a few important categories, without any subcategories or verification objects. Naturally, such estimation has low credibility, due to the lack of detailed information and verifications - but the uncertainty is reflected in the verification level and is visible in the graphical representation.

4.5 Interactive Visualization Model

Information visualization can provide useful tools and techniques for gaining insight and understanding of data [Zuk et al. 2006], and can amplify cognition [Card et al. 1999]. The purpose of our visualization model is to display competence structures and estimated competence levels for individuals, aggregated to group and organizations. It is based on a hierarchical, details-on-demand graphical representation of data. Effective overviews of data are equally important as rich descriptions of specific data elements [Ahlberg 1996]. Only showing details on requested are instrumental for the concept. To enable an explorative support for analysis and decision making at all three levels, a direct manipulation [Schneiderman 1983] model is used. Direct manipulation provides efficient and intuitive interaction methods [Baudel 2006].

5 The Visualization Model

The visualization model is based on a new graphical diagram (see Figure 1), which is a modification of the radar diagram in Excel. The "spider web" is divided into compartments, one per main category. The levels in the compartments represent estimation levels. Each triangle represents a competency where the height represents estimated competence level and the opacity illustrates verification level. Opacity is a natural mapping of uncertainty [Tufte 1990]. Levels are aggregated. A compound competence can be unfolded and folded, to provide details-on-demand (diagram B and C). An atomic competence can display verification objects (diagram D). The colour coding is optional. It differs from radar diagrams by being hierarchical, and to provide the detail-on-demand principle. This representation respects the semantic notion of levels being independent of their surrounding levels, which is not true for radar diagrams. Semantic soundness is an important quality of visualizations.

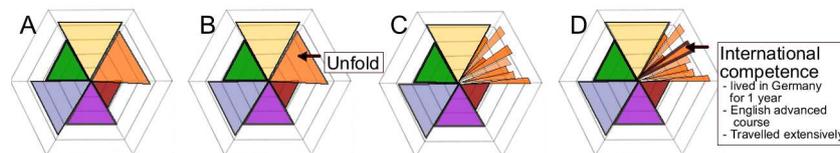


Figure 1. The hierarchical, graphical visualization model

Figure 2 shows an illustration of how the visualization model can support activities such as employee-manager meetings discussing individual progress and competence development plans. Diagram A illustrates the competence status from last year, diagram B the competence goal for this year, and diagram C the difference between the two. Note that with our model, three different dimensions of competence development are present (see diagram C): *competence deepening*, *competence broadening*, and *experience deepening*. Competence deepening refers to increasing the competence level (the enlarged triangle at the bottom), broadening of new emerging competence areas (the new thin triangle in the upper-right category) and experience deepening (new verification objects and the added transparent triangle at the same level). Note that diagram B is the sum of A and C, both semantically and graphically. Diagram D illustrates a comparison of two competence profiles in the same diagram, comparing actual status and planned status in this case.

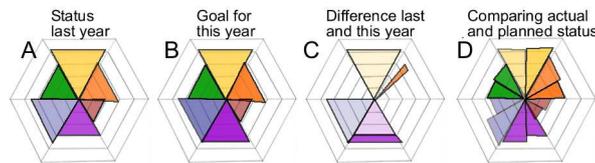


Figure 2. The visualization model when used in CM activities

Figure 3 illustrates different ways of specifying search for matching profiles: where A is an approximate goal profile, B is a requirement specification of certain competencies, and C is criteria for min and max values of competence levels.

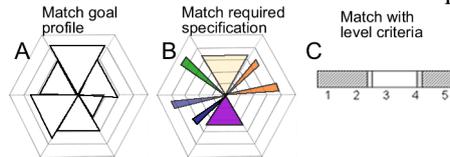


Figure 3. Matching principles

In Figure 4, searching and search results are illustrated, in a different view. In diagram A results are presented as individual profiles. Diagram B shows a comparative view where all results are presented in one large diagram and each category is sorted on competence and verification level, in that order. This can be used for showing aggregated group or organizational competence, or as a result of a search. Diagram C is the same as B with one individual selected, shown by the marked triangles.

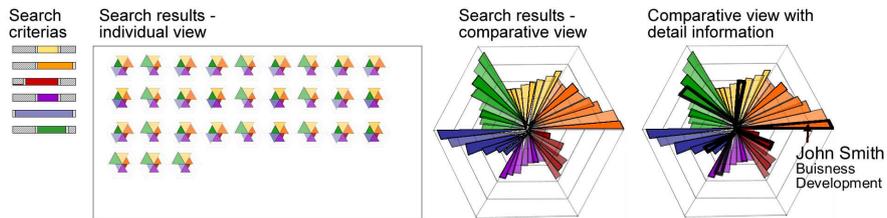


Figure 4. Presentation of search results

The visualization model is general since it can display a wide range of data, and it is constructed from a few basic operations: show a profile, fold- and unfold, show details, add or compute the difference of profiles, compare one or several profiles side by side or in a large diagram, and different ways of specifying a search. Development over time as a sequence of diagrams is achieved by appropriate search criteria.

5.1 Summary of Visualization Model Benefits

The visualization model is a communication tool in the sense that a wide range of competence profiles can be displayed, compared and explored interactively. Different views of the same profile, different profiles or different points in time can be shown and compared, which is valuable as a mean to communicate different competence situations according to [Lindgren et al. 2004]. Acquired, required, planned for and desired competence can all be described with the same type of profile and in the same framework. It allows for a transparent usage: competence profiles of individuals, groups or organizations are all in the same framework. All variations use competence estimation together with verification objects. For planned competence development or interest declarations, verification objects can be used as planned learning activities or motivations of the interest value to the company, respectively. The problem of competence scales being too instrumental is highly reduced by providing verification objects. Employees are empowered since they can be more aware of their role in the organization, more aware of company strategic directions, more in control of opportunities and potential threats in advance. Users may see transparency as an opportunity to market their competence rather than as a threat, allowing for more active development either motivated by their own competence interest or organizational competence needs [Lindgren et al. 2004].

6 Conclusion and Future Work

The purpose of our CMS is to integrate organizational strategic planning, group planning and individual career planning, in the same transparent framework. The role of IT is to provide support for communicating situations and for explorative analyses in current CM activities. Our view on competence as negotiable estimations with varying verification levels is novel and has far reaching consequences. It raises fundamental questions about competence and challenges the current division of power in many enterprises. Future work includes finishing a working prototype and conducting pilot studies with two companies: one which will evaluate incremental construction of competence information from scratch, and the other which will evaluate the visualization model as an add-on to an existing CMS. Both evaluations involve visualization effects, which is challenging and require unusual methods according to [Plaisant 2004].

References

[Ahlberg 1996] Ahlberg, C. 1996. Spotfire: an information exploration environment. *SIGMOD Rec.* 25, 4 (Dec. 1996), 25-29.

- [Baudel 2006] Baudel, T. 2006. From information visualization to direct manipulation: extending a generic visualization framework for the interactive editing of large datasets. In *Proceedings of the 19th ACM Symp UIST '06*, 2006. ACM Press, New York, NY, 67-76.
- [Berio & Harzallah 2005] Berio, G. Harzallah, M. 2005. Knowledge management for competence management, In *Proceedings of I-KNOW'05 - 5th International Conference on Knowledge Management*, Graz Austria, 2005.
- [Berio & Harzallah 2007] Berio, G. and Harzallah, M. 2007. Towards an integrating architecture for competence management. *Comput. Ind.* 58, 2 (Feb. 2007), 199-209.
- [Boucher et al. 2007] Boucher, X. Bonjour E, Grabot, B. 2007. Formalisation and use of competencies for industrial performance optimisation: A survey. *Comp In Industry* 58(2) (2007) 98-117.
- [Harzallah et al. 2006] Harzallah, M. Berio, G. Vernadat, F. 2006. Analysis and modelling of individual competencies: towards better management of human resources, *IEEE Transactions on Systems, Man and Cybernetics: Part A* 36 (1) (2006).
- [Höfferer & Hiermann 2003] Höfferer, M Hiermann, W. 2003. BEKO-SMS: a practical Knowledge-based Approach to Skill Management. In *Proceedings of I-Know 03 - Third International Conference on Knowledge Management*, Graz Austria, 2003.
- [Card et al. 1999] Card, S. K. Mackinlay, J. D and Shneiderman, B. 1999. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco, CA, US.
- [Lindgren et al. 2004] Lindgren, R., O. Henfridsson, and U. Schultze, 2004. Design Principles for Competence Management Systems: A Synthesis of an Action Research Study, *MIS Quarterly* Vol. 28 No. 3, pp. 435-472.
- [Lindgren 2002] Lindgren, R. 2002. Competence Visualizer: Generating Competence Patterns of Organizational Groups. In *Proceedings of the 35th Annual Hawaii international Conference on System Sciences (Hicss'02)-Volume 4 - Volume 4* (January 07 - 10, 2002). HICSS. IEEE Computer Society, Washington, DC, 108.
- [Lundh Snis et al 2007] Lundh Snis U, Pareto L, & Svensson L, 2007. Competence Management Systems in Networking Organisations: Designing for Empowerment? To appear in *Proceedings of IRIS'30* (Tampere, Finland, August 11-14, 2007).
- [Plaisant 2004] Plaisant, C. 2004. The challenge of information visualization evaluation. In *Proceedings of the Working Conference on Advanced Visual interfaces* (Gallipoli, Italy, May 25 - 28, 2004). AVI '04. ACM Press, New York, NY, 109-116.
- [Schneiderman 1983] Schneiderman, B. 1983. Direct Manipulation: a Step beyond programming languages. *IEEE Computer*, Vol. 16, No. 8 (August 1983), pp.57-69.
- [Tufte 1990] Tufte, E. 1990. "Envisioning Information" Graphics Press, 1990.
- [Yen et al. 2002] Yen, C-C., Woolley, M. & Hsieh, K-J. 2002. Action case research: a method for the accumulation of design theory/practice knowledge in practice. *Working Papers in Art and Design* 2, ISSN 1466-4917.
- [Zuk et al. 2006] Zuk, T., Schlesier, L., Neumann, P., Hancock, M. S., and Carpendale, S. 2006. Heuristics for information visualization evaluation. In *Proceedings of the 2006 AVI workshop on Beyond Time and Errors: Novel Evaluation Methods For information Visualization* (Venice, Italy, May 23 - 23, 2006). BELIV '06. ACM Press, New York, NY, 1-6.

Visualizing Semantic Structures to Support Reading

Yongqiang Qiu

(M3C Lab, University of Bolton, UK
yqq1ect@bolton.ac.uk)

Adel Elsayed

(M3C Lab, University of Bolton, UK
a.elsayed@bolton.ac.uk)

Abstract: Reading text for knowledge acquisition is a linear process in which a user reads verbal content by scanning text in a sequential manner. It is an established fact that computer users prefer to read text from papers than off computer screens. Hence, computers must take a more active role in supporting knowledge acquisition through reading than just displaying text. This paper reports on a research program that aims to find a new role for computers in supporting this process. One way to do this is by eliciting the semantic and conceptual content of text in an explicit representation, the semantic and conceptual structure. Providing appropriate means for visualizing such structures can facilitate non-linear navigation of text, which may prove to be a more effective and efficient way for acquiring knowledge from orthographic content. Based on this, we develop semantic structures as cognitive tools to support text-based reading on screen.

Keywords: semantic structure, semantic visualization, computer-based reading, cognitive tools

1 Introduction

Acquiring knowledge by reading text is an all-familiar cognitive operation. It is a linear process in which a user reads verbal content by scanning text in a sequential manner. It is an established fact that computer users prefer to read text from papers than off computer screens [Mills et al. 87; O'Hara and Sellen 97]. Hence, computers must take a more active role in supporting knowledge acquisition through reading rather than just displaying text.

Reading text on screen brings two issues. One issue is how to support the cognitive processes of reading. Schema Theory indicates two types of reading processes: bottom-up (decoding meanings from text) and top-down (using existing knowledge to predict and assimilate new knowledge) [Carrell and Eisterhold 88; Landry 02]. Both processes will be used in reading and need to be supported. Another issue is how to enable the interaction between readers and reading materials. The active role of computer-supported reading is to enhance psychophysical properties like readability, engagement, and responsiveness of the text on screen [Longhurst 03]. The 3Book, a 3D interactive visualization virtual book, has features to enhance reading, such as ScentIndex for conceptual organisation and spreading activation for automatic recommendation [Card et al. 04; Chi et al. 04; Woodruff 00].

Our approach is to use semantic technology to construct a new environment for reading on screen. In this paper, we present an explicit semantic representation in supporting the cognitive processes in knowledge acquisition by reading [Elsayed and Qiu 06]. The semantic representation takes a hierarchical form which is constructed by document mapping [Ou and Elsayed 06] and semantic condensation from the actual verbal content. Different to other semantic applications, such as Semantic Wikipedia [Völkel et al. 06] which embed hyperlinks within the text, our approach uses semantic visualisation technology [Chase et al. 98; Bederson et al. 96; Suh and Bederson 01] to facilitate non-linear navigation of text by establishing cross-window mappings.

This paper is organised as follows. The next section will be on the analysis of cognitive processes of reading. Then we will show our approach which includes the construction of the structures for semantic representation and the visualization for non-linear navigation. The last part will introduce the cognitive functions of the under-developing cognitive tools.

2 Cognitive process analysis for reading

Knowledge acquisition by reading involves several cognitive processes. Reading involves identification of letters, recognition of words, visual discrimination, predicting the forthcoming content and decoding the intended meaning of the message [Varzegar 93]. It is a ‘bottom-up’ process which is to sequentially receive and interpret the encoded information [Urquardt and Weir 98; Landry 02]. Reading is also a ‘top-down’ process, in which the reader makes contextual guessing to obtain contextual meanings [Carrell and Eisterhold 88]. Varzegar [Varzegar 03] asserts that reading is a process of extracting the contextual and pragmatic meanings which lead to a more global and holistic comprehension. Such process is called Conceptualization [Balmer 81]. The means the reader reads sentence by sentence and comprehends the individual sentences, but may fail to extract the meaning of the context because the pragmatic residues are not grasped [Levinson 83].

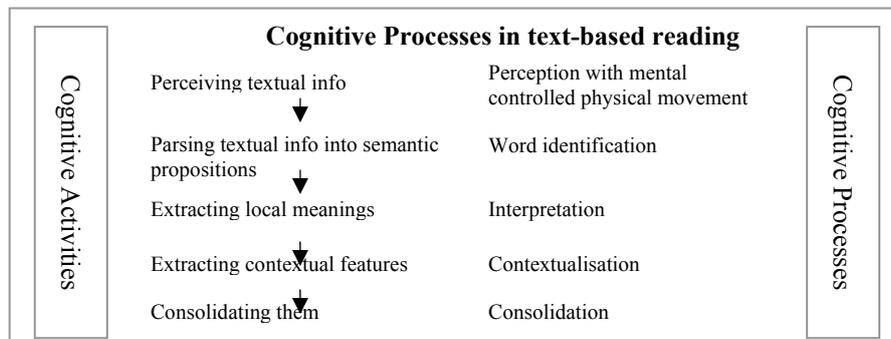


Figure 1: Cognitive process analysis for text-based reading

On the other hand, Royer [Royer and Sinatra 94] pointed that the reading information in the memory will decay quickly. It’s necessary to have a post-reading process to review and consolidate them associated with personal thoughts [McNeil

84]. In general, the process of reading can be broken down to several cognitive activities associated with individual cognitive processes. Figure 1 illustrates the cognitive activities and cognitive processes in reading.

The first step is to perceive the textual information, which relates to mental-controlled physical movement such as eye-movement. The second involves the syntactic processing for word identification. The third step has got a specific cognitive process, 'interpretation', which interprets the perceived textual content. The fourth step involves a high-level cognitive process, 'contextualization', which links 'interpreted' semantic information to contextual meanings in order to extract 'global meanings'. The consolidation process is to rehearse the 'acquired' information and organise them into 'schematic' structure.

Computer doesn't support those cognitive processes [Fig. 1]. In paper-based reading, the on-hand pens can be used to write hints and annotations, make marking or take notes on the paper directly. The on-hand tool can support above cognitive processes. But the computer-screen based reading is limited to get that kind of support.

3 Our Approach

In this research, we suggest to use semantic structures to support cognitive processes. By providing appropriate visualisation technology, the reading content would be navigated in a non-linear way.

3.1 Semantic Structures

Semantic structure, in cognitive science, is a type of knowledge representation to represent meaning. Although it may not be directly visible, it follows the generation of a syntactic representation by the human parser in reading. In an extreme view, except language study, the mental effort for physical perception and syntactic processing is redundant in reading. Only semantic representation might be useful for knowledge acquisition. However, the linearly constructed reading materials force us to follow the linearly reading style: perception, syntactic processing, and semantic-lexical processing.

Therefore, our approach aims to develop semantic structure out of the original text. That semantic structure is a hierarchical structure, which takes a triangular form where the bottom level represents the actual verbal content of the document and the top level provides its contextual representation of the main topic. Between these two extreme levels there would be a number of discrete representations in which the semantic content of the document is successively reduced, retaining only core semantic attributes until the contextual level is reached. To achieve this form of representation, the original text is transformed into elementary propositional units, which are successively analyzed and reduced using text summarization techniques. The resulting two-dimensional triangular representation can be expanded into a three-dimensional pyramid representation, in which the propositional units at each level are replaced by corresponding two-dimensional conceptual maps.

The pyramid representation can be generated by automatic text summarisation [Mani 01] and document mapping [Ou and Elsayed 06] techniques with the ontology of subject domain. At this moment, we use semi-automation

process to obtain the semantic structures. The structures are encoded by using RDF format. The next section will discuss how to use visualisation technique to support non-linear navigation.

3.2 Visualisation for non-linear navigation

The semantic hierarchical structure enables the navigation on the reading materials in a non-linear way. For example, the structure can provide the direct relations of relevant topics (semantic propositions) associated with the current reading sentence. The relevant topics can be in the local area (the same level propositions in the structure) or in the regional and global area (the higher level propositions in the structure). When they are selected in the structure, the associated sentence(s) in the actual verbal content will be highlighted for continuous reading. Furthermore, the relevant conceptual maps associated with the propositions in the structure are also available for the reader to view the conceptual structure and the distributed concepts (or keyword) in both original content and the structure. That will support a wider non-linear navigation. This non-linear navigation is similar to Semantic Zooming [Bederson et al. 96], which provides zooming in/out based on their semantic relations.

The key to successful non-linear navigation depends on the usability afforded by the interface that will relate the reader to the document and its semantic structure. The usability will no doubt be enhanced if supported by appropriate visualisation techniques [Chase et al. 98]. We are currently exploring an interface prototype based on four windows [Fig. 2]. Since these four windows provide four different perspectives of the content, they can support non-linear navigation by establishing cross-window mappings, allowing the identification of keyword distributions and propositional threads, in addition to supporting the completion of user-created conceptual maps that may be developed during the reading process.

4 Cognitive tools for computer-screen based reading

We are developing a prototype including a set of cognitive tools to implement above approach. The following graph [Fig. 2] shows the interface of the prototype:

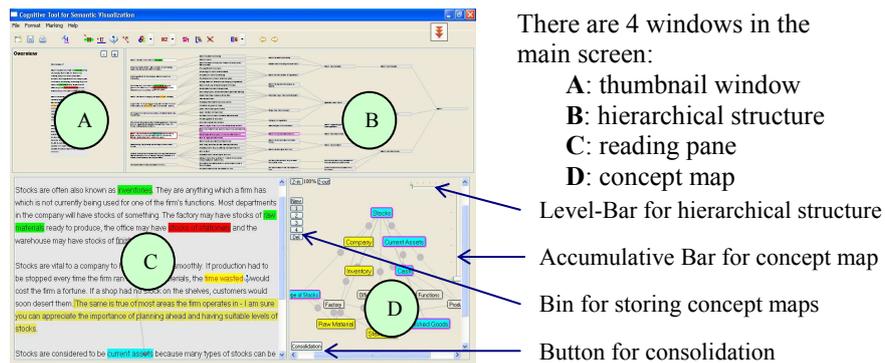
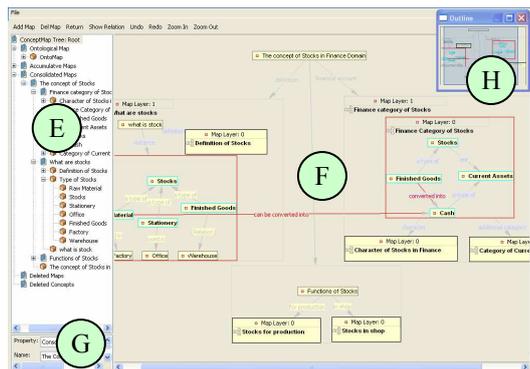


Figure 2: main screen of the cognitive tool

The main screen includes four windows introduced in Section 3.2. Thumbnail window displays the overview of the document. Reading pane shows the actual verbal content and supports eye-focus and marking. The third window shows the hierarchical structure, currently limited to a two-dimensional triangular form, and the last window displays the accumulative concept maps corresponding to the selected text in the structure and reading pane. In the last window, there are some special navigation functions: level-bar for selecting a level in hierarchical structure, accumulative bar for selecting the accumulative process of the concepts, and map bin takes managerial role for managing the accumulative maps.

The following graph [Fig. 3] shows the screen of the consolidation tool:



There are 4 windows in the workspace of consolidation screen:

- E: tree structure of hierarchical combined maps
- F: workspace for combining maps with semantic zooming function
- G: properties of concept maps or concepts
- H: thumbnail window of workspace

Figure 3: interface for consolidation to build combined maps

The consolidation tool provides 'editing' and 'zooming' functions to support organising the concepts into multi-layer maps. Each map is a cluster of some information and the embedded sub-maps can be zoomed in/out [see 'Consolidation mapping tool' below].

As we discussed in the Section 2, reading involves several cognitive processes. In this research, we have developed a set of cognitive tools to support those cognitive processes. The following table [Tab. 1] shows the use of cognitive tools for individual cognitive process in reading.

<i>Cognitive process</i>	<i>Cognitive tools for computer-screen based reading</i>	<i>Pen-paper based reading</i>
Perception	Eye-focus tool;	Pen-guided eye-movement;
Word identification	Marking tool	Free-marking
Interpretation	Marking tool; Word distribution; Semantic conceptual structure; Pyramid hierarchical structure; Non-linear navigation tool	Free-marking; Annotation
Contextualization	Pyramid hierarchical structure; Non-linear navigation tool	Annotation
Consolidation	Accumulative conceptual maps; Combination of conceptual maps; Non-linear navigation tool	Notes

Table 1: the use of cognitive tools for supporting cognitive processes in reading

The process and functions of above cognitive tools are described below:

- *Eye-focus and Marking tool* provides a ‘focal window’ to highlight the current reading sentence. The reader can move the window from one to another. It’s used to enhance the ‘perception’ ‘lexical identification’. The marking tool provides several marking functions to help the reader to identify the important, interested, or ambiguous information on the content. It’s used support to obtain the ‘interpretation’ from the neighbourhoods and relevant information. It also supports backward reading or review.
- *Keyword distribution tool* supports a quick search of the selected keyword within the document and semantic structure. This distribution or statistic information supplements the ‘interpretation’ to view the importance of the word and other possible meanings or usage in the document. It indirectly helps non-linear navigation.
- *Conceptual structure* explicitly represents the semantic relations between the concepts. When it’s visualised, it will help ‘interpretation’. When it supports user-creation, it supports ‘consolidation’ of knowledge.
- *Pyramid hierarchical structure* represents the multi-level summaries of the document. When it’s navigated, it will help to extract the global meanings (bottom-up) or explain the details (top-down).
- *Non-linear navigation tool* supports different cognitive processes: interpretation (navigation between reading pane, semantic structure and concept maps), contextualization (navigation between reading pane, semantic structure), and consolidation (navigation within the hierarchical form of combined conceptual maps and the accumulative map associated with original content).
- *Accumulative conceptual mapping tool* supports the reader to visualise the accumulative process of acquiring concepts. It will support the process of interpretation and initial consolidation.
- *Consolidation mapping tool* is used to consolidate the ‘acquired’ knowledge. It includes three main functions for the consolidation. The first is to support the reader to edit the maps which reflect the ‘acquired’ information. The second enables the user to combine the concepts and maps into higher layers. The third function is to support semantic zooming function which enables zooming in/out across the different layer maps. Those consolidated maps are the e-Notes which may help for the further reading. Furthermore, it can be used for articulating/externalizing what they have acquired from the text using their own language [Elsayed and Qiu 06].

In general, those cognitive tools are not isolated in reading. They are associated with each other. The readers can use them as a navigation tool to operate the document. Same as navigating on an electronic map, the readers are able to zoom in and zoom out using those cognitive tools, especially by the semantic structures.

5 Further research

This research is a part of research project in M3C (Machine-Mediated Multimodal Communication) Lab in the University of Bolton. The prototype is currently at the last stage of development. The next step of this research is to test and evaluate these cognitive tools in both technical and psychological level. The tool may support all kind of computer-screen readers and all kind of reading materials. At this moment, we are only testing based on adult readers and a domain of 'business/finance'. Our further research would be in three research areas:

- Automation of generating semantic structures: it will use natural language processing technologies, automatic text summarisation and document mapping, to generate semantic structures automatically.
- Cognitive support for articulation: it will use document mapping technology to map the user's articulation and compare them with existing conceptual structures. In fact, the user-created conceptual maps will be a base for the users to articulate the 'acquired' knowledge after reading.
- Pen-based learning: it's possible to use graphic tablet as a pen to operate the cognitive tools developed in this research.

6 Conclusion

Semantic technology will be widely used in next generation of web systems. Semantic Web technology as a fancy term will enable the web-based information more meaningful and navigable. By providing semantic navigation cognitive tools, we are hoping the users could easily access, search, browse, and manage the meaningful information online.

Acknowledgements

The M3C Lab is a Machine-Mediated Multimodal Communication research Laboratory in the University of Bolton. This research program is funded by M3C Lab (<http://data.bolton.ac.uk/staff/elsayed/m3c/m3c/index.html>).

Reference

- [Ballmer 81] Ballmer, T.: "Words, sentences, texts, and all that", *Text*, 1, 2 (1981), 163-189.
- [Bederson et al. 96] Bederson, B. B., Hollan, J. D., Perlin, K., Meyer, J., Bacon, D., Furnas, G.: "PAD++: A Zoomable Graphical Sketchpad for Exploring Alternate Interface Physics"; *Journal of Visual Languages and Computing*, 7, (1996), 3 – 31.
- [Card et al. 04] Card, S. K., Hong, L., Mackinlay, J. D. and Chi, E. H.: "Proceedings of the Human Factors in Computing Systems"; Conference (CHI2004) Conference Companion, Vienna, Austria (2004).
- [Carrell and Eisterhold 88] Carrell, P., & Eisterhold, J.: "Schema theory and ESL reading pedagogy"; In P. Carrell, J. Devine, & D. Eskey (Eds.) *Interactive approaches to second language reading*, (1988), 37-55. Cambridge: Cambridge University Press.

- [Chase et al. 98] Chase, P., et al: "Semantic Visualization"; ACL-COLING Workshop on Content Visualization and Intermedia Representations, (1998)
- [Chi et al. 04] Chi, Ed H., Hong, L., Heiser, J., Card, S. K.: "eBooks with Indexes that Reorganize Conceptually"; CHI 2004, ACM Conference on Human Factors in Software (2004). ACM Press.
- [Elsayed and Qiu 06] Elsayed A., Qiu, Y.Q.: "Cognitive support: a Machine-Mediated Communication perspective"; Advanced Technologies for Life-Long Learning, 6th IEEE International Conference on Advanced Learning Technologies (ICALT 2006), Kerkrade, Netherlands July 5 - 7, (2006)
- [Landry 02] Landry, K. L.: "Schemata In Second Language Reading"; *The Reading Matrix*, 2, 3 (2002)
- [Levinson 83] Levinson, S.: "Pragmatics"; Cambridge: Cambridge University Press (1983).
- [Longhurst 03] Longhurst, J.: "World History on the World Wide Web: A Student Satisfaction Survey and a Blinding Flash of the Obvious"; *The History Teacher*, 36, 3 (2003), 1.
- [McNeil 84] McNeil, J.D.: "Reading comprehension: New directions for classroom practice"; Glenview, IL: Scott Foresman (1984).
- [Mani 01] Mani, I.: "Automatic Summarization"; John Benjamins Publishing Company, Amsterdam, Philadelphia (2001).
- [Mills et al. 87] Mills, C.B. and Welon, L.J.: "Reading Text from Computer Screens"; *ACM Computer Surveys*, 19, 4 (1987), 329-331.
- [O'Hara and Sellen 97] O'Hara, K. and Sellen, A.: "A Comparison of Reading Paper and On-Line Documents"; *ACM CHI 97* (1997), 335-343.
- [Ou and Elsayed 06] Ou, W. and Elsayed, A.: "A knowledge-based approach for semantic extraction"; International Conference on Business Knowledge Management, Macao (2006).
- [Royer and Sinatra 94] Royer, J. M., & Sinatra, G. M.: "A cognitive theoretical approach to reading diagnostics"; *Educational Psychology Review*, 6, (1994), 81-114.
- [Suh and Bederson 01] Suh, B., Bederson, B.: "OZONE: A Zoomable Interface for Navigating Ontology"; *Proceedings of International Conference on Advanced Visual Interfaces (AVI 2002)*, ACM, Trento, Italy, (2002), 139-143, ACM Press.
- [Urquardt and Weir 98] Urquardt, S., & Weir, C.: "Reading in a second language: Process, product, and practice"; New York: Addison Wesley Longman Limited (1998).
- [Varzegar 93] Varzegar, M.: "The componential features of comprehension in reading"; Tabriz University Press (1993)
- [Varzegar 03] Varzegar, M.: "The features of the compensatory-interactive reading comprehension techniques"; NJTESOL/BE. Somerset, NJ (2003).
- [Völkel et al. 06] Völkel, M., Krötzsch, M., Vrandečić, D., Haller, H., and Studer, R.: "Semantic Wikipedia"; In *Proceedings of the 15th International Conference on World Wide Web*, Edinburgh, Scotland, WWW '06. ACM Press, New York, NY, (2006), 585-594.
- [Woodruff et al. 00] Woodruff, A., Gossweiler, R., Pitkow, J., Chi, E. H. and Card, S. K.: "Enhancing a Digital Book with a Reading Recommender"; CHI 2000, ACM Conference on Human Factors in Computing Systems (The Hague, The Netherlands), (2000), 153-160. New York: ACM Press.

Visual Tools Decipher Historic Artefacts Documentation

Iwona Dudek, Jean-Yves Blaise
(UMR CNRS/MCC 694 MAP, France
{idu, jyb}@map.archi.fr)

Abstract: Analysing and understanding the evolution of historic artefacts requires the cross-examination of indications ranging from specific pieces of data (remains of the edifice, archival materials, etc.), to generic pieces of knowledge (historical context, comparable cases, theory of architecture, etc.). This research is based on the premise that the artefact's acts as a media allowing the integration of the above-mentioned heterogeneous indications. Consequently, they may enable information visualisation and retrieval through 2D/3D dynamic graphics. In this contribution, we discuss four SVG-based graphic tools aiming at exploiting visually the relations between an artefact and the above-mentioned indications, *i.e.* its documentation.

Key Words: Architectural heritage, Ontology, Interfaces, Information visualisation, Spatio-temporal data, Document analysis and retrieval.

Categories: H.3.3, H.5.2, J.5, H.1, H.2.5, H.2.8, H.3.1

1 Introduction

Analysing and understanding the evolution of an historic artefact requires the cross-examination of various *clues*, ranging from specific pieces of data (remains of the edifice, archival materials, etc.), to generic pieces of knowledge (historical context, theory of architecture, etc.). Such clues, that we will call the artefact's documentation, may vary in type, relevance or spatial granularity (*i.e.* they may concern a whole town or an isolated architectural object). Therefore one single artefact may be described by numerous heterogeneous documents. In parallel, a single piece of documentation – a manuscript for example - may concern various artefacts and various spatial granularities. The analysis of the documentation helps to put in relation architectural elements at various scales (from architectonic details to edifices) with pieces of information, and eventually allows researchers to understand and represent the successive spatial configurations of artefacts. From then on, the artefact's representation can be used as a mean to retrieve/visualise the documentation. It should be stressed however that the artefact's representation acts as a metaphor: its role in the investigation process is *not narration but explanation* in the sense of [Tufte, 01].

In this contribution, we present four SVG-based (*cf.* [Geroimenko, 05]) graphic tools which exploit visually the relations between architectural and urban components of Krakow's urban fabric and their documentation. Four major questions were posed:

- Can we interface the documentation using the artefact's spatial distribution?
- Can we retrieve from a document its *architectural content* and visualise it?
- Can we visualise the reliability of the information extracted from documents and justify the artefact's representation by a "bibliographic" context?

- Finally, in the case of visual documents (a painting for example), how can we describe what we can guess about “where the creator “was standing” and what he was “looking at” while taking his sketches” and use this information in a map representation ?

The resulting developments we present should be understood as a way to reveal through visual means the analysis of artefacts’ documentation, *i.e.* an attempt to *amplify cognition* [Kienreich, 06] about the artefact’s bibliography.

2 Scientific background

As demonstrated in J.Bertin’s “graphic semiology” [Bertin, 98], attempts to exploit the spatial distribution of data sets raises numerous methodological questions about the efficiency and the readability of graphics. In response, we have in recent contributions [Dudek, 05], [Dudek, 05b] tried to take our inspiration from ideas and methods developed and used the field of *infovis*, while keeping in mind the particularity of our domain – an approach centred on the study of individual objects and their particularities. In this contribution we focus on one of these particularities: a massive yet highly questionable documentation.

The heterogeneous nature of documentation poses various problems to the analyst - different types of documents pose dissimilar problem (*i.e.* ambiguity of the textual descriptions, exactitude of artistic representations benefiting largely from *licencia artistica*, etc.). A document’s interpretation is drawn from its reading as well as from the experience, knowledge and intuition of the analysts. The result of this process is a hypothesis about artefact changes over time. In “historical sciences” however, one cannot use experiment in order to verify a hypothesis (an experiment in the past is not very realistic!). Only through intersubjective comparisons with the results of other scientists (*cf.* intersubjectivity [Bocheński, 88]) can the likelihood of a hypothesis be amplified. Furthermore, we propose to focus on visual means of comparison (access to documents and data, accompanied with critical analyses - analysis of a source/data, its architectural content, its context of creation, etc.) by switching, as often as possible, from ethnical languages to a visual language - more hard one [Lem, 96], more explicit and clear-cutting.

Which tools can we exploit in order to promote the above-mentioned ambition? Researchers dealing with the architectural heritage are offered an array of tools (GIS, laser scanning, virtual reality, etc.) and confronted to renewed methodological challenges (growing mass of documents, XML-based modelling languages, DC Standard, etc.). But in practice, do these opportunities really help understanding architecture in its historical/geographical/cultural context, or do they only reduce the artefacts, here to polygons, 3D clouds of points and meshes [De Luca, 05], there to tags, and all-purpose typologies and links [Pajares, 00]? Most notably because of the narrowness of technological solutions and lack of a sustainable research methodology, clues in the understanding of architectural changes over time remain scattered pieces of information. Consequently, the architectural heritage regularly appears a playground for other disciplines, where architecture appears by way of illustration of a tool or a technology. The tools are not the heart of heritage sites, their heart is a history, partly written in architectural shapes. Beyond the shapes there is also

information, and the bibliographic justification of a spatial configuration of an artefact is for any scientist a very important aspect (see [Alkhoven, 93]) - providing the means of visual lecture of the relation document(s)-artefact(s) is therefore a priority.

3 Method and hypotheses

The idea of document retrieval and visualisation supported by assessment of its architectural content is based on two simple observations:

- an artefact's documentation is an information to visualise;
- this information can be found behind an artefact's shape.

Briefly speaking, our objective is to exploit the relation document(s)-artefact(s) in order to produce a *display of bibliographic evidence* [Tuft, 01]. However, this goal can be achieved only if the architectural content of a source has been formerly isolated and described. By "*architectural content of a historical source*" we understand the ensemble of information concerning architectural (or urban) elements one can derive from a source (*i.e.* a list of represented objects, date(s) at which they are represented, etc.). Visual sources theoretically do not present any particular difficulties in the determination of a visible objects (ex. *the Royal Castle, St. Peters Cathedral*). Dating may pose more problems, especially in the case of old artistic representations (ex. *the end of XVIth century?*). Textual sources pose even more problems, (*poli-interpretation, vagueness*), natural for the ethnical languages.

It has to be stressed that in any serious investigation of historic artefacts, the analysis of documentation is carried out in depth. Consequently, our approach does not introduce more documentation analysis, but only proposes a renewed structuring and memorisation of the documentation analysis. This approach is implemented on a case study - architectural and urban analyses of Krakow's medieval - on which we have several years of research with notably as results:

- an architectural ontology (implemented as a set of classes in the sense of OOP. Various evolutions of an artefact are represented as a chain of independent instances, allowing the system to handle morphological and documentary changes through time;
- VIA artefacts' database (RDBMS). It stores for each artefact a set of descriptive criteria - typological specificity, dating, alternative denomination, etc., with a special focus on a level of certainty of information (*To which extent can we be confident that this artefact was created before year 1345? Who said so?*). It contains at this stage 817 evolutions of 335 object;
- SOL bibliographic, iconographic and cartographic database (RDBMS), (stores data about documentation - editorial details, physical format, content and context description, etc.), 761 sources have been described;
- XML/XSLT, SVG (2D), VRML (3D) outputs (graphics produced in real-time represent the artefact at any time of its evolution: they implement J.Bertin's view : graphics as "an answer to a question").

Our aim is to perform document retrieval and visualisation by assessing the architectural content of a document. Using M. Eibl [Eibl, 02] definition, the document(s)-artefact(s) relation can be seen as enabling basic Boolean retrieval (yes/no bi-directional relation of the document to the artefact). It should also be

considered as enabling vague retrieval since it allows users to expand searches either on ontology similarity, on period expansion, on authorship (etc.). The following section introduces examples of these mechanisms, with a focus on recent visual displays we have developed within the above-mentioned research context.

4 Case study and implementation

4.1 Visual result of basic query mechanisms

As mentioned in section 3, artefacts are represented as a time-chain of independent instances, each of it having its own period of relevance, its own morphology, and its own justification. Basic queries visualise our state of knowledge for one time slot, and provide links to the artefact's related sources and to their analysis (SOL: the sources themselves, VIA: what they tell us of the artefact). The query uses a criterion of time (a period of time) and ontology (a type of artefact) to select artefacts. Colour codes are used to visualise ontological differences and certainty differences.

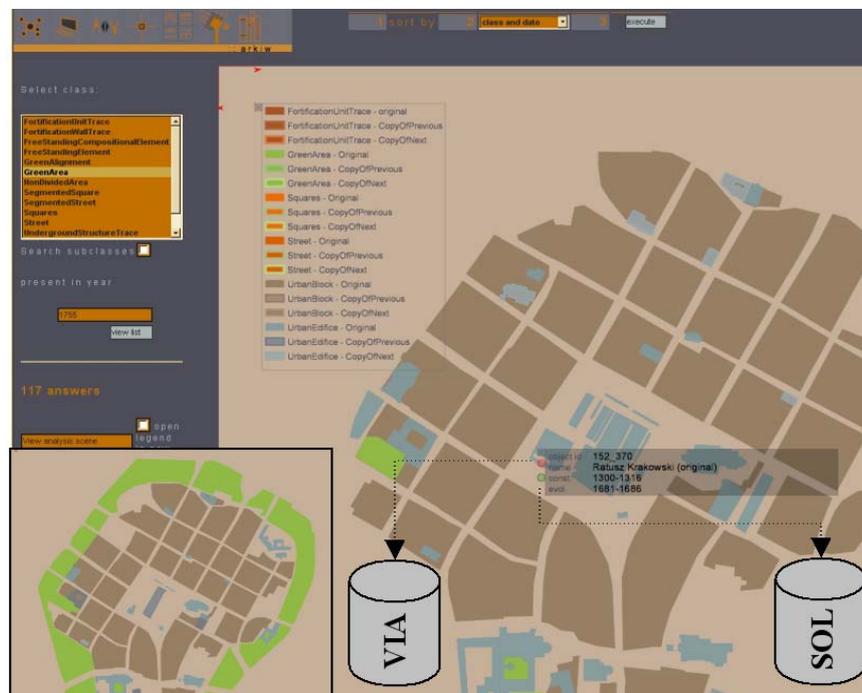


Figure 1: The XML/XSLT/SVG basic query interface (parameters date 1755 with 3 classes of objects, bottom left, in black outlined box, same query for period 1890)

4.2 Visualising the architectural content of sources

In section 4.1 the search is carried out on properties of artefacts and the resulting representation linked to the documentation. In this second case the document(s)-

artefact(s) relation is exploited the other way round. The search is carried out on properties of the documentation, and outputs an SVG “architectural content visualisation” which represents the architectural content of the selected source(s).

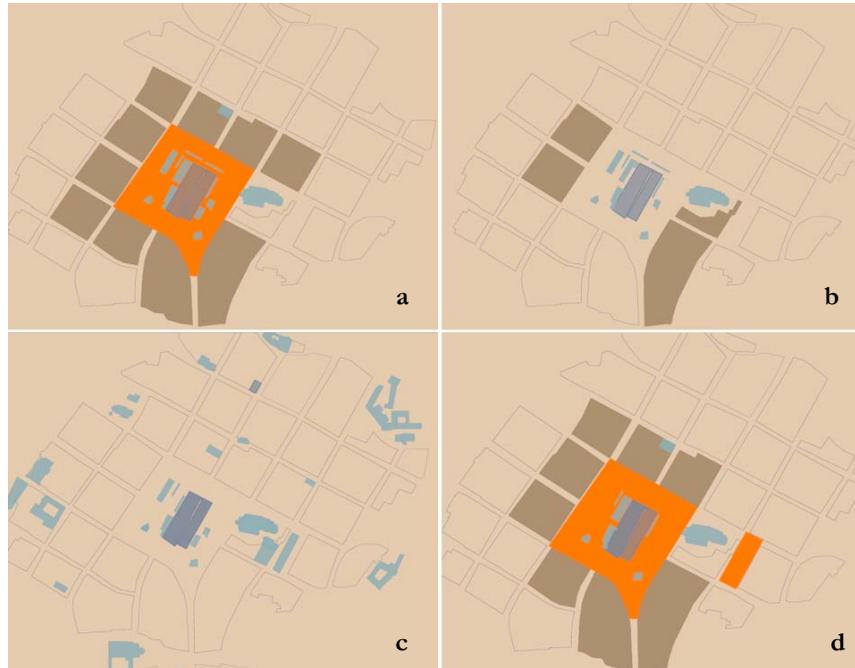


Figure 2: (a,b) - queries on a single source, (c,d) - queries on authors. The typical visual patterns for the content of: (a) - a historical map with a wide ontological coverage (variety of types of artefacts) and for (b) - a painting, (c) an architect's publications (narrow ontological coverage - only monumental buildings, but throughout the city) and (d) a historian investigations (encompassing ontological differences –urban blocs, monumental buildings, squares, etc.)

4.3 Cross examining instance justification and documentation

The documentation's analysis helps researches justifying choices they make when describing how an artefact evolved throughout its history. What we call *justification* is implemented as lexical scales (in the sense of [Zeitouni, 00]) stating "how likely" the information we have is. It is most important to find visuals means to represent this justification. Furthermore, this justification varies with object changes : some periods of evolution are better known (and documented) than others. As an answer we have (see figure 3) developed a three-component visual disposal associated with each artefact (selection of artefacts as in section 4.1):

- an horizontal timeline puts in parallel the chronology of evolution of the artefact, and the chronology of its related sources (in order to point out documentation gaps);
- a vertical line divides (left) total amount of documents for the whole history of the artefact and (right) amount of documents for the time slot under scrutiny;

- four series of six arcs mark the justification level of the artefact (selected value: red).

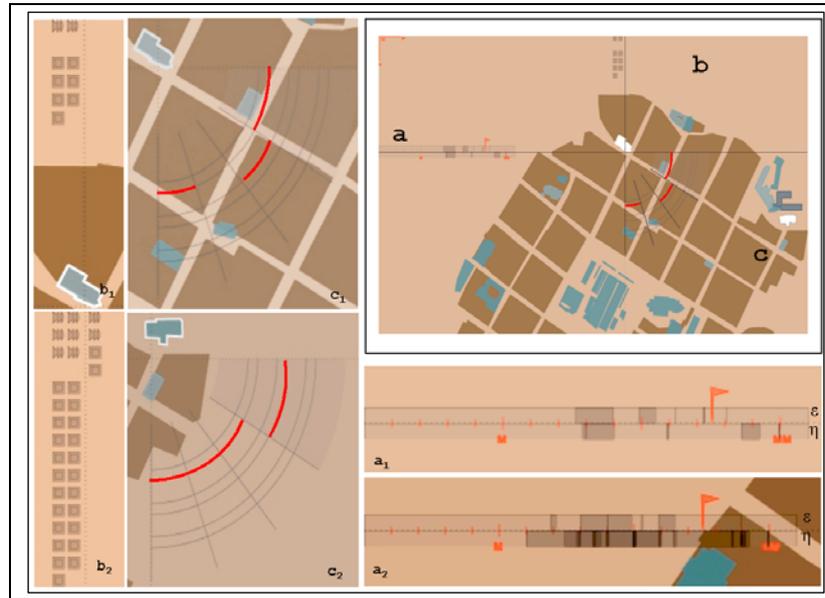


Figure 3: Top right, composition of the three-component visual disposal (period under scrutiny 1755, disposal over church Sw. Marka); churches Sw. Marka (left, 1) and Sw. Krzyza (right, 2) marked in white; **a** - visual comparison of the artefact's chronology of evolution (ϵ -line) and of the chronology of sources (η -line). For church Sw. Marka (a_1) documents do not cover each evolution of the artefact. For church Sw. Krzyza (a_2) each evolution has corresponding documentation. **b** - overall number of sources available vs. number of sources available for the period of the query (underlines moments in the artefact's history than would need further documentary investigation). (b_1) Sw. Marka (poorly documented, no document for period under scrutiny), (b_2) Sw. Krzyza - better documented. **c** - justifiers - visual display. Each sector corresponds to a certainty assessment (period, morphology, structure, function). In each sector an arc corresponds to a value in the lexical scale, the closest to the centre being the "most certain" (active values marked in red). No underlining means absence of data.

4.4 Current work

When the amount of documents handled for a given artefact is important, or when a document is related to many artefacts, we would need to further filter the documentation. The idea we are currently working on would be to analyse sources in order to know which "part" of the artefact they precisely mention (ex. northern façade). The basic parameters are the viewer's estimated position and an estimated target point. The first may be described by a point (x,y,z) (cf. Fig 4a). This parameter requires a coefficient of precision (visualised for example by a cloud-of-uncertainty

(cf. Fig 4b)). The definition of a z co-ordinate may require equally a special treatment (cartography for example) (cf. Fig 4cd).

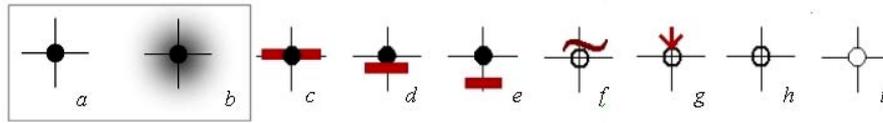


Figure 4: Proposed symbols for viewer's estimated position - the origin (a), the same symbol with a coefficient of precision represented as a cloud-of-uncertainty (b). A cross defines x,y co-ordinates, a circle represents the status of x,y co-ordinates (defined/undefined), a red symbol expresses the z co-ordinate (lexical scale).

The estimated target point can be represented – according to the data's precision by: a *point*, a *bearing* (horizontal angular direction of view), or by a *geographical bearing* defining the orientation. For some visual sources it may be useful to provide a *vertical bearing* – vertical angle of view (cf. Fig 5).

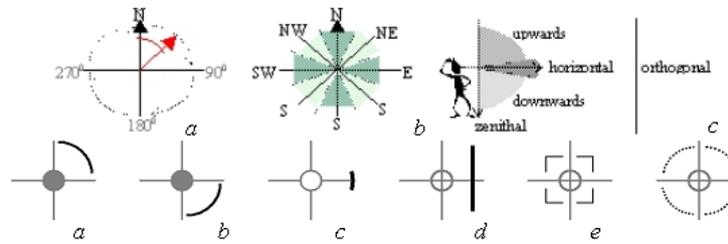


Figure 5: Top, Bearing (a), geographical bearing(b), vertical bearing(c). Bottom, Symbols of vertical bearing are represented with a symbol of viewer's position (a) upwards, (b) downwards, (c) horizontal, (d)orthogonal, (e) zenithal, (f) undefined

A visual representation of image's ratio and a source type are also the important indicators for a user (cf. Fig. 6). Two examples of how sources are projected on contextual maps are given below (cf. Fig. 6, right).

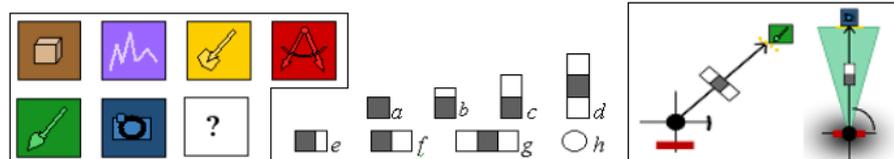


Figure 6: Left, source type - identified by a colour and pictogram, examples: physical model, diagram, archaeological drawing, architectural drawing, artist representation, photograph, undefined. Centre, image's ratio : (a) square a/a , (b) vertical postcard $a/a < b < 2a$, (c) vertical $a/2a$, (d) vertical extensive $a/b > 2a$, (e) horizontal postcard $a < b < 2a/a$, (f) cinema $2a/a$, (g) panorama $b > 2a/a$, (h) undefined. Right- first example: a panoramic artistic representation, from an elevated point, horizontal angle of view, fuzzy target point; second example: a vertical photograph, low point of view, an upwards angle of view, fuzzy view direction, fuzzy origin point.

5 Limits and conclusions

We present ideas to retain through visual means the artefact's documentation analysis, and provide graphic tools to reveal it. Our implementation identifies three steps:

- analyse the architectural content of a document;
- cumulative analyses a spatial configuration for the artefact;
- artefact's spatial representation in order to exploit the document(s)-artefact(s) relations.

Accordingly, graphics go beyond mimicking how an artefact is or has been, they deliver an information, and therefore can be considered as a scientific visualisation disposal in the sense of [Spence, 01]. At this point in time, numerous obstacles remain to be solved (ex. the *spatial location* of textual sources) but the gain of insight one gets about the artefacts and their documentation appears promising. Moreover, the document(s)-artefact(s) relation could be exploited beyond the field of architecture and probably find applications in the field of geospatial studies, with as a major constraint the necessity of a long-term cumulative approach in knowledge gathering.

References

- [Alkhoven, 93] Alkhoven P.: "The changing image of the city"; PhD University of Utrecht
- [Bocheński, 54] Bocheński, J.M: "Die zeitgenössischen Denkmethoden", Bern 1954
- [Bertin, 98] Bertin, J: "Sémiologie graphique", EHESS, *sine loco* 1967/1998
- [De Luca, 05] De Luca L, et al.: "Reverse engineering of architectural buildings based on a hybrid modeling approach", *Computers & Graphics*. Vol. 30, 2006
- [Dudek, 05] Dudek I., et al: "Using abstraction levels in the visual exploitation of a knowledge acquisition process", Proc. I-KNOW 05, Graz, JUCS, pp. 543-552
- [Dudek, 05b] Dudek I., et al: "From artefact representation to information visualisation: genesis of informative modelling", Proc. SmartGraphics, LNCS Volume 3638/2005, 2005
- [Eibl, 02] Eibl, M.: "DEVID: a media design and software ergonomics integrating visualisation for document retrieval", Information visualisation vol.1 Number 2, Palgrave 2002
- [Geroimenko, 05] Geroimenko, V., Chen C (Eds): "Visualizing information using SVG and X3D" Springer, *sine loco* 2005
- [Kienreich, 06] Kienreich, W.: "Information and Knowledge Visualisation: an oblique view", MIA Journal Vol0 Numb.1, 2006
- [Lem, 96] Lem, S.:, "Tajemnica chinskiego pokoju", Universitas, Kraków 1996
- [Pajares 00] Pajares Tosca, S: "A pragmatics of links", Journal of Digital Information, vol. 1/6, 2000
- [Spence, 01] Spence R.: "Information vizualisation", Addison Wesley, Harlow 2001
- [Tufte 01] Tufte, E.R.: "Envisioning information", Graphics Press, *sine loco* 1990-2001
- [Zeitouni, 00] Zeitouni, K. et al.: "Join indices as a tool for spatial data mining", Int. Workshop on Temporal, Spatial and Spatio-Temporal Data Mining, Springer, pp.102-114, 2000

Visualization of Rule Bases - The Overall Structure ¹

Valentin Zacharias

(FZI Forschungszentrum Informatik, Germany
zacharias@fzi.de)

Abstract: In this paper we describe novel ideas and their prototypical implementation for the visualization of rule bases. In the creation of the visualization our approach considers not only the structure of a rule base but also records of its usage. We also discuss methods to address the challenges for visualization tools posed by rule bases that are large, created with high level knowledge acquisition tools or that contain low level rules that should remain hidden from the users.

Key Words: visualization, rule bases, knowledge visualization, visualization of knowledge structures, f-logic

Category: H.5.2, H.5

1 Introduction

The visualization of entire rule bases, independent of the answer to any particular query, is a problem that has so far been largely ignored by the research community. The goals for such a visualization are the same as for UML class diagrams and other overview representations of programs: to aid teaching, programming, debugging, validation and maintenance by facilitating a better understanding of a computer program by the programmers. The importance of such representations cannot easily be overstated and is probably higher than in object oriented programming; In object oriented programming the overall structure is explicitly created by the programmer with links between objects, the inheritance hierarchy and method calls; in rule bases the interaction between rules is only decided by the inference engine and usually never shown to the user. An overview representation of the entire rule bases is necessary to make this hidden structure visible. Before suitable visualizations for rule bases can be created, however, we need a definition of the *overall structure of the rule base* - doing this with respect to usage data is the main contribution of this paper.

This paper starts with a definition of rule bases. It then goes on to define the overall structure of a rule base, first the static structure and then the dynamic structure as a combination of all proof trees from uses of the rule base. In the end we shortly sketch some advanced problems, such as hiding language axioms or showing rule bases at different levels of abstraction.

¹ An extended version of this paper is available from <http://vzach.de/papers/visualizationLong.pdf>

2 Terminology - Rule Bases

A rule base is a finite set of rules. The rules are of the form:

$$p(A) \leftarrow q(B), \text{not } r(C, d)$$

The part to the left of the \leftarrow symbol is the *goal* or *head* of the rule, the part to the right the *body* or *subgoals*. The head of the rule consists of one *atom*, the body of a conjunction of *literals* (atoms that may be negated). The arguments of the predicates that make up head and body of the rule are called *terms*. Variables and constants as well as function symbols with terms as arguments are terms. A term without arguments is called a *ground term*. Variables are written with uppercase letters, ground terms with lowercase letters.

The meaning of a rule can be intuitively understood as *the statement in the head is true if the body can be proven* or even shorter as *if body then head*.

3 The static structure of rule bases - the rule graph

The core concept for the static structure of rule bases is the *depends-on* connection between rules. Such a link indicates that two rules seem to be able to work together. Consider the following rule base:

$$\text{father}(A) \leftarrow \text{male}(A), \text{parent}(A, B) \quad (1)$$

$$\text{parent}(X, Y) \leftarrow \text{child}(Y, X) \quad (2)$$

$$\text{employer}(X) \leftarrow \text{owns}(X, C), \text{employed_at}(A, C) \quad (3)$$

It can be seen that, with the right facts, rule (1) could work with the result of rule (2). The following fact base consisting of only two facts should serve as an example:

$$\text{male}(\text{mike}), \text{child}(\text{michelle}, \text{mike})$$

With these facts rule (2) could deduce that $\text{parent}(\text{mike}, \text{michelle})$ from $\text{child}(\text{michelle}, \text{mike})$ and rule (1) could then deduce that $\text{father}(\text{mike})$. In such a case we say that rule (1) *depends-on* rule (2). We say depends-on because the rule (1) could not have made the inference that Mike is a father without rule (2) being present; for this conclusion it depended on the other rule. For the depends-on connection we only consider the cases where a rule works directly on the conclusion of another rule, not the cases where there are intermediary rules. The depends-on connections are calculated on the rules without consideration for the actual facts that are available: a depends-on connection exists independently of the facts. A depends-on connection between a rule A and a rule B states that there exists a set of facts such that an inference of rule A directly depends on rule B being present.

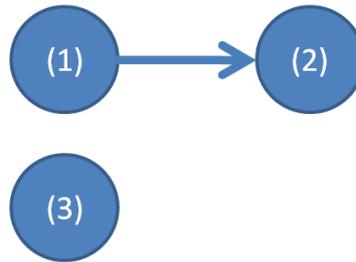


Figure 1: *The static structure for the example rule base in section 3.*

The rule (3) on the other hand, has no depends-on connection to any of the other rules - there exists no set of facts such that rule 3 could directly work on data inferred by one of the other two rules. Also note that depends on connections are not symmetric: rule (1) depends-on (2) but not the other way around.

A simple way to calculate an approximation, of the depends-on connection is the following: a depends-on connections exists from rule A to rule B, iff rule A has a body atom that can be unified with the head atom of B.

The static structure of the rule base then, is a directed graph $G_R = (R, A)$ with

- A set of vertices R , there is one vertice for each rule in the rule base.
- A set of ordered pairs A of vertices called arcs or arrows. An arc $e = (x, y)$ is considered to mean that the rule represented by the vertice x depends on the rule represented by y. Such an arc exists for each depends-on link between rules.

The rule graph for the simple rule base example in this section is shown in figure 1. It quickly conveys information such that $\{1,2\}$ and $\{3\}$ are independent parts of the rule base, that a deletion (2) will affect (1) and that someone interested in understanding rule (1) should also look at (2). For a rule base consisting only of three rules such a diagram is obviously not needed, but it can be seen how this kind of information can aid the creation, debugging, maintenance and reuse of larger rule bases.

4 Dynamic Structure

The static structure defined in the previous section represents the potential of rules to interact, without making assumptions/using information about the queries and facts a rule base is used with. The dynamic structure of a rule base complements this by combining the proofrees from all known queries to the rule

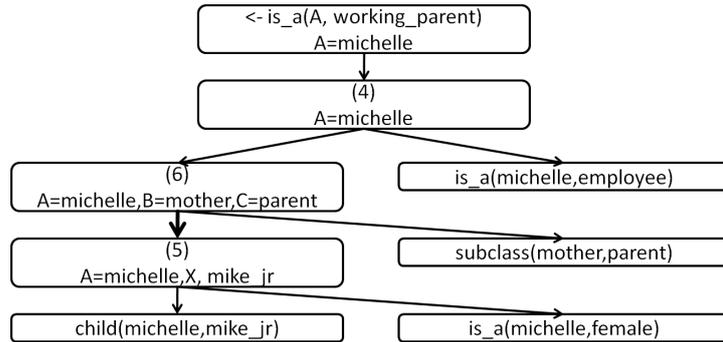


Figure 2: A proof tree for the example in section 4.

base into an overview picture. The resulting picture shows which rules and rule connections actually matter in the use of a rule base. When used to visualize the proof trees from the tests for a rule base it also gives a picture of test coverage.

Consider the following rule base. This rule base makes heavy use of a type hierarchy encoded in special predicates - a structure typical for rule bases created in F-logic [6].

$$is_a(A, working_parent) \leftarrow is_a(A, parent), is_a(A, employee) \quad (4)$$

$$is_a(A, mother) \leftarrow is_a(A, female), child(X, A) \quad (5)$$

$$is_a(A, C) \leftarrow is_a(A, B), subclass(B, C) \quad (6)$$

The static structure of this rule base is confusing (see figure 3) as almost everything depends on everything else; looking at the rule interactions in the actual use of the rule base can help here.

The structure that represents the interaction of rules and facts in the inferring of a result is called a proof tree. The root node of a proof tree is always the query, its variables bound to the result that has been returned. The children of this node are the rules that were directly needed to prove the query. The children of these rules are again the rules needed to prove them. Leafs of the proof tree are formed by the facts in the knowledge base. A bit more formally we can say that a proof tree is a directed graph $G_P = (V, A)$ with

- A set of vertices V . Each vertice can be understood as the application of one rule during the inference process. There is also (partial²) function $f_r : V \rightarrow R$ that gives the rule that was applied for a vertice in the proof tree.

² The proof tree also contains nodes that represent the “firing” of the query and the usage of facts

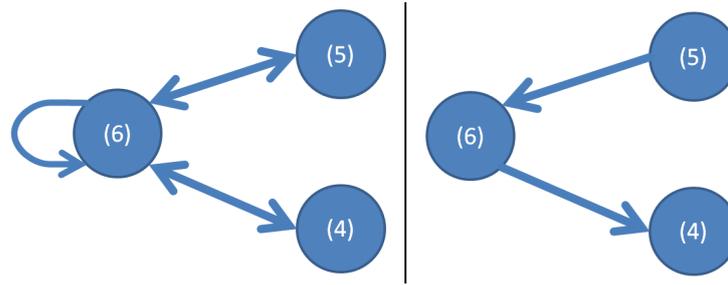


Figure 3: *The static structure (to the left) and the dynamic structure (right) of the example in section 4. In this example the dynamic structure is very similar to a proof tree - because the usage data consists only of this one proof tree; this is not the case generally.*

- A set of arcs A . An arc $e = (x, y)$ is considered to mean that the application of a rule represented by vertex x depended on the rule firing represented by y .

An example proof tree is created by the query $is_a(A, working_parent)$ posed against the rule base above and the following facts:

```
subclass(mother, parent), is_a(michelle, female)
is_a(michelle, employee), child(michelle, mike_jr)
```

A graphical representation of this proof tree is shown in figure 2. It shows the query on top with the variable A bound to michelle. This result of the query directly depended on the firing of rule (4) that in turn depended on rule (6) and the fact $is_a(michelle, employee)$ etc. Proof trees are well known tools to represent the inference process that lead to a particular result. Our novel approach however, combines all known proof trees for a rule base to create an overview representation of the entire rule base, independent of any particular query or result. To do this we define usage data U as the multiset of all known proof trees for a rule base: $U = \{G_{P,1}, G_{P,2}, \dots, G_{P,i}\}$ and the function f_r defined earlier. The proof trees could be created by tests of the rule base or be collected during the actual use of the system. We can then add a weight to each arc/rule in the rule graph (see section 3) based on how often this arc/rule appears in the proof trees in the usage data. Graphical representations of the rule graph can then hide unused depends-on links and highlight important rules and links.

The dynamic structure of the example in this section is shown in figure 4. Compared to the static structure it gives a much better picture of the interactions between these rules as they would happen for real world data.

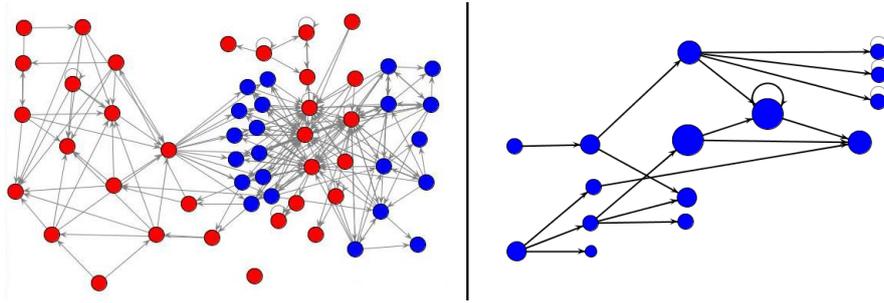


Figure 4: *Visualizations of a small working rule base. The left picture shows the unaltered static structure, the right picture shows the structure after rule nodes have been scaled based on their use, low level axioms have been hidden and some rules representing high level entities have been joined.*

Displaying the dynamic structure in this way is not the only possible way to use the usage data for a rule base. The following section will shortly discuss how it can also aid transformations of the rule graph to hide certain rules or to display the rule graph at different levels of abstraction.

5 Hiding and Joining Rules

The techniques presented so far work well enough for showing the structure of toy rule base. Even small real life rule bases, however, pose additional challenges: they are too large to show every rule, very general rules/axiom like rules (similar to rule (6) in the previous section) confuse the picture and sometimes graphical editors are used that create multiple rules for one high level entity known to the user. For instance figure 4 shows the static structure of a small but functional rule base - the structure is much too confusing to be of any use.

We selectively hide and join rules to deal with these problems. Figure 4 shows how this techniques work together. The right side shows the same rule base as the left side after usage data has been used to scale the rules nodes, axioms have been hidden and some rule sets (reflecting rules automatically created from one high level entity) have been joined. These transformations have uncovered the relatively simple, layered and ordered structure of this diagnostic rule base.

5.1 Hiding Rules

For certain very general, axiom like rules it makes sense to completely hide them from the view of the user. This is particular important for F-logic rule bases that

add a number of axioms to the rule base that are transparent to the user. Again the usage data is used to ensure that only relevant links are displayed.

Usage data in the form of proofrees can help to better deal with this problem because it contains *paths* through the rule base, not only arcs between two rules. When hiding rules this enables to retain only the actual paths through the hidden node(s). A detailed description of this can be found in the extended version of this paper available from <http://vzach.de/papers/visualizationLong.pdf>.

5.2 Joining Rules

There are a number of cases where it makes sense to join a number of rules for display purposes:

- Some engineering environments create more than one rule for a high level entity edited by the user. Here the visualization should only shows the high level entity without losing information.
- In a large rule base there is often some hierarchical structure on top of the rules (like rule packages or different files defining rules). The system can join all rules in each package and thereby give a high level view of the rule base, allowing to *zoom in* at one package, to show the rules in a package.
- In large rule bases without auxiliary structure clustering algorithms could be used to identify and the jointly display rule clusters, again allowing to expand the clusters.

Joining of rules is done by replacing a number of nodes in the rule graph with one new node. The detailed algorithm for this has to be omitted for brevity but can also be found in the extended version of this paper.

6 Related Work

To our best knowledge no approach that uses usage data in the form of proofrees for the visualization of rule bases exists. In fact we are not aware of any approach that uses the runtime rule interactions to create visualizations of entire rule bases.

There is a large number of approaches that visualize the inference process that lead to a single result (e.g. [2]) and some that show the static structure of rule bases [7, 4] but none that uses runtime rule interactions to create visualizations of entire rule bases. The approaches that show the static structure of rule bases also do not consider the challenges posed by large rule bases, high level editors and the hiding of rules.

Further visualizations of rules exist within the data mining community [1, 3, 5, 8, 9]. These approaches visualize association rules in order to facilitate the

analysis of these rules and the extraction of the most important information from them. They also face the problem of a large set of rules from which a few interesting ones need to be selected. These approaches, however, are mostly concerned with the problems posed by the statistical nature of association rules. The data used for the creation of the visualization, and its goal is very different.

7 Conclusion

A visualization of an entire rule base is an important tool for the creation, maintenance and reuse of rule bases. A rule base can be visualized based on its static structure or based on the actual rule interactions that happen during queries to the rule base. The joining and hiding of rules can be used to deal with the challenges posed by large rule bases and those that contain low level rules that should remain hidden from the user's view. The data about the actual rule interaction is necessary to join and hide rules while keeping the overall picture intact.

Acknowledgements

The author thanks Imen Borgi for doing a large part of the implementation reported on in this paper.

References

1. D.Bruzzese and P.Buono. Combining visual techniques for association rules exploration. In *Proceedings of the working conference on advanced visual interfaces*, pages 381 – 384, Gallipoli, Italy, 2004.
2. M. Eisenstadt and M. Brayshaw. The transparent prolog machine (tpm): an execution model and graphical debugger for logic programming. *Journal of Logic Programming*, 5(4):277–342, 1988.
3. Usama Fayyad, Georges Grinstein, and Andreas Wierse. *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann, 2001.
4. C.A.McK. Grant. *Software Visualization in Prolog*. PhD thesis, University of Zurich, 2001.
5. Jiawei Han and Micheline Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 1st edition, 2000.
6. M. Kifer, G. Lausen, and J. Wu. Logical foundations of object-oriented and frame-based languages. *Journal of the ACM*, 42(4):741–843, 1995.
7. J.W. Lewis. An effective graphics user interface for rules and inference mechanisms. In *Conference on Human Factors in Computing Systems Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 139 – 143, Boston, Massachusetts, United States, 1983.
8. M.Burch, S.Diehl, and P.Weiger. Visual data mining in software archives. In *Proceedings of the 2005 ACM symposium on Software visualization*, pages 37 – 46, 2005.
9. M.Kreuseler and H.Schumann. A flexible approach for visual data mining. In *IEEE Transactions on Visualization and Computer Graphics*, volume 8, pages 39 – 51, Germany, 2002.

A Semantic-based Search Engine for Professional Knowledge

Simona Colucci
(Politecnico di Bari, Bari, Italy
s.colucci@poliba.it)

Tommaso Di Noia
(Politecnico di Bari, Bari, Italy
t.dinoia@poliba.it)

Eugenio Di Sciascio
(Politecnico di Bari, Bari, Italy
disciascio@poliba.it)

Francesco M. Donini,
(Università della Tuscia Viterbo, Italy
donini@unitus.it)

Azzurra Ragone
(Politecnico di Bari, Bari, Italy
a.ragone@poliba.it)

Michelantonio Trizio
(Politecnico di Bari, Bari, Italy
mikelantonio.trizio@gmail.com)

Abstract: The search for professional knowledge is affected by an endemic ambiguity in the definition of required and provided competences. Ontologies represent a mean for disambiguation, by providing a shared vocabulary for job market knowledge domain. In this paper we propose an ontology based search engine for curricula, which exploits the semantic annotation of available curricula to rank them with respect to a knowledge request and implements novel technologies for dynamic and interactive web applications development.

Key Words: Skill management, Semantic Search, Knowledge Representation

Category: I.2.4, K.6.1

1 Introduction

The main distinguishing element of knowledge with respect to any other kind of good is in its intangibility, which makes subjective and ambiguous all of transactions in job market. A step toward disambiguation consists in trying to share as

much as possible the vocabulary used to describe required and provided knowledge. The use of ontologies [4, 3], nowadays widely employed in knowledge representation, may realize such a sharing objective. The technologies developed for the Semantic Web initiative [2] make knowledge availability independent of the physical location of the knowledge provider. In the introduced scenario the deployed resources of interest contain curricula vitae: by annotating curricula vitae according to a given ontology, the disambiguation objective may be achieved. We present a semantic-based search engine for curricula vitae, exploiting Semantic Web technologies for curricula annotation according to the vocabulary provided by a shared ontology. The search process calls for reasoning tasks on formalized knowledge, which we solve by exploiting formalism and reasoning services provided by Description Logics(DL) [1].

2 The Search Engine

The decentralized architecture of the proposed search engine is sketched in Figure 1. The engine searches for a CV answering the knowledge request among the

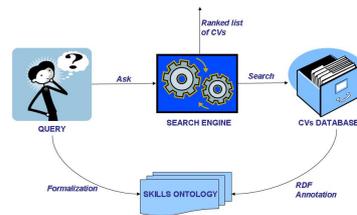


Figure 1: Search engine architecture

ones addressed by the database. The query, represented as an OWL-DL concept [7], is answered by a retrieval process exploiting RACERPRO [5] to rank CVs w.r.t. the query and taking the semantics of each CV into account. CVs semantics is conveyed through their annotation w.r.t. the given ontology in a RDF file, whose URL is stored in the CVs database together with the URL of the web page containing the CV itself. The decentralized architecture allows such resources to be not necessarily resident on the same server as the search engine, given their accessibility through URLs stored in the Database.

In order to sort CVs on the basis of their semantic similarity to required task, we propose here a ranking function, intuitively incrementing CV scores at each

occurrence of an owned skill relevant for the contextual query. More formally, consider the knowledge query Q formalized w.r.t. an ontology \mathcal{T} as conjunction of concepts C_i , representing the required skills, that is to say $Q = \sqcap C_i$ with $i = 1, \dots, n$. Consider also the set $D = CV_j$ with $j = 1 \dots m$, including CVs annotations according to \mathcal{T} . The scores computation mechanism is based on the algorithm detailed in the following:

```

Input : ontology  $\mathcal{T}$ ,  $CV_j$  with  $j = 1 \dots m$  and  $Q = \sqcap C_i$  with  $i = 1, \dots, n$  concepts in DL,
satisfiable w.r.t.  $\mathcal{T}$ 
Output:  $V = (CV_j, score_j)$ : score vector
 $i = 0$ ;
foreach  $i \leq n$  do
   $j = 0$ ;
  foreach  $j \leq m$  do
    if  $CV_j \sqsubseteq C_i$  then
       $score_j = score_j + 1$ ;
    end
     $j = j + 1$ ;
  end
   $i = i + 1$ ;
end
return  $V = (CV_j, score_j)$ ;

```

For each element C_i of concept in the knowledge request Q , RACERPRO is queried to compute CV scores as number of C_i concepts subsuming each curriculum CV_j . The search process returns curricula sorted by score and the list of URLs corresponding to such sorted list is retrieved by querying the DBMS. The system provides a query GUI implementing a novel technology for Web 2.0 increasing the interactivity and the dynamism of web applications: Ajax [6]. The GUI also provides multilingual support, by exploiting information for translation stored in the database. The layout of the home page providing the GUI for the proposed search engine, shown in Figure 2, is divided in two main sections. The left one contains the menu through which the user performs an intensional



Figure 2: Main page layout

navigation of the knowledge domain by browsing the ontology, which is shown in a tree-like representation. Such a visualization is application independent and is built on the fly. The column on the right is aimed at containing the chosen search parameters, selected by using the Drag and Drop functionalities allowed in the GUI.

Consider for example the request for a task requiring *competences in engineering, ASP technology, and advanced knowledge about object oriented programming*. Consider also that *candidates who lived in United Kingdom(UK) for an unspecified period of time are preferred*. In order to formulate such a query the user has to drag concepts **Engineering**, **ASP**, **UK**, **OOB** from the taxonomy to the query panel. The selection of a concepts may cause the visualization of an option menu with the relations associated with the selected concept. In the introduced example the relations **livingIn** and **hasAdvancedKnowledge** have to be chosen for the concepts **UK** and **OOB**, respectively. The resulting conjunctive query is formalized as follows:

$$Q = \mathbf{Engineering} \sqcap \mathbf{ASP} \sqcap \exists \mathbf{livingIn.UK} \sqcap \exists \mathbf{hasAdvancedKnowledge.OOB}$$

Without loss of generality we show the application of the score computation algorithm to only one annotated curriculum, *CV*, belonging to a *Graduate in Engineering, living in UK*. The algorithm returns the vector $V = \{(CV, 2)\}$, because both the **Engineering** and the $\exists \mathbf{livingIn.UK}$ conjuncts of Q subsume *CV*. If several CVs are available, their scores are computed according to the proposed algorithm and the sorted list of URLs to the curricula web pages is finally shown in the results page.

3 Conclusions

We proposed a semantic based search engine for curricula supporting the process of retrieving professional knowledge. The search engine is provided with a GUI for query formalization exploiting the novel AJAX technology for dynamic and interactive web applications development.

References

1. F. Baader, D. Calvanese, D. Mc Guinness, D. Nardi, and P. Patel-Schneider, editors. *The Description Logic Handbook*. Cambridge University Press, 2002.
2. Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, 248(4):34–43, 2001.
3. T.R. Gruber. Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human Computer Studies(IJHCS)*, 43(5/6):907–928, 1995.
4. N. Guarino. Formal ontology in information systems. 6-8 June 1998.
5. Volker Haarslev and Ralf Möller. RACER system description. *Lecture Notes in Computer Science*, 2083:701–706, 2001.
6. Tim O'Reilly. O'reilly – what is web 2.0.
7. OWL. www.w3.org/TR/owl-features/, 2004.

Benefits of an SOA-Implementation for KM - Illustrated by a Case Study

Wilko Krass

(University of Erlangen-Nuremberg, Germany
krass@like.e-technik.uni-erlangen.de)

Abstract: Service oriented architectures (SOAs) are a new paradigm in IT-management. This paper presents an SOA-implementation in the intranet portal of a 520-employee research institute, which was launched to accompany parallel knowledge-management (KM) activities. Furthermore it explains the benefits of an SOA-based IT-landscape for KM-activities.

Keywords: Service Oriented Architecture, Intranet Portal, Knowledge Management, Yellow Page Service, Knowledge Directory, Case Study

Categories: D.2.11, D.2.13, H.3.4, H.3.5, H.5.1, H.5.2, H.5.3

1 Introduction

Nowadays, *modern* organizations have to be able to change their IT-infrastructure fast. Service oriented architectures (SOA) support easy adaption and fast development by modular architectures, loosely coupled services, reusable functionalities, standardized interfaces and open protocols. Complex business logic is realized by joining services („composition of services“) and not inside the services [Krafzig, 05]. These qualities are useful for an infrastructure in support of KM-activities, too.

Because human-oriented KM is difficult as it depends on the voluntary participation of the employees and cannot simply be ordered top-down [Drucker, 93], certain KM-activities will work and others will not. Success and failure depend on a huge amount of factors reaching from intraorganisational culture and politics to the design of the user interface. Therefore an iterative - not to say almost a trial and error - strategy has to be applied for the introduction and evolution of KM-activities. This leads in conjunction with software development to considerable costs.

In order to keep costs and effort manageable, it makes sense to build the software modularly, step by step as separate services, prove their acceptance by the users and link them later on.

This paper presents an SOA-implementation in an intranet portal of a 520-people research institute, which was launched to support KM-activities. Mainly, these activities are:

- Companywide standardization of terms with a Topic Map.
- Knowledge identification in the sense of Probst [Probst, 99], realized by a yellow page service in order to build a human-oriented knowledge directory [Lehner, 06].

- Knowledge sharing in the sense of Web2.0: A web-based document storage platform is installed and centrally administrated, while the employees themselves provide the content, i.e. knowledge-intense documents.

2 The Services

This section presents the services for the KM-activities and their interconnection.

2.1 Directory Listing of the Staff Members

The first important service installed was a directory listing of the staff members. It provides portrait pictures, phone numbers, location data and further employee related information.

2.2 Topic Map

The next important step was the introduction of a polyhierarchical topic map. Its major task is to standardize technical terms companywide. Its major service is a web-based listing of terms that can be – in the sense of keywords - utilized for specifying any kind of object.

Furthermore, the topic map offers a pseudo-hierarchical navigation, a search interface and an index - similar to the index at the end of a book.

2.3 LDAP-based Authentication Service

A frequently underrated service - because taken as a matter of course, is authentication. It will become a considerable problem, if not well introduced to the staff. If an employee voluntarily wants to classify his knowledge for a yellow page service and needs to get new account data (username and password), he might decline, and the knowledge directory stays empty: the hurdle is too high.

At the institute a new LDAP-based authentication service was rolled out companywide. LDAP (Lightweight Directory Access Protocol) is a standardized application protocol for querying and modifying directory services running over TCP/IP and is therefore ideally suitable for reusable user management.

2.4 Linking Topic Map and Employees' Directory Listing

As authentication is possible now, a user can authenticate and then specify his knowledge by using the topic map. The resulting information gets stored in a dedicated database, which can be utilized by the topic map and the directory of employees.

When navigating through the topic map, employees who have knowledge in the present chosen topic will be displayed. On the other hand, when moving through the extended telephone directory of a department, not only pictures, names and phone numbers are provided, but also the knowledge of the displayed people. A classical yellow page service is derived.

2.5 Project Map

In order to improve transparency about projects and project members a polyhierarchical project map was built and integrated into the portal. This project map supplies information about projects as well as a navigation and search interface. In order to specify a project more precisely, the project manager can use the topic map.

The project map's eminent service is access control. Other services can query the project map, whether the user is a member of a project, and grant access to content or modules related to the project or reject it.

2.6 Document Storage

Document storage is a common service in a web-based intranet portal. It simplifies file exchange among a distinct group of persons and is independent of the operating system the user has on his PC.

In the portal the installed document storage solution uses the project map access control interface, if a document should only be seen by a distinct group of persons.

Moreover it was registered at the topic map. After uploading a document the user is led to the keyword list view of the topic map in order to specify the document content. After choosing suitable terms, the user is conducted back to the metadata view of the document, where the document storage software offers further functionalities on the file, like e.g. mailing, commenting, etc.

2.7 Component-Ordering System

The portal offers a stand-alone service for ordering electronic parts. This service again was linked with the project map for access control, because only members of a project should be able to order parts on the project's account.

In order to specify the component itself as well as some of its properties - if necessary, the user can utilize certain parts of the topic map.

2.8 Architecture Overview

Recapitulating, the architecture mainly offers the following functionality (Fig. 1):

- Persons can assign topics themselves in order to specify their knowledge (C1).
- They can upload documents (C2) and specify the content using the keyword service of the topic map (C3).
- Additionally, if access control is necessary, documents can be assigned to projects (C4). This way, only project members can access the documents.
- Employees can be assigned to projects (C6) in order to enable access to project-related content.
- Projects can be specified by keywords from the topic map (C5).
- In principle, other services, like e.g. the mentioned component-ordering system, play the same role inside the architecture shown in Figure 1 as the document storage solution.
- As all of the described activities need authentication, the LDAP-service is always involved (not shown in Figure 1).

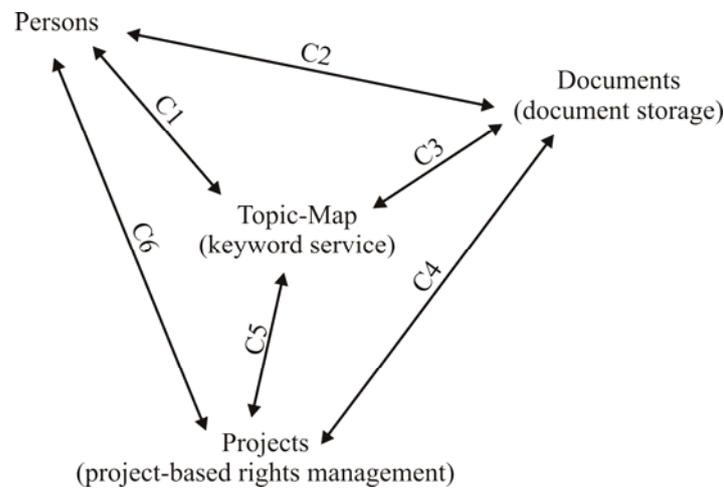


Figure 1: Interconnected Services

3 Summary and Benefits

The provided case study presents an implementation of an SOA-based IT-infrastructure. The topic map and the project map provide interfaces which other services can use for their own means. They do not have to implement the desired functionality on their own or care about derivation of necessary data. This prevents building piles of redundant data. The maintenance of software as well as data becomes easier and cheaper.

Secondly, through daily usage the employees get familiar with the centrally placed Topic Map. This is an important factor for KM-activities: It automatically involves marketing for KM and makes the employees aware of the activities.

Thirdly, the software components get more independent from each other. This does not only enable parallel software development, but it eases adding or removing additional services asynchronously.

Lastly, the connection of new services to already established and well-known services eases the gain of acceptance by the user.

References

- [Drucker, 93] Drucker, P.F.: Die postkapitalistische Gesellschaft. Econ, Düsseldorf (1993)
- [Krafzig, 05] Krafzig, D., Banke, K.: Enterprise SOA. Prentice Hall, New Jersey (2005)
- [Lehner, 06] Lehner F., Scholz, M.: Wissensmanagement. Hanser, München (2006)
- [Probst, 99] Probst, G.; Raub, S.; Romhardt, K.: Wissen managen. Gabler, Wiesbaden (1999)

Conception of Knowledge Management Supported by Information Technologies

Inga Beleviciute

(Vilnius Gediminas Technical University, Lithuania
inga@gama.vtu.lt)

Abstract: There are many technologies calling themselves knowledge management systems in the market but most of them deal only with information management. The main difference between them is that knowledge management systems are oriented into people and participated in tacit knowledge capturing. The paper describes set of information technologies which could take part in all process of knowledge management and especially in collecting tacit knowledge.

Keywords: Knowledge management technologies, knowledge creation, knowledge transfer, knowledge applying, knowledge store, retrieval

Categories: H.1, H.3, H.4.2, H.4.3, H.5.1

1 Introduction

Knowledge includes both the experience and understanding of the people in the organization and the information artefacts, such as documents and reports, available within the organization and in the world outside [Marwick, 2001]. Polanyi divided knowledge into two types [Polanyi, 1962, 1996]: explicit knowledge, the knowledge included in documents or books and tacit knowledge, the knowledge that can be acquired by experience, communication.

Knowledge management considers mainly the four basic processes: knowledge creation, storing, transferring and applying [Alavi, 2001]. Each of these steps could be supported by technologies which are detailed further. These processes do not present a monolithic set of activities, but an interconnected and intertwined set of activities.

There are many technologies calling themselves knowledge management systems in the market but most of them deal only with information management. The main difference between knowledge management and information management is that working with *objects* (data or information) is information management and working with *people* is knowledge management [Grey, 1998]. Also knowledge management deals with tacit, unstructured knowledge which is one of the hardest tasks for information technologies, to capture it and convert it into an explicit and searchable form.

The paper describes set of information technologies which could take part in all process of knowledge management and especially in collecting tacit knowledge.

2 Knowledge creation

Knowledge creation Nonaka [Nonaka, 1995] describes as the conversion between explicit and tacit knowledge and divides it into four stages: socialization, externalization, combination, internalization. IT could participate in this process by collecting tacit and explicit knowledge during knowledge creating, then to structure and convert them into a searchable form and use it in other KM processes.

Socialization happens when people interact with each other and share their knowledge. This is one of hardest tasks for IT to take part in this process, but still some technologies could be applied such as Chat, Videoconferencing and etc. [Figure 1] shows the conversion of tacit knowledge within a Chat. Each chat participant chats and writes his tacit knowledge in text which after could be summarized, structured by text mining technologies. A mechanism should be provided to let participants choose themselves which chat should be handled by technologies not to make people avoid communication.

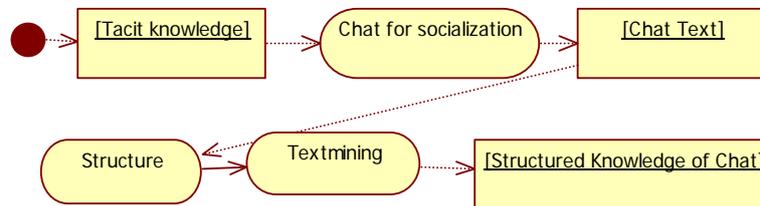


Figure 1: Socialization using chat

Socialization can be supported by videoconference. Persons discuss and share their tacit knowledge in videoconference. Voice records collected in this process could be arranged by speech recognition and after processed by text mining tools. With the examples of technologies which could be used for socialization mentioned above only a small part of tacit knowledge could be collected by IT.

Converting tacit knowledge to explicit is externalization. Information technologies could support this process by Forums, Questions-Answers bases, Web 2.0 social software and etc. [Figure 2] shows an example how the social software Blogging takes part in externalization processes.

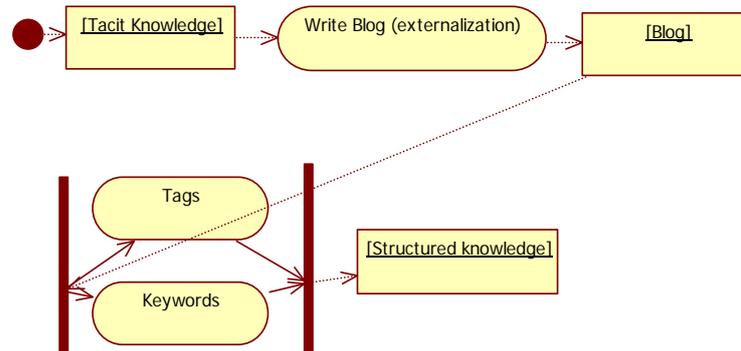


Figure 2: Externalization in Web 2.0

Combination could be influenced by IT the most because it is an interaction between explicit knowledge. Here many information management technologies play a big role. For example an explicit knowledge artefact document could be processed by many technologies shown in [Figure 3] to get more structured knowledge which could be easier to search and use for other purposes like detect experience of a person in a particular area .

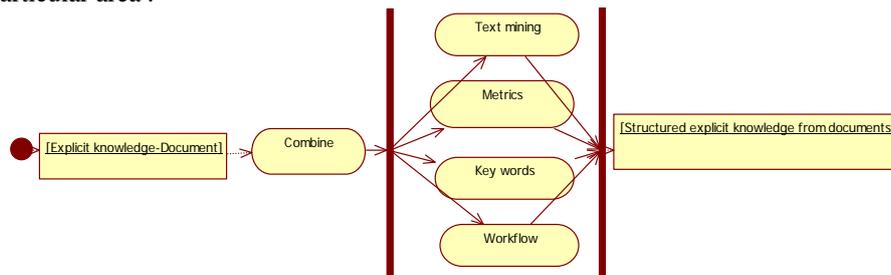


Figure 3: Combination of explicit knowledge

Internalization happens when people learn from documents, books and etc. and explicit knowledge becomes tacit. This process could be supported for example by learning technologies [Figure 4].



Figure 4: Internalization via learning activities

3 Knowledge storing/ retrieval

Organizational knowledge is stored in various forms like documents, multimedia bases, knowledge artefacts which come from knowledge creating process and etc.

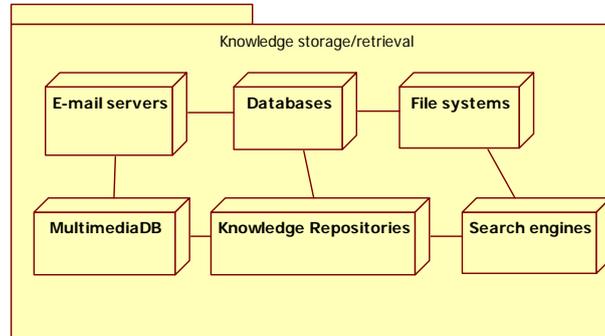


Figure 5: Knowledge storage/retrieval

4 Knowledge transfer

Knowledge transfer occurs in different levels: between individuals, from individuals to explicit knowledge, from individuals to groups, between groups, across groups, and from the group to organization. Information technologies can support knowledge transfer by creating collaboration environments, video conferences and etc.

5 Knowledge applying

Knowledge applying happens when existing knowledge are added, used and reused by other knowledge. IT can have a positive influence on knowledge application. Examples are workflow, knowledge bases and etc.

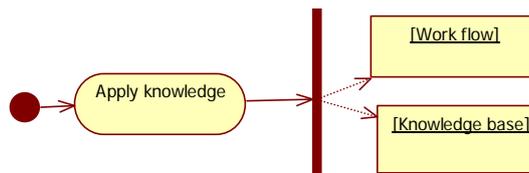


Figure 6: Knowledge application process

6 Conclusion

Many technologies could be applied in different steps in knowledge management to facilitate these processes. The particular set of them depends on situations, requirements and needs and could vary in different ways. KM applications could be used to solve or improve one particular problem or be comprehensive solutions in organizations. Though, despite growing information technologies possibilities to influence and improve knowledge management in organizations, successful implementation of the latter depends on the culture and understanding about the necessity of it in organization. The structured knowledge from chat, videoconference, documents, forums and individual person profile (which could be filled up by the person himself and combined with data from personal DB) could be reused to estimate experience and knowledge in particular fields.

References

- [Alavi, 2001] M. Alavi, D. E. Leidner, Review: Knowledge management and knowledge management systems: conceptual foundations and research issues. *MIS Quarterly* Vol. 25 No 1, pp. 107-136, March 2001
- [Grey, 1998] D. Grey Knowledge Management and Information Management: The Differences. August 1998
- [Marwick, 2001] A. Marwick, Knowledge management technology. *IBM systems journal*, Vol. 40, No 4, pp. 814-830, 2001
- [Nonaka, 1995] I. Nonaka and H. Takeuchi, *The Knowledge Creating Company*, Oxford University Press, Oxford, UK (1995)
- [Polanyi, 1996] M. Polanyi, *The Tacit Dimension*. Routledge & Kegan Paul London, 1996
- [Polanyi, 1962] Polanyi, M., *Personal Knowledge*, Corrected edition, Routledge, London

Intellectual Capital Report Controlling

Wolfgang Halb

(Graz University of Technology, Austria
wolfgang.halb@gmx.at)

Manfred Kraudinger

(Infineon Technologies Austria AG, Graz, Austria
manfred.kraudinger@infineon.com)

Abstract: This paper evaluates methods that can be used for effectively generating and utilizing intellectual capital reports (ICRs) for external and internal reporting as well as for internal knowledge-sensitive management decisions. Research is accompanied by a case study in a high-tech enterprise (Infineon Technologies Austria AG). An intellectual capital report controlling process has been developed. The findings of this research will be of use to all knowledge-based companies and institutions as they demonstrate how and why to measure intellectual capital. The paper discusses the proposed controlling system and the most important findings. An ICR is a multi-indicator system used to depict the development of intangible assets. As there are currently only unsatisfactory accounting guidelines for the reporting of intellectual capital available, an emphasis on techniques for selecting indicators for the report has been made. As a result, convincing reports for external addressees can be generated. Furthermore, methods for using this information as a foundation for knowledge-sensitive management decisions are drafted. The ideas presented in this paper are based on a master's thesis written at Infineon Technologies Austria AG.

Keywords: Intellectual Capital Report, Intellectual Capital Statement, Indicator selection, Knowledge Measurement

Categories: A.1, H.1.0, H.3.2, H.3.5, H.5.1

1 Introduction

Intellectual capital reports (ICRs) are going to become a popular means of measuring knowledge. At present, almost all common accounting guidelines recommend the publication of information about intangible assets. Some guidelines also exist that explicitly recommend the reporting of intellectual capital/assets (e.g. [DRSC 2005], [METI 2005] and [SKE 2005]). An intellectual capital report can not only be used to inform external parties, but also as internal steering instrument.

This paper evaluates methods that can be used for effectively generating and utilizing intellectual capital reports for external reporting as well as for internal knowledge-based management decisions. The research is accompanied by a case study at Infineon Technologies Austria AG. An intellectual capital report controlling process has been developed during this project. This process controls the creation and utilization of a meaningful report. The ideas presented in this paper are based on a master's thesis [see Halb 2007] which contains more detailed information. There is still potential for future research though.

2 Intellectual capital report controlling framework

The intellectual capital report controlling framework is an extension to a basic intellectual capital report model like the one described in [Koch, Leitner, Bornemann 2000]. The controlling process adds features to generate a meaningful report that can also be used for knowledge-sensitive management decisions. Research was based on the ICR 2005 of Infineon Technologies Austria AG which also follows the fundamentals of the basic model mentioned above.

The basic model does not provide sufficient instructions on how to select meaningful indicators and lacks certain capabilities of a steering instrument. With our controlling process methods are introduced that aid in the selection of an appropriate indicator set and that prepare information as a foundation for knowledge-sensitive decisions. In order to get the controlling process started, the management has to be motivated to actively support the intellectual capital reporting. At Infineon Technologies Austria the initial impetus for implementing this process came from top executives. The next important step is to designate responsible controllers that track the execution of the whole process and act as an interface to the management. We have developed software tools that support the controller in collecting and analysing the data. Then the company's knowledge goals have to be defined. The following selection of indicators is discussed in detail in [Section 3]. Further details about the subsequent stages, where the data is collected and utilized, are given in [Section 4].

3 Selection of indicators to report

The selection of an initial set of indicators is a lengthy but important task. As stated in [Ley, et al. 2002] the use of a sample set of indicators can aid in finding suitable indicators. There are also some guidelines available that can aid in the selection of an initial indicator set (e.g. [BMW 2005] and [DMSTI 2003]).

The initial indicator set is then analysed to find missing, isolated and redundant indicators. Based on assessments cross-influences are calculated. Indicators with low overall influence are isolated and therefore candidates for being removed from the set. Categories or knowledge goals with few related indicators need to be reviewed and it might be necessary to add more relevant indicators for this category/knowledge goal. All dependencies are also visualized using automatically generated dynamic graphs with radial layout that represent all direct and indirect influences of a given indicator or knowledge goal. An example screenshot is shown in [Figure 1].

It is possible to dynamically navigate through the data and examine each item individually. The main advantage of this visualisation is that it can give a comprehensible overview of all relations for each indicator or knowledge goal. The node-size is determined by the overall influence of the indicator and edges are also coloured and sized according to the strength of the influence. Items located on the inner ring have a direct relation with the item selected in the centre.

The indicator set from the ICR 2005 of Infineon Technologies Austria has been analysed using these methods. As a result, very important indicators for the attainment of previously defined strategic knowledge goals could be identified. It was also possible to locate indicators that are not sufficiently related to knowledge goals.

Finally a set of indicators exists, that covers all important areas and that is relevant to the company's strategic knowledge goals. These figures are capable of delivering an objective, transparent and convincing overview of the company's intellectual capital and resources. This is important for external reporting and internal steering activities.

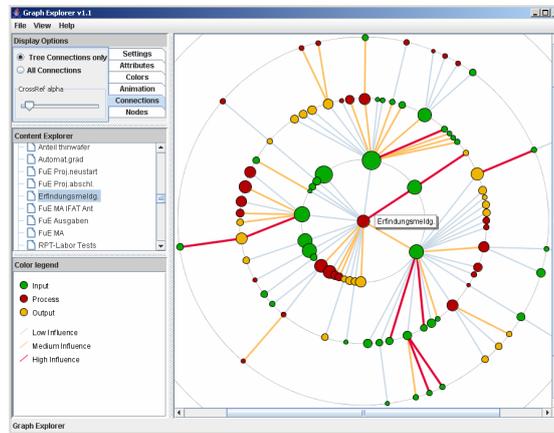


Figure 1: Screenshot of GraphExplorer showing an example indicator set analysis.

4 Data processing

During the case study a software solution was built to support the data collection process. During design and development, lessons learned from a previous data collection have been considered. Main challenges faced in the earlier creation of an ICR were the precise definition of calculation rules for the indicators, the data collection from decentralized departments and the validation of data collected.

The "ICR-Tool" is an intranet-application that can be accessed by every authorised employee using only a web-browser. The main feature of the software is that data can be centrally collected, stored and analysed. Once a responsible owner for each indicator is entered in the system, the owners get notified by e-mail and can either supply the requested data personally or delegate this task to other employees. After the data has been entered in the system, the data has to be verified by another employee before it is released. Advantages of using the ICR-Tool instead of manually collecting the data are that it is a decentralised process where several employees can supply/verify data simultaneously and that every input/verification is traceable.

In the case study we analysed the results using a combination of the coverage analysis as described in [Section 3] with a classic sensitivity analysis. The attainment of previously defined target values can also be considered in the analysis. This analysis then reveals areas for improvement. We also gain information about the qualification of indicators for reaching a desired result and possible side-effects. The information from the graphical analysis can be used to visually identify all relevant inter-relations, connections and dependencies.

5 Conclusions

The work presented in this paper demonstrates a process to create a convincing intellectual capital report for external addressees as well as for internal purposes. By extending a basic reporting model with controlling features, a reporting process was developed that makes continuous reporting on intellectual capital possible.

By analysing and peer reviewing the set of indicators, the set can be refined iteratively. The coverage analysis therefore leads to a balanced selection of indicators. As a result, a transparent and convincing intellectual capital report can be generated. The use of a purpose-built distributed software-solution for collecting the data proved to be very convenient. The software-implemented workflow of data supply and verification by different employees secures data accuracy which is crucial. We also drafted methods for analyzing the data collected.

6 Future Work

The next step for improving the benefits of intellectual capital reporting to the company will be an extended research on the utilization of the data collected. The process model with its associated methods presented in this paper also has to be evaluated and refined empirically. As the model was evaluated in a case study in a large enterprise, the applicability of this model should also be examined for SMEs.

References

- [BMWA 2005] Bundesministerium für Wirtschaft und Arbeit: "Wissensbilanz – Made in Germany. Leitfaden 1.0 zur Erstellung einer Wissensbilanz"; BMWA, Berlin (2005)
- [DMSTI 2003] Danish Ministry of Science, Technology and Innovation: "Intellectual capital statements. The New Guideline"; DMSTI, Copenhagen (2003)
- [DRSC 2005] Deutsches Rechnungslegungs Standards Committee: "Deutscher Rechnungslegungs Standard 15"; DRSC, Berlin (2005)
- [Halb 2007] Halb, W.: "Intellectual Capital Report-Controlling: Conceptualization of a software-supported Intellectual Capital Report-Controlling process for a high-tech enterprise"; MSc Thesis, Graz University of Technology (2007)
- [Koch, Leitner, Bornemann 2000] Koch, G. R.; Leitner, K.; Bornemann, M.: "Measuring and reporting intangible assets and results in a European Contract Research Organization"; Proc. Joint German-OECD Conference Benchmarking Industry-Science Relationships, Berlin (2000)
- [Ley, et al. 2002] Ley, T.; Rollett, H.; Dösinger, G.; Tochtermann, K.; Bruhnsen, K.; Droschl, G.: "Implementing Instruments for Managing Intellectual Capital"; Proc. Third European Conference on Knowledge Management, Dublin (2002), 400-412.
- [METI 2005] Ministry of Economy, Trade and Industry: "Guidelines for Disclosure of Intellectual Assets Based Management"; METI, Tokyo (2005)
- [SKE 2005] Society for Knowledge Economics: "Australian Guiding Principles on Extended Performance Management"; SKE, Melbourne (2005)

Knowledge Management Systems: Towards a Failure Theory

Bashar Ababneh

(Aston Business School, UK.
ababneba@aston.ac.uk)

John S. Edwards

(Aston Business School, UK.
j.s.edwards@aston.ac.uk)

Abstract: Knowledge Management literature suggests that failure is a non trivial issue. Crucial to researchers' understanding of this phenomenon are: definition, causality, and predictability. These issues are addressed in this paper, as it aims to propose a methodology to examine Knowledge Management Systems Failure. The position advocated here is that Knowledge Management Systems are descendents of Information Systems; thus, we take up the Systems Failure Approach from the Soft Systems school of thought, as a potentially fruitful methodology to be implemented in Knowledge Management Systems failures studies.

Keywords: Knowledge Management Failures, Systems Failure Approach, Knowledge Management Failure Analysis, Failure Analysis Methodology, Knowledge Management Failure Theory.

Categories: A.0, A.1, H.1.0, H1.1

1 Introduction

This paper approaches knowledge management systems failure by capitalising on the much larger literature of information systems failures. This is not to claim that information systems are the only important aspect in Knowledge Management, but merely that information systems has contributed to Knowledge Management evolution and that some of the issues in the two fields are similar. The authors propose using the Systems Failure Approach which allows a clearer road map for future Knowledge Management failure research and praxis. The theory promoted here can be viewed from Gregor's (2006) taxonomies to be; explaining, predicting, acting and designing theory for Knowledge Management Systems failure, or type IV & V theory.

2 Systems failure approach (SFA)

“The systems failure approach has two key features: conceptualisation and modelling of the failure situation as a system(s); and comparison of that system(s) with a model of a robust system that is capable of purposeful activity without failure in order to reveal lessons about the failure(s)”[Fortune and Peters 2005. P.97]. SFA's roots are both systemic and pragmatic, making it, we believe, an ideal KMS failure study

approach. Figure 1 shows a version of the SFA modified for analysing knowledge systems.

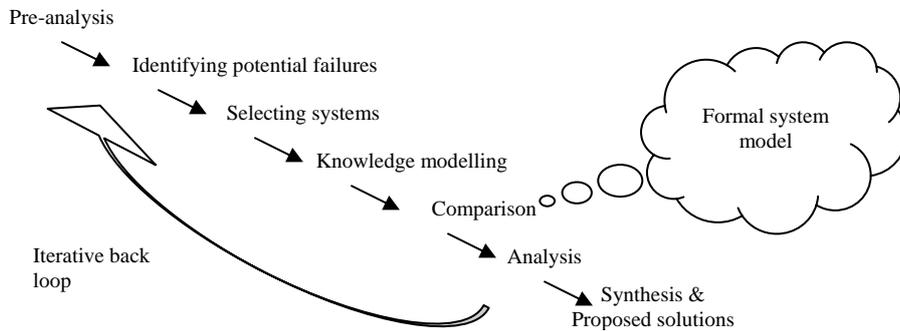


Figure 1: Knowledge systems failure approach, adapted from [Fortune and Peters 2005]

2.1 Pre-analysis

A holistic picture is derived by balancing the following perspectives: purpose, scope, time and money resources pertinent to the understanding of the Knowledge flow, which minimises the risk of omitting relevant details. In figure 2 an example of a spray diagram based on Newell et al’s [2002] BankCo case study shows different factors and how they were building up towards the failure of using iWeb effectively.

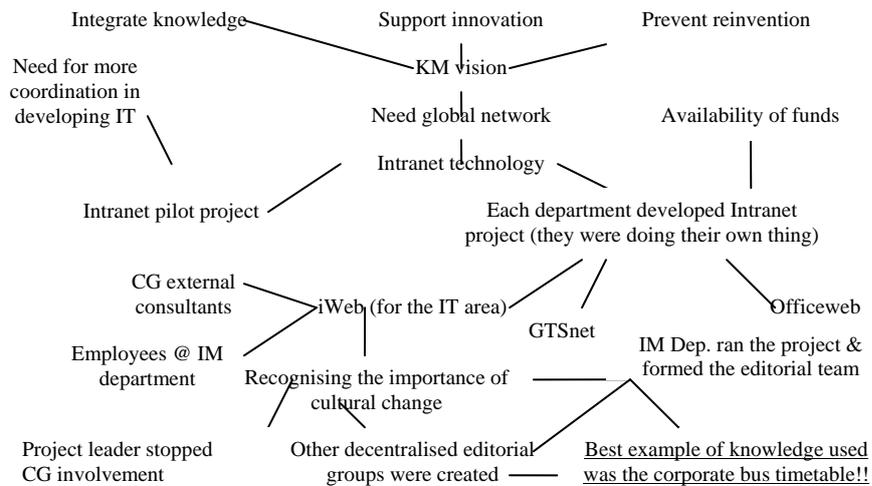


Figure 2: Spray diagram for BankCo

2.2 Identifying potential failures and selecting systems

Identifying failures builds on the pre-analysis by trying to focus more on the areas that are going to receive further attention and analysis. Recognising the immaturity of

knowledge shared lies at the core of this stage, which was demonstrated when employees indicated that retrieving the corporate bus timetable is the best example for knowledge looked up through the system!! (See figure 2). SFA of BankCo poses the question of what system should be studied in order to analyse this potential failure, here the focus is placed on the iWeb initiative.

2.3 Knowledge modelling

Figure 3 represents how KM at BankCo should be in an ideal system. Clearly the vision of KM at BankCo is communicated through this process as knowledge should be gathered from different geographic branches, and scrutinized using the bank's existing experience to offer service for a local client about an overseas business opportunity.

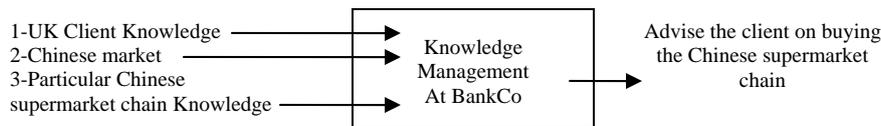


Figure 3: Input-output Diagram for BankCo

2.4 Comparison

Developing a Formal System Model (FSM) is the first step in comparing the actual system, the one conceptualised from failure, to an ideal KM system. Figure 4 portrays FSM through a hierarchical structure of four levels of interacting systems (namely; environment, wider system, iWeb system, and three sub-systems). The FSM explains how the iWeb KM initiative is dependent on its ability to react to surroundings' changes; this requires continuous reflective communication that feeds back into system and control system that directs future plans.

2.5 Analysis and synthesis

The prescribed model in Figure 1 offers a methodical, yet flexible approach, demonstrating an ability to adjust to the situation critical details, as the feed back loop can be carried out at any phase; however, it is crucial for analysis where possible alternatives are prioritised according to the issues identified in the pre-analysis phase,

A synthesis for action or design plans should be adequately explained by the findings of the analysis. As suggested in the literature [Edwards and Kidd, 2003] a staged approach where change is incrementally implemented is a worthy candidate.

3 Conclusion

By translating the current situation into a spray diagram the analysis of the failure has drawn (with sufficient detail) attention to the troubling areas of the system, while not losing the holistic picture of the situation. The input-output diagram was used to conceptualise the abstract system's goal, which fostered the way to build the Formal System Model that revealed different hierarchical interactions between systems on

different levels that is expected to lead to a successful KMS. Finally, comparing the FSM with the actual KMS at BankCo reveals the possible areas of improvement that are both desirable and feasible, which makes a strong case for any action or remedy to overcome the current failure. Such an approach relies on interacting with failure closely to obtain the needed details; however, this was not possible in this study as the authors relied on published case study. Hence, a tentative proposition is put forward for future practice and research to judge the suitability of SFA for KMS.

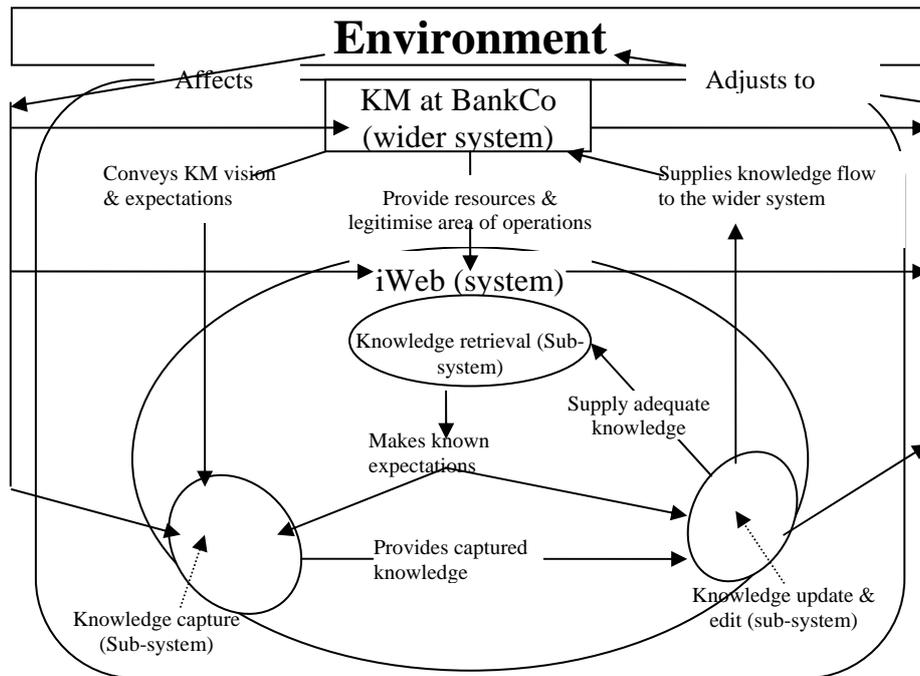


Figure 4: Formal system model for BankCo

References

- [Edwards and Kidd 2003] Edwards, J., Kidd, J.: "Knowledge management sans frontieres": Journal of the Operational Research Society 54, 2 (2003), 130-139.
- [Fortune and Peters 2005] Fortune, J., Peters, G., Information systems: achieving success by avoiding failure, John Wiley & Sons, Chichester (2005).
- [Gregor 2006] Gregor, S.: "The nature of theory in information systems." MIS Quarterly, 30, 3 (2006), 611-642.
- [Newell and Robertson, et al 2002] Newell, S., Robertson, M., et al.: Managing knowledge work. Palgrave, New York (2002).

Reconsidering Relationships for Knowledge Representation

Katrin Weller

(Heinrich-Heine-University Düsseldorf , Germany
weller@uni-duesseldorf.de)

Isabella Peters

(Heinrich-Heine-University Düsseldorf , Germany
isabella.peters@uni-duesseldorf.de)

Abstract: Classical knowledge representation methods traditionally work with established relations such as synonymy, hierarchy and unspecified associations. Recent developments like ontologies and folksonomies show new forms of collaboration, indexing and knowledge representation and encourage the reconsideration of standard knowledge relationships. In a summarizing overview we show which relations are currently utilized in elaborated knowledge representation methods and which may be inherently hidden in folksonomies and ontologies.

Keywords: Knowledge relationships, knowledge representation, ontologies, folksonomies.

Categories: H.3.1, H.3.3, I.2.4, E.2.

1 Introduction

Most methods for knowledge representation organize knowledge in form of concepts and the relations between them, though the level of differentiating relations varies. In current classical methods of knowledge representation only a small number of paradigmatic semantic relations is differentiated and applied. This is somewhat contrary to the current detailed and highly-professional theoretical reflections on relationships, e.g. in [Green, et al., 02], [Khoo, Na, 06], [Storey, 93]. The efforts to establish a *Semantic Web* demand for more sophisticated depiction of knowledge relationships. Ontologies, the core technique for semantic annotation, can make use of a wide range of relations. Furthermore, folksonomies show implicitly existent relations between user-created tags. A closer look at logic-based ontologies and social tagging will help to redefine a classification of relationship types.

2 Relation Types in Knowledge Representation Methods

2.1 Types of Relations and Classical Methods of Knowledge Representation

Two fundamental relations can be distinguished, based on their strength in bonding terms together: *Paradigmatic relations* are fixed concept relations applied to controlled vocabularies. *Syntagmatic relations* are not attached to concepts but originate merely in the actual co-occurrence of terms within a certain setting, e.g.

within the same document. In knowledge representation *generalizable* relations play a major role; these are paradigmatic relations that can be meaningfully used in all (or in most of all) knowledge representation and organization models, such as [Bean, Green, 01]: a) *relations of equivalence*, e.g. synonyms and quasi-synonyms, b) *hierarchical relations*, e.g. meronymy (mereology, part-of relation, part-whole relation, partonomy) [Gerstl, Pribbenow, 96], [Pribbenow, 02] and hyponymy (kind-of-relation, taxonomic relation, taxonomy) [Cruse, 02], c) *associative relations*, which are unspecified connections of concepts that can have any kind of relation.

In information science and documentation, major classical methods of knowledge representation are controlled keyword indexing (*Schlagwortmethode*), classifications and thesauri. *Controlled keyword indexing* is a rather simple form of knowledge representation which works with controlled keywords extracted from natural language and is focused on capturing synonyms. *Classifications* consist of non-verbal notations which represent concepts and relations between them. They are structured hierarchically, without distinguishing different types of hierarchies. *Thesauri* are built of controlled terms (*descriptors*) extracted from natural language. They pay much attention to the relation of equivalence, differentiate hyponymy and meronymy and make use of (entirely undifferentiated) associative relations.

2.2 Additional Relations in Ontologies

Elaborated ontology languages such as OWL support detailed knowledge representation and as many as needed types of relations. Still, some ontologies only make use of hyponymy and meronymy. There are no modeling conventions or other common construction guidelines yet; discussions have just begun [Schulz, et al., 06]. In ontologies, associative relations can be split into a diversity of domain-dependent, specified paradigmatic relations. The investigation of existing ontologies regarding the relationships in use may now lead to new views on how relationships can be distinguished. On the other hand the community of ontology designers needs further evaluation and studies on advantages and disadvantages of certain relational constructions. Guidelines for adequate knowledge modeling have to be derived by analyzing semantic relations in more detail, as also proposed e.g. by [Hovy, 02].

2.3 Syntagmatic Relations in Folksonomies

In latest *Web 2.0* applications users themselves can index documents with freely chosen tags (*folksonomies*) [Mathes, 04]. There are no explicit rules or guidelines on how to annotate documents in social tagging applications. As no structures or formats are provided in folksonomies, all relations between the tags are by definition syntagmatic ones, i.e. they implicitly exist in the co-occurrence of tags. Relationships between tags are not labeled at all, they are not even discussed by the users. And still we can assume that a user does keep the idea of certain relationships in mind during the process of tagging content with their own words. For example, a certain picture on the platform www.flickr.com might be tagged with “Düsseldorf, Germany, Altstadt”, which implicitly includes a part-of relation. Folksonomies include data on how non-experts compose content-descriptive tags, which should be evaluated to gain a better understanding of underlying ideas. We are particularly interested in kinds of relations that can be identified and how frequently they appear.

3 Revision of Standard Relations for Knowledge Representation and Retrieval

3.1 Subdividing Traditional Relations

For years hardly any doubts were cast whether synonymy, hierarchy (with its specifications hyponymy and meronymy) and (unspecified) associative relations are the only paradigmatic relations generally applicable for practical use. In times of ontologies and folksonomies, we have to ask whether new interconnections between concepts should be distinguished. The next steps will be the subdivision of existing relations followed by practical application of fine-grained relationship construction in systems in use and evaluation of the effects on information retrieval efficiency.

As already mentioned, ontologists have started to decompose what was previously summed up as associative relations in thesauri. Hierarchical relations have been a subject to similar theoretical investigations [Winston, et al., 87], [Cruse, 02]. Potential subdivisions of meronymic relations have been gathered [Winston, et al., 87]: 1. component - integral object; 2. member - collection; 3. portion - mass; 4. stuff - object; 5. feature - activity; 6. place - area. Refinements have been provided by [Storey, 93] and [Johansson, 04], more possible differentiations should now be investigated and finally be transferred into practical use.

3.2 Transitivity in Meronymy and Associative Relations

One important trait of knowledge relations is transitivity, because it is practically relevant for information retrieval, more precisely for the conduction of query expansion to improve quality of search results in terms of recall. A relation is considered as transitive if the following holds true: 1) *concept x is related to another concept y in a relation R*; 2) *y is also related to a third concept z in the same relation R*; 3) *x also relates to z via the same relation R*. Thus we have a three-level relation $x \rightarrow y \rightarrow z$ (with \rightarrow as a fixed type of relation), that allows to “skip” one level, as the relation holds also true for $x \rightarrow z$. This works for example in the relation-chain “vehicles \rightarrow automobiles \rightarrow racing cars” (with \rightarrow indicating a kind-of relation). In query expansion for information retrieval, expansions along transitive relation chains may include *several levels* (though it might not always be appropriate), for intransitive relations only a *directly* related concept can be used for expansion.

Some relations are transitive, others are not. The question is what determines the transitivity or intransitivity of relations? Hyponymy can be regarded as transitive; but principles of hyponymy have to be followed very strictly, and there is a lack of guidelines for deciding what justifies establishing a kind-of relation [Cruse, 02]. Meronymy cannot be regarded as generally transitive or intransitive. Both cases can be found for different examples [Pribbenow, 02]. Here, the question arises whether intransitivity in meronymy is an effect of summing up different aspects into one general relation. What if part-ofs were separated into specific types of meronymic relations? For example, all partitive relations regarding geographic entities are transitive. Thus if “geographic components” was established as specific subtype of meronyms, it could be labelled as transitive relation. Similar considerations regard the associative relations. Here transitivity will only be of use, if the associations are

specified in detail. A statement on transitivity cannot be given (or is at least not meaningful) while it is only indicated, that some concept is somehow related to another one – without characterizing the kind of relationship.

4 Conclusions and Future Work

The advent of ontologies and folksonomies asks for reconsidering traditional types of relations in knowledge representation from a practical point of view. Differentiated relations can be explicitly formalized in ontologies or may be inherently hidden in folksonomies. We believe in the necessity of reconsidering relationships regarding their classification and generalizability for the use in knowledge representation and information retrieval. Theoretical studies on semantics of relations are available and may be used as starting points. Future work will have to regard existing approaches in extracting relations from text with automatic extraction tools. Structured representations of relations may then be used for knowledge mining.

References

- [Bean, Green, 01] Bean CA, Green R (Eds.): *Relationships in the Organization of Knowledge*. Dordrecht: Kluwer, 2001.
- [Cruse, 02] Cruse, DA: Hyponymy and its Varieties. In: Green, R, Bean, CA, Myaeng, SH, (ed.): *The Semantics of Relationships*. Dordrecht: Kluwer, 2002, 3-21.
- [Gerstl, Pribbenow, 96] Gerstl, P, Pribbenow, S: A Conceptual Theory of Part-Whole Relations and its Applications. In: *Data & Knowledge Engineering*, 20, 1996, 305-322.
- [Green, et al., 02] Green, R, Bean, CA, & Myaeng, SH, (Eds.): *The Semantics of Relationships*. Dordrecht: Kluwer, 2002.
- [Hovy, 02] Hovy, E: Comparing Sets of Semantic Relations in Ontologies. In: Green, R, Bean, CA, Myaeng, SH (Eds): *The Semantics of Relationships*. Dordrecht: Kluwer, 2002, 91-110.
- [Johansson, 04] Johansson, I.: On the Transitivity of the Parthood Relations. In: Hochberg, H, Mulligan, K (Eds): *Relations and Predicates*. – Frankfurt: Ontos, 2004, 161-181.
- [Khoo, Na, 06] Khoo, CSG, Na, JC: Semantic Relations in Information Science. In: *Annual Review of Information Science and Technology*, 40, 2006, 157-228.
- [Mathes, 04] Mathes, A: Folksonomies – Cooperative Classification and Communication Through Shared Metadata, 2004. <http://adammathes.com/academic/computer-mediated-communication/folksonomies.html> (retrieved March 30, 2007).
- [Pribbenow, 02] Pribbenow, S: From Classical Mereology to Complex Part-Whole-Relations. In: Green, R, Bean, CA, Myaeng, SH (Eds.): *The Semantics of Relationships*. Dordrecht: Kluwer, 2002, 35-50.
- [Schulz, et al., 06] Schulz, S, Kumar, A, Bittner, T: Biomedical Ontologies: What Part-of is and isn't. In: *Journal of Biomedical Informatics*, 39, 2006, 350-361.
- [Storey, 93] Storey, VC: Understanding Semantic Relationships. In: *VLDB Journal* 2, 1993, 455-488.
- [Winston, et al., 87] Winston, ME, Chaffin, R, & Herrmann, D: A Taxonomy of Part-Whole-Relations. In: *Cognitive Science*, 11, 1987, 417-444.

Tacit Knowledge Management (T-KM): The Hidden Agenda of Knowledge Management

Franz Lehner

(University of Passau, Germany
lehner@uni-passau.de)

Abstract: Knowledge Management is meanwhile generally recognized as a discipline established in many organisations and no longer needs a special justification. It is certainly not unrestrictedly positively seen and even some of its supporters saw a hype in the development of the last few years, which is fading and is being replaced by a realist view of what is feasible. In the attempt to gain a proof of success, one bumps into further difficulties since the effects of KM activities are often not easily or explicitly measurable. It is not clear if this would not have developed as positively when not measures would have been taken for the development of knowledge management. The goal of this paper is the attempt of a comprehensive explanation of a new theme which is incorporated in Tacit Knowledge Management (TKM), as well as a proposal for the integration of TKM in the traditional understanding of knowledge management.

Keywords: knowledge management, tacit knowledge, organisational knowledge, KM paradigm, KM views, KM activities, TKM

Categories: A, H

1 Background and relevance of the topic

Knowledge Management is meanwhile generally recognized as a discipline established in many organisations and no longer needs a special justification. It is certainly not unrestrictedly positively seen and even some of its supporters saw a hype in the development of the last few years, which is fading and is being replaced by a realist view of what is feasible. In the attempt to gain a proof of success, one bumps into further difficulties since the effects of KM activities are often not easily or explicitly measurable. It is not clear if this would not have developed as positively when not measures would have been taken for the development of knowledge management. The goal of this paper is the attempt of a comprehensive explanation of a new theme which is incorporated in Tacit Knowledge Management (T-KM), as well as a proposal for the integration of T-KM in the traditional understanding of knowledge management.

Based on the predominate positive experience with knowledge management, in spite of the criticism, and the fact that knowledge management is, in the meantime, for some tasks indispensable, the hitherto development shouldn't be basically questioned. The discipline of Knowledge Management seems to have come to a point, however, in what a methodical further development and a precise understanding of the addressed phenomena is imperative. At the same time, included is the analysis of the until now neglected aspects of Knowledge Management.

Knowledge is the recognition of the relationship between cause and effect and is based on a systematic connection of information. The capability to use, adapt and further develop knowledge, to render it into a profitable, innovative product and service will mark the competition positions of the companies in the next years.

Polanyi originally divided knowledge into two forms, namely explicit and implicit knowledge. Although, one has a differentiated understanding of knowledge today, these differences still play an important role.

Explicit knowledge is knowledge, which can be verbalised and which can be stated. It can be described, questioned, discussed and transported through phrases in different forms, e.g. texts documents, data bases, mathematical formulas. Other people can work with this knowledge, it can be easily transmitted, and it can be understood through logical thinking

Implicit knowledge is however the part of knowledge, which cannot be totally expressed or understood in words. It includes skills, experiences, beliefs etc. and doesn't need extensive formal recording. But it is an important part of human behaviour. The existence and the importance of hidden knowledge (tacit knowledge) was first recognised and described by Polanyi (1966). Implicit knowledge is based on individual experience, personal ideas, beliefs, perspectives, philosophy of life, ideals, values and emotions. It includes skills, action routines, convictions, doctrine and mental schemas.

Knowledge should be treated as a production factor and be used as efficiently as possible. Knowledge management in connection with the modern management practice traditionally includes (among other activities)

- The „management“ of externalised knowledge,
- supporting of the externalization of knowledge,
- the use of the externalised knowledge in connection with the operational tasks by the employees or software.

The externalised knowledge of an organisation is e.g. in its data bases, archives etc. and can in this way be optimised or examined. Implicit knowledge about methods of Knowledge Engineering are transformed through specific processes into explicit knowledge with varied success. Interesting is the not externalised knowledge (similar as with the human memory) whereby emergent structures have the greatest influence (e.g. memory, will of learning, absorption speed). Especially with these questions in the background, a concentration on externalised knowledge oversimplifies the real matters and blinds out many aspects. Knowledge itself is only found in humans resp. in the interaction of humans. Based on the foundation of this realization, a broader understanding of Knowledge Management has already been propagated in the last few years.

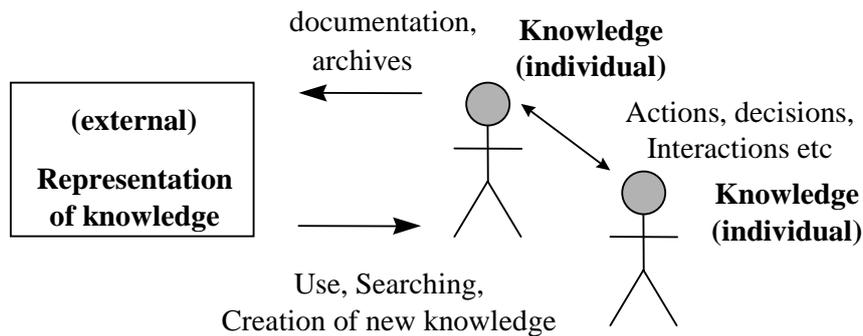


Figure 1: Context of knowledge management activities

2 Initial and traditional understanding of tacit knowledge management (TK-M)

As a result of a growing interest in the existing knowledge of a company or an organisation as decisive success factor within economic competition, "tacit knowledge" is experiencing greater attention next to explicit knowledge. The goal of a European research projects about „Tacit Managerial Knowledge“ (cf. <http://www.tacit-knowledge.org>) is to create a theoretical foundation for tacit knowledge in the area of management, which reflects the significance of the construct in professional success as well as containing the separation of different connected purposes and an explanation of terms.

The main interpretation of Tacit Knowledge Management aims so far at the management of hidden knowledge and the term TK-M was used for the first time in this context, (cf. e.g. Kreiner 2002). Tacit or hidden knowledge, according to a study by Daniel Goleman, makes up around 80% of the knowledge in organisations. By succeeding at least a part of this hidden knowledge to be made transparent, a considerable development potential would be assumed for the business.

A special significance is given in a business context next to the individual or personal knowledge in the collective and organisational knowledge. Schneider describes collective knowledge as a mixture of explicit and hidden or implicit knowledge in a relationship net. Collective knowledge is, in this process, an important source for competitive gain. Because of its nature, it is very difficult to dismantle, reproduce or transfer to other organisations. (cf. Schneider 1996, 21-22 as well as "Paradox of Replication" and the "Paradox of Intelligence"). Nonaka/Takeuchi define collective organisational knowledge to be understood as similar to „ ... capability of company as a whole to create new knowledge, disseminate it throughout the organisation and embody it in products, services and systems“ (Nonaka/Takeuchi 1995, 12). The process of development of collective knowledge requires the continuous cultivation of four forms of knowledge creation. In the shaping of each one of these four forms, it's a matter of a transition between explicit and hidden or implicit knowledge. Figure 2 gives an overview of these connections.

Transition	transition form	processes and examples
explicit -> explicit	<i>combination</i>	Traditional learning process, information Exchange, scientific research
explicit -> hidden	<i>Internalisation</i>	Automatisation, Transition to skills / expertise
hidden -> explicit	<i>Externalisation</i>	Reflexion, Reconfiguration
hidden -> hidden	<i>Socialisation</i>	Shared experience, imitation, exercises

Figure 2: Typology of the knowledge formation (Schneider 1996, 22)

Figure 3 shows a further precision of explicit and implicit forms of knowledge. Wiegand points out specifically, in this connection, to the similarity with other diagrams (e.g. with Hedlund/Nonaka). In this case, though the ability to transfer knowledge is the main point of interest, based on Wiegand's assumption, that implicit knowledge generally cannot be transferred without intensive and personal contact. Furthermore important is the distinction of the different learning levels, which are also supported by different learning processes. Learning is then a further mechanism to collect knowledge.

Learning level	explicit knowledge	implicit knowledge
individual	Phone numbers	Biking, playing an instrument
group	scope and goal of a project	Group norms
organisation	company goals	organiasational norms, culture
knowledge community	qualification for an occupational group	procedural routine in a specific situation

Figure 3: Examples of personalized knowledge forms (according to Wiegand 1996, 329)

Coming back to the initial idea a conclusion is that the multiple aspects and phenomena around tacit knowledge so far are not addressed sufficiently. There is much evidence that a focus on tacit knowledge would be too narrow.

3 Proposal of an extended view of tacit knowledge management

Based on previous considerations, it is becoming clear that there are more tasks and activities in knowledge management than have been consciously perceived until now. A closer look at the postulated tasks and concepts of knowledge management brings us to a phenomenon that could be described in the analogy of „tacit knowledge“ as Tacit Knowledge Management (T-KM). Compared to the previous usage of the term, a new and broader view of T-KM will be suggested, whereby the following tasks and focus areas could be differentiated (including some overlaps). The topics identified so far shall serve as proposal for initiating a more substantial discussion and can be described as follows:

- There are no conscious KM-activities – KM “happens“ but nevertheless; here we are basically dealing with the question; in which situations an active consciousness of the task of knowledge management is above simply letting things happen (i.e. trusting in self organization).
- The actual necessary activities do not refer to shared knowledge, resp. require just about the contrary measures to be taken as what is printed about Knowledge Sharing in the traditional knowledge management books (cf. e.g. Brauner/Becker)
- Tacit Knowledge means, simplified, that we know more than we are conscious of or that we can consciously express. This applies naturally also in the organizational context, not only relating to the individual knowledge, but also for organizational knowledge. T-KM concentrates on the explanation of this knowledge (traditional understanding of T-KM)
- KM-activities are consciously introduced but are not known to all (resp. not to all that should know about them). The consequence, especially in bigger organisations, can be, among other things, uncoordinated KM-activities, „the wheel“ when it comes to KM is always reinvented. – T-KM in this sense can mean an elimination of deficits in meta knowledge about KM-activities.
- The activities (of knowledge management) concentrate on information sharing, so that the actual dealing with knowledge and knowledge processes is neglected.
- Consequences of existing but not explicitly communicated goals of knowledge management
- essential KM-processes are understood as “autopoietical” (self-organising)
- there are more KM-activities as is known to the entrepreneur
- significance of hidden knowledge structures; i.e. informal structures and relationships, which have a specific meaning and which are actually more important as the formal structures and tasks, are taken care of through KM.

Tacit Knowledge Management has to do with the circumstance that in a company or an organisation more processes of knowledge management take place than are consciously initiated or known of (i.e. that the generally shared knowledge about KM or KM activities is too small). As there is tacit knowledge, there will also be TKM because of self organization. It will be important in the future to not let it have an accidental development. The setting up of Tacit Knowledge Management makes it

possible for the businesses to understand knowledge management in its entirety and to take definite measures to e.g. eliminate the deficit of meta knowledge about KM activities. The initial thoughts presented here could form a base for a more general concept which in the long run should lead to more empirical evidence in the whole field of knowledge management. The success of knowledge management should not be left to an accidental development, but extended to an important and until now unnoticed field.

References

- Brauner, E., Becker, A.: Beyond Knowledge Sharing: The Management of Transactive Knowledge Systems. In: Knowledge and Process management, Vol. 13, 1/2006, 62-71
- Goleman, D.: Emotional Intelligence, dtv Publ. Comp, 1977
- Hedlund, G., Nonaka, I.: Models of Knowledge Management in the West and Japan. In: Lorange, P. et al. (Hrsg.): Implementing strategic processes. Oxford 1993, 117-144
- Kreiner, K.: Tacit knowledge management: the role of artifacts. In: Journal of Knowledge Management, Vol. 6, 2/2002, 112-123
- Lehmann, H., Lehner, F.: Perspectives of knowledge sharing for the global firm: A transatlantic perspective Seventh Annual Global Information Technology Management (GITM) World Conference on June 11, 12, & 13, 2006 in Orlando, Florida
- Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company. How Japanes Companies Foster Creativity and Innovation for Competitive Advantage. Oxford University Press, London et al 1995
- Nonaka, I.: The Knowledge-Creating Company. In: Harvard Business Review, Nov. - Dec. 1991, 96-104
- Polanyi, M.: The Tacit Dimension. New York 1967
- Schneider, U. (Hrsg.): Wissensmanagement. Die Aktivierung des intellektuellen Kapitals. Frankfurt 1996
- Wiegand, M.: Prozesse organisationalen Lernens. Wiesbaden 1996

The Implications of Expert Systems in Knowledge Management

Ileana Hauer

(West University of Timisoara, Romania
ileana.hauer@fse.uvt.ro)

Cornelia Muntean

(West University of Timisoara, Romania
cornelia.muntean@fse.uvt.ro)

Abstract: Many organizations are in front of most competitive economic environments, where, in order to survive, they must reduce costs all the time and adopt the most intelligent business strategy. We decided to design an expert system for choosing the most adequate accountancy program for an enterprise and that will serve as an effective tool for knowledge management. The purpose of this paper is to highlight that an Expert System is an innovative tool for managing knowledge in any domain.

Keywords: Artificial Intelligence, Expert Systems, Knowledge-based Economy, Knowledge Management

Category: H.4.m

1 Introduction: Knowledge Management

Knowledge Management (KM) has got outline since the 1990s and now it becomes the most developed branch of management. The crystallization is happening in the context of knowledge revolution, of passing to the new type of economy, knowledge-based economy.

The interest in KM has increased significantly and a lot of organizations are focusing on KM to solve business problems. Peter F. Drucker said that knowledge will be the only competitive resource for companies in the future.

Knowledge management is essentially about knowledge and about the transfer of knowledge. In general, members of an organization possess different kinds of knowledge. The purpose of KM is to facilitate effective transfer of the knowledge to others who have a need for the knowledge in carrying out their responsibilities in the organization [Mahesh, 2004].

The application KM successfully needs the understanding and constructive use of information flows within the organization.

Given the importance of knowledge in virtually all aspects of daily and commercial life, two knowledge-related aspects are vital for viability and success at any level [Wiig, 2004]:

- o *Knowledge assets - the valuable knowledge available to be used or exploited - must be nurtured, preserved, and used to the largest extent possible by both individuals and organizations.*

- *Knowledge-related processes - to create, build, compile, organize, transform, transfer, pool, apply, and safeguard knowledge - must be carefully and explicitly managed in all areas affected.*

In practice KM is the approaches, methods and techniques from companies centered on producing and using the knowledge, providing an higher utilization from the previous period, with multiples valences of knowledge. Successful KM strategies involve a combination of people, process, technology and culture.

Knowledge management draws from a wide range of disciplines and technologies, we will only underline Expert Systems and artificial intelligence (AI). These technologies are applied widely and they are directly applicable to knowledge management. *“Information technology is often cited as an important KM infrastructural capability, enabling or supporting core knowledge activities such as knowledge creation, knowledge distribution and knowledge application”* [Gold et al, 2001].

2 Theory fundamentals: Expert Systems

The raising of decisional act complexity and the necessity of using in short time personal experience or methods from managerial process, required the use of Information Technology. Artificial Intelligence (AI) is a branch of computer science research from which there are derived special computer programs called Expert Systems. An Expert System is a knowledge-based computer program containing expert domain knowledge about objects, events, situations and courses of action, which emulates the process of human experts in the particular domain.

Although the concept “expert system” is quite simple, the building of an ES is very complex because it involves two important and difficult tasks:

- *Knowledge acquisition* from human experts together with the methods used by these experts in problem solving;
- *Knowledge representation*, which means to restate knowledge and methods in an organized manner for their later use.

These two activities are very complex and time-consuming and together they are known as *knowledge engineering*.

2.1 How an expert system provides recommendations

Knowledge representation formalizes and organizes the knowledge. The probably most widely used representation is the *production rule*, or simply *rule*. Knowledge representation by rules uses two structures: *facts* and *rules*.

Facts are statements, which could be true or false, referring to a given aspect of the reality. The set of facts stored in the system forms the *facts base* of the knowledge base. In contrast to traditional informatics applications, an ES builds the solution of a problem by a chaining of deductive processes. But the facts alone do not permit the inference of new knowledge.

A *rule* consists of an IF part and a THEN part (also called a *condition* and an *action*). The IF part lists a set of conditions in some logical combination. The rules are analyzed by the expert system inference engine, which combines the individual rules together to solve a much larger problem. Expert systems whose knowledge is

represented in rule form are called *rule-based systems*. As with a human expert, the system can logically explain the basis for its conclusions.

Knowledge is almost always incomplete and uncertain. To deal with uncertain knowledge, a rule may have associated with it a *confidence factor* or a weight. The set of methods for using uncertain knowledge in combination with uncertain data in the reasoning process is called *reasoning with uncertainty*. Because of the importance of knowledge in expert systems and because the current knowledge acquisition method is slow and tedious, much of the future of expert systems depends on breaking the knowledge acquisition bottleneck and in codifying and representing a large knowledge infrastructure.

3 ES implemented in Corvid for choosing the most adequate accountancy program for an enterprise. Discussion and results

We used Corvid, an expert system generator developed by EXSYS Inc., for implementing our application presented in the this paragraph.

Analyzing the existing offer of accountancy programs on the software market and taking into account some aspects referring to the necessities of the enterprise we implemented an expert system to recommend the most adequate programs using Corvid metablocks. MetaBlocks provide a way to build systems that put generic decision-making information in Logic Blocks that interact with spreadsheet files that contain all of the detailed product data. This approach makes it easy to update and maintain a system by adding or deleting products in the spreadsheet or changing product features or data.

The user interface puts forth some questions along with some options. The users should select the option and from the interaction, the system will draw conclusions and try to recommend the most adequate accountancy software programs for the enterprise. This type of advisory system provides a probabilistic “best fit” recommendation of products to meet a user’s overall needs. The results are quite different from a database approach, which requires a match on every feature.

The database were we stored the features for the programs offered by the software market contains following fields : *Denumire* (the name of the soft program), *Producator* (the name of the soft producer), *Modul monopost* (the list of program modules from which the enterprise will choose the module that fits best with the specific activity of the firm , *Cost* (the price of the transaction of the program), *Viteza de lucru* (the working speed, expressed as an average number of documents that could be done in one minute), *Flexibilitate* (the possibility of the program to allow structure changes in connection with the enterprise activity), *Asistenta post-instalare* (the average number of month/year in which there is assistance and service for the program and training for the employees for working with it), *Lucreaza cu mouse* (the percent of working with the mouse), *Lucreaza cu tastatura* (the percent of working with the keyboard), *Lucreza in retea* (the capacity of the program to adapt to a network), *Procesor “minim”*(minimum requirements cerinte minime referring to the computer processor where the program would run), *Memorie “minim”* (minimum requirements referring to the computer memory where the program would run, *Hard disk “minim”* (minimum free space on the hard disk for an efficient run), *Sistem de*

operare (the operating system on which the program could run), *Distribuito*r (the name of the distributor firm for the program), *Foto* (the image of the program sigle), *Contact* (telephone numbers for more details about the program).

We tried to design a very friendly interface, by asking the user many questions concerning the desires and needs of the enterprise. Finally we presented a recommendation.

4 Conclusions

Expert Systems are versatile tools, which can be used as multi-purpose systems such as decision support system, diagnostic system, self-learning tool, and teaching aid etc. Apart from the user interface with the system, the system itself puts forth some questions along with some options. The users should select the option and from the interaction, the system will draw conclusions and recommend the accountancy programs that fit best with the requirements of the enterprise users and their desires. This paper has highlighted that Expert Systems are innovative tools for managing knowledge in any domain.

References

- [Andone and Mockler 2001] Andone, I.; Mockler, R.; Tugui, Al.: "Developing inteligente systems in economy, *Economica*, 2001
- [Gnana Sekari et al 2005] Gnana Sekari, G. et al, "Expert System-an innovative tool for managing indian traditional medical knowledge", *Proceedings of the 9th World Congress on Health Information and Librarias*, available at <http://www.icml9.org/program/track7/person.php?lang=es&id=3>
- [Gold et al 2001] Gold, A. H.; Malhotra, A.; Segars, A. H. "Knowledge anagement: An organizational capabilities perspective", *Journal of Management Information Systems*, Vol. 18, No. 1, Summer, pp.185-214.
- [Jones 2006] Jones, D.: "Defining Knowledge Management", *KMPRO Journal*, Vol.3, No.1, Spring 2006
- [Mahesh and Suresh 2004] Mahesh, K.; Suresh, J.K.; "What is the K in KM Technology", *The Electronic Journal of Knowledge Management*, Vol. 2, No. 2, pp 11-22, available online at www.ejkm.com
- [Power 2001] Power, D.: "What is the difference between knowledge management and decision support?", *DSS News*, Vol. 2, No. 14, July 1, 2001
- [Prusak 2001] Prusak, L. "Where did knowledge management come from?", *IBM System Journal*, Vol.40, No.4, 2001
- [Radu and Ursacescu 2005] Radu , I.; Ursacescu Minodora : "Information technology and management", *Universitara, Bucuresti*, 2005
- [Rosca et al 2006] Rosca, Gh. et al, *The knowledge society*, Ed. *Economica*, Bucuresti, 2006
- [Wiig 2000] Wiig K, "On The Management of Knowledge" available online at <http://www.km-forum.org/wiig.htm>

Author Index

A

Ababneh, B., 489
Andrews, K., 373
Aqqal, A., 87
Auer, S., 144

B

Balancieri, R., 152
Bayer, F., 118, 200
Becks, A., 8
Beleviciute, I., 65, 480
Benteler, F., 408
Blaise, J., 456
Blazevic, V., 25
Blumauer, A., 71
Böszörmenyi, L., 376
Brandt-Herrmann, G., 136
Braun, S., 321
Bresciani, S., 416

C

Cao, Y., 384
Carlos dos Santos Pacheco, R., 152
Chanduvi-Suarez, C., 263
Colucci, S., 313, 472
Cuel, R., 152

D

Derntl, M., 49
Di Noia, T., 313, 472
Di Sciascio, E., 313, 472
Dieng-Kuntz, R., 348
Doehring, M., 281
Donini, F., 313, 472
Driessen, S., 168
Dudek, I., 456
Durst, C., 111
Duval, E., 392, 297

E

Edwards, J., 489
Efimova, L., 33
El Ghali, A., 348
Elsayed, A., 448
Enoksson, F., 289
Eppler, M., 7, 416
Erren, P., 57

F

Faatz, A., 281
Fill, H., 25
Fink, K., 176

G

Giboin, A., 348
Globocnik, D., 233
Godehardt, E., 281
Goertz, M., 281
Grahl, M., 356
Granitzer, M., 373, 330

H

Halb, W., 485
Hampel, T., 49
Hauer, I., 503
Hefke, M., 128
Herrmann, T., 16, 136
Hesselmann, C., 8
Hildebrandt, T., 192
Hoffmann, R., 365
Hotho, A., 356
Huijsen, W., 168
Huster, J., 8

I

Ihmig, F., 111
Ivanova, M., 65

J

Jahnke, I., 136

K

Karagiannis, D., 25
Keen, P., 209
Kienreich, W., 373
Klamma, R., 384
Kleiner, F., 128
Klerkx, J., 392
Klieber, W., 365
Kotis, K., 340
Kouba, Z., 332
Krass, W., 476
Kraudinger, M., 485
Krellner, B., 424
Kremen, P., 332
Kroha, P., 424
Krüger, M., 118
Kulathuramaiyer, N., 225

L

Latif, K., 400
Lauenroth, K., 144
Lehmann, H., 245
Lehmann, J., 144
Lehmann, L., 192
Lehner, F., 497, 245
Leng, Y., 384
Ley, T., 271
Liegl, J., 330
Lindstaedt, S., 271, 330
Lundh Snis, U., 440
Lux, M., 330

M

Maier, R., 118, 200
Martin, G., 297
Maurer, H., 1, 432, 225
Mayer, R., 400
Mommel, M., 41
Meuthrath, B., 103
Mödrischer, F., 365

Motschnig-Pitrik, R., 49

Müller, C., 103

Muntean, C., 503

N

Nitz, H., 118

Novak, F., 71

O

Oberbichler, A., 376

Ölçer, F., 253

Olivella, J., 305

Oprea, M., 184

P

Pammer, V., 330

Pareto, L., 440

Paulzen, O., 128

Peters, I., 493

Pitner, T., 49

Ploder, C., 176

Povalej, R., 95

Prilla, M., 16

Q

Qiu, Y., 448

R

Ragone, A., 313, 472

Reichel, T., 424

Reiser, P., 6, 71

Rensing, C., 79, 87, 192

Riechert, T., 144

Riegler, W., 217

S

Sabol, V., 373

Sandow, A., 118

Scheir, P., 330

Schirru, R., 41

Schmidt, A., 321

Shirley, S., 111

Slagter, R., 33

Slijp, D., 168

Smith, M., 5
Spaniol, M., 384
Spiliopoulos, V., 340
Stein, B., 408
Steinmetz, R., 79, 87, 192
Stoll, M., 273
Storckenmaier, A., 128
Story, H., 71
Stumme, G., 356

T

Tan, M., 209
Tanasescu, A., 184
Thalman, S., 289, 118
Tifous, A., 348
Tochtermann, K., 1
Trinkaus, H., 160
Trizio, M., 472
Tudor, I., 184

U

Ugbor, U., 263
Ulbrich, A., 271

V

Valarakos, A., 340
Vasilyeva, E., 65
Vorbach, S., 233
Vouros, G., 340

W

Weiß, P., 95
Weller, K., 493
Woitsch, R., 25
Wolpers, M., 297

Z

Zacharias, V., 464
Zaka, B., 432
Zimmermann, B., 79
Zimmermann, H., 111