Mátyás Török

Organizational knowledge extraction from business process models

Ph.D. Thesis

Summary

Supervisor: dr Andrea Kő, Department of Information Systems

Budapest, 2014.
I. Background and overview of the Research .................................................................3
  I.1 Aims, Premises and Main Questions of the Research .....................................4
II. Method of the Research .........................................................................................8
  II.1 Fundamental of social science research .............................................................8
  II.2 Exploratory and validatory research – inductive or deductive logic ..........8
  II.3 Qualitative and quantitative research ...............................................................9
  II.4 Research based on case studies ......................................................................10
III. Results of the Research, Contribution of the Thesis ........................................13
  III.1 Complementary modeling layers .................................................................13
  III.2 Mapping of knowledge elements .................................................................14
  III.3 Multilateral process views – process coupling via semantic transformations ...............................................................15
  III.4 Further Research Questions and Future Directions .............................16
IV. References .........................................................................................................18
V. References of the Author ....................................................................................20
I. BACKGROUND AND OVERVIEW OF THE RESEARCH

In today’s dynamic environment all organizations need up-to-date knowledge for their operations that are based on business processes. Complex organizations use Business Process Management (BPM) tools to model and manage these processes. BPM applications tend to model the organizational processes, together with the required information and other resources needed to perform each activity. BPM yields an overall context, but focuses on a high level of process representation.

My research focuses on extracting, organizing and preserving knowledge embedded in organizational processes captured with BPM techniques to enrich organizational knowledge in a systematic and controlled way. The proposed solution is to extract the knowledge from information stored in the process model in order to articulate, externalize and transfer it. The thesis focuses on the BPM aspects of the solution as I strive to investigate it from the information systems perspective.

The novelty of the solution is based on the connection between process model and corporate knowledge, where the process structure will be used for building up the knowledge structure. Common form of managing knowledge within an organization is the ontology, which provides the conceptualization of a certain domain. By using the ontology and combining it with the process models, we connect knowledge management and business process management in a dynamic, systematic and well-controlled solution.

The proposed outcome is a process ontology – domain ontology duplet, where the domain ontology serves as a knowledge repository, and the process ontology holds the multilateral process information incorporating process structure with the viewpoints of organizational stakeholders and IT implementation.

Structure of the Thesis

The structure of the dissertation is presented in this section. The first chapter deals with problem statement, background, research questions and research methodology. I give an overview about the premise of my work, about the methodology being applied. My main research questions and statements are discussed. The second chapter is about the theoretical background of my work, so I give a detailed description about business process modelling, process modelling standards and languages. The third chapter is
Aims, Background, Premise and Main Questions of the Research

dealing with semantic business process modelling. I discuss ontologies and their role in general in semantic business process modelling. Research questions and statements are proved and detailed in the following chapters. Chapter four describes the proposed solution for capturing every aspect of a business process, extended with the identification and mapping of the knowledge items. The modeling procedure set forth in this section is applied in the case study of the thesis, in chapter five. Chapter six gives an overview about the insurance ontology, which was applied in my solution. Chapter seven deals with the case study process model. I give a detailed description about the ontology in the appendix.

I.1 Aims, Premises and Main Questions of the Research

Enterprises have to operate in a dynamic environment, affected by several external and internal factors. They are acquiring organizational knowledge from numerous sources, whether they know about it or not. In this volatile context of the organizational knowledge creation, it is hard to influence knowledge conversion, maintain a healthy rate of tacit and explicit knowledge as it is discussed in the knowledge conversion theory of Nonaka and Takeuchi (Nonaka & Takeuchi, 1995). One of the main threats from organizational knowledge management aspect is staff movement and mobility. The main challenge is the “wall-to-wall” knowledge articulation in order to provide the organization with up-to-date knowledge. In this way the internal training of employees has to be fully supported. The other dimension of the same problem is supporting the IT systems creation to fit the current requirements of the organization determined by business processes.

Complex organizations use to model and manage their processes with the help of business process management (BPM) tools. These applications help to describe the organizational processes, together with the required information and other resources (amongst other human resources) needed to perform each activity. BPM yields an overall context, but it tends to be static.

Business processes are defined as a sequence of activities. Business processes represent dynamic perspective in enterprises, while the embedded knowledge remains hidden in many cases. From the human resource management view it is required to define unambiguously, who is responsible for the execution of each activity. The RACI matrix (Responsible, Accountable, Consulted, Informed) is used for grouping role types,
bridging the organizational model and the process model. Since we need to acquire knowledge belonging to the job roles, in this sense RACI assigns only job role types to the tasks. The RACI is often used for job role discovery, but it lacks the description of the knowledge elements related to tasks and activities. My research area is dedicated to the challenges of knowledge extraction from business processes.

My goal is to analyze the opportunities of knowledge extraction and to develop a solution to extract, organize and preserve knowledge embedded in organizational processes. This knowledge extraction process will enrich organizational knowledge in a systematic and controlled way. The proposed solution will extract the knowledge from information stored in the process model in order to articulate, externalize and transfer it. Since the business process models are used for the execution of processes in a workflow engine, another very important source for gathering useful knowledge are real-time instantiations of the business processes, that gives a view on the dynamic knowledge, usually represented in the form of different business rules. My other research problem is how to organize the extracted knowledge, what are the appropriate ICT solutions, environment for it.

The novelty of my proposed solution is based on the connection between process model and corporate knowledge repository, where the process structure will be used for building up the knowledge structure. Common form of knowledge representation is the ontology. My research focuses on a framework to build ontologies for both process and domain. In the context of this work, I provide a distinction for the two terms:

Process ontology: Identifies all the artifacts that describe a process, regardless of whether it is structured or not. It allows building clearly and unambiguously all process elements, linked with the domain ontologies that specify enterprise concepts, as well as the business rules, roles, outcomes, and all other inter-dependencies.

Domain ontology: The domain ontology provides vocabulary of concepts and their relationships, captures the activities performed on the theories and elementary principles governing that domain. It is not a glossary of terms, it is what defines the company sphere and represents what the company does.

According to these research challenges, my first research question is investigating the relation of processes and organizational elements:

**Research question 1**: How can we determine the connection between process elements
and other organizational phenomena?
To answer this question, I will analyze the main BPM methodologies and their organizational dependencies. Common BPM methodologies provide the methods and tools to identify several dimensions of organizational environment, such as IT infrastructure elements, or organizational stakeholders as human actors closely related to the organization. Every perspective has its procedures and the knowledge behind them. The challenge lies in a systematic and gapless integration of these viewpoints.
The following research question is dealing with my main research issue; discussion of knowledge extraction methods from business processes:

**Research question 2:** What are the possible approaches of extracting domain specific knowledge embedded in BPM process models?
Answering this question starts with clarifying how can we articulate the hidden knowledge in BPM. I will review theoretical foundations of related fields, like business process management, semantic technology and ontologies.
In my thesis emphasis is given to enrich process models with organizational knowledge, in more strict terms to include knowledge elements in business process models at different levels of granularity. I have to examine what are the preconditions and requirements against processes and how can we organize the extracted knowledge in a most effective and efficient way. The following research question is dealing with the possibilities of the knowledge extraction automation.

**Research question 3:** Is there any possibility for semi-automatic or automatic solution for knowledge extraction from business process models?
To answer this research question I will overview and analyze the semantic business process management and semantic web services literature, and based on that, I will propose my approach for knowledge extraction. Justification of the ontological approach in knowledge management is proved through the presentation of case studies.
I will utilize my research projects experiences, especially which I gained in Prokex (PROKEX, 2013) and eBEST projects (Ternai & Török, 2012).

**Research question 4:** What is the potential for organizations in having knowledge-enriched process repositories?
From the case studies, I will strive to answer the following questions:
How can a proposed method ease the problem of fluctuation? Can it lead to more targeted training? Is a multi-lateral view on business processes enhances the improvement of processes?
II. METHOD OF THE RESEARCH

In reviewing my thesis research methodology I had to comply with the nature of the research as well as the requirements of the Ph.D. School. In case of IT related theses written under the aegis of accredited Ph.D. schools it is a common occurrence for candidates to define solvable tasks in the form of setting up a series of research related questions and providing answers to them instead of making hypotheses. In contrast to theses aiming to prove hypotheses leaving a problem unsolved is not acceptable, but rather it is taken as a failure.

The Business Informatics Ph.D. School of Budapest Corvinus University has been classified to the IT discipline that belongs to the field of social sciences and as such, applying research methods in a kind of ‘hybrid’ way can hopefully be considered to be accepted.

II.1 Fundamental of social science research

Basically all research works have the goal either to explore new theories by searching for unknown relations or to prove discovered but still unproved theories, thus adding to the general knowledge of the given field. These two aims necessitate a different logical approach: while a validatory research requires deductive logic, an exploratory research follows inductive logic.

II.2 Exploratory and validatory research – inductive or deductive logic

The validatory approach is suitable for testing assumptions and hypotheses deducted from the accepted theoretical background of the field of research. It uses deductive logic which is applied to test research theories based on hypotheses. Thus it is clearly visible that making hypotheses is inevitable in a validatory research. Only after having the hypotheses put down in black and white can the researcher proceed to the observatory part of the research and the evaluation of the hypotheses.

The exploratory approach is a good choice in cases when the field of research is completely or largely unexplored. Exploratory researches are carried out typically with three main goals (Szabó, 2000):

- ensure a better understanding of the topic,
- serve as testing the feasibility of future, more thorough researches,
- develop applicable methods for further researches.
In fields where this approach is appropriate, making testable hypotheses would often be too early and untimely. Moreover the process through which theory development takes place is less strict by its nature (Benbasat, Goldstein, & Mead, 1987; Babbie, 1989). Exploratory research is based on inductive logic which says that theories can be developed by analyzing research data and generalization.

When examining Ph.D. theses of our faculty it must be noted that Klimkó doesn’t make any hypotheses in his Ph.D. thesis (Klimkó, 2001), but instead he draws up his research-related expectations. He however emphasizes that it is the inductive approach that makes this possible because his thesis is not of validatory nature. “Amongst the questions there are no deductive ones that could be aimed at validating hypotheses. All questions are of inductive nature. That is why my research questions are about “expectations” instead of “hypotheses” (Klimkó, 2001).

My present research is of exploratory nature and follows inductive logic. In my thesis I am going to identify research questions and tasks along with hypotheses and will explain the importance of the questions. Also, by reaching the goals set in the questions, I am also going to give an explanation on the importance of the chosen topic itself.

II.3 Qualitative and quantitative research

From a methodological point of view, we can take the qualitative and quantitative approaches commonly used in organization evaluation methods as a basis (Balaton & Dobák, 1991). Quantitative methods include the application of mathematical and statistical means for data processing, so these methods can be used in researches where a lot of measurable data are available.

If we want to explore and understand the deeper relations within a discipline without trying to analyze numerical data sets, it is reasonable to use qualitative methods. These are suitable for research fields where a well-founded knowledge base hasn’t been established yet or when the aim is to solve a problem and theory is built based on this solution.

In order to avoid the drawbacks of the methods it is recommended to use methodological triangulation (the application of different research methods and perspectives for analyzing the same question) (Balaton & Dobák, 1991). Types of triangulation are:

- simultaneous application of various quantitative procedures
- simultaneous application of various qualitative procedures
Method of the Research

- combination of quantitative and qualitative methods

My present research is based on qualitative methods because it follows an exploratory, deductive logic without having access to large, measurable data sets.

II.4 Research based on case studies

According to Yin (Yin, 1994) basic research strategies can be based on

- experiments
- questionnaire surveys
- secondary analyses
- historical analyses
- procession of a case study

Yin asserts that it is expedient to use case studies when “…questions of ‘how’ and ‘why’ are asked in relation to current events over which the researcher has little control”.

Case studies examine phenomena in their natural environment and apply several different data acquisition methods with a small number of examination subjects (Benbasat, Goldstein, & Mead, 1987).

The application of case studies is preferred to other methods when researched concepts and relations can’t be examined in an isolated manner. In such situations it is only the method of case studying that can guarantee the necessary depth for a theory’s evolution.

This method has a long tradition in IT literature (Lee, 1989).

The case study approach has many strengths: it provides an overall perspective and enables a more thorough, in-depth understanding. It also helps to reveal such relationships that would remain hidden if a different method was applied (Babbie, 1989)(Galliers, 1992). Bensabat et. al. (Benbasat, Goldstein, & Mead, 1987) make substantial statements in respect to case study based research that, as being idiographic, tries to understand problems in their own context.

Bensabat et. al. summarize main features of the case study based research strategy as follows:

- examines a phenomenon in its natural setting
- employs multiple methods of data acquisition
- gathers information from one or a few entities
- is of exploratory nature
Method of the Research

- no experimental control or manipulation is used
- neither dependent nor independent variables are predefined
- results are highly dependent on the researcher’s ability to integrate
- data acquisition methods can change during the research
- the nature of the phenomenon and the reason for it is the question, not the frequency of its occurrence

Case studies may relate to a single or multiple events and there are countless possible levels of analysis in the research. Case studies are usually based on combined data acquisition methods (archives, interviews, questionnaires, observations), in which results can be both qualitative and quantitative.

The case study approach can be applied in order to reach at least three goals (Eisenhardt, 1989):
- with the intention to illustrate (to explain a theory),
- create an applicable theory,
- test a previously worked out theory.

Case studies can also be used to evaluate whether practice corroborates main theoretical concepts. Eisenhardt and Bensabat et. al. provide a detailed guidance to planning a theory development research based on case studies.

In order to avoid any threats while applying this method, five criteria have to be met (Babbie, 1989):
- a relatively neutral aim should be defined
- known data sources should be used
- an adequate time frame should be examined
- known data acquisition methods should be applied
- consistency with the currently accepted knowledge base should be ensured

The main advantage of a case study based research is its flexibility. It enables the interaction between data acquisition and data analysis. This approach has an outstanding validity: instead of defining concepts, case studies provide detailed illustration.

However the case study approach may come with quite a few drawbacks: it rarely provides an accurate description on the state of a large population and the deductions are rather to be considered as suggestions than definitive conclusions. Reliability may also be an issue in a case study based research, just like its inadequacy to generalize the
findings. The personal nature of observations and measurements can lead to results that can’t be reproduced by others. Secondly it is harder to generalize the in-depth, overall understanding than those results that are based on a strict model and standardized measurements. Thirdly there is a big chance to distort the model (Babbie, 1989).

- As it is of exploratory nature, my present research uses a case study based approach in validating hypotheses.
III. RESULTS OF THE RESEARCH, CONTRIBUTION OF THE THESIS

The basis of my multi-lateral approach is general control-flow oriented business process models. The process modeling starts with the close observation of an existing, real-life process at the given organization. The first step is to conduct interviews with all of the stakeholders of the process to be captured at the company, assess already existing process documentation, document the process development meetings and materials prepared during the actual project. A thorough inspection of the underlying IT infrastructure is also necessary.

The ever-recurring problem of capturing processes is the level of granularity. Setting this appropriate level can be thought of as an optimization problem in itself. If a process model is too superficial, it will not contain enough information to draw conclusions, conduct redesign or utilize it in any other ways. A modeling architecture with unnecessarily frittered details or a model with inhomogeneous granularity results in a confusing process architecture, and consumes unnecessary resources to create, maintain and manage. Ternai et al. collects the parameters have to be set in order to use a process model as a base of semantic transformations (Ternai, Szabó, & Varga, 2013), I abide myself to the guidelines in this work. The level of granularity in modeling a process is set to grant the ability to attach corresponding concepts like roles or information objects to the model.

At this point, the process structure, and meta-information for the IT and organizational viewpoints are recorded, all relevant information resources are elaborated, but organizational knowledge is unstructured, hard to identify and has various, heterogeneous sources.

III.1 Complementary modeling layers

After finalizing the basic process flow, the specific activities within the process model have to be aligned with roles and responsibilities. We capture a view of the inner stakeholders of the organization. We start by collecting all the roles that are related to the given process, and gradually examine, which roles have any relation with a given activity. This task is carried out on the theoretical ground of the RACI responsibility matrix. We determine which the explicit roles are being played by which stakeholder at
the level of a given activity. More precisely, we define according to the RACI, which role is Responsible for the performing of the activity, which role is Accountable for it, which are the roles needed to be Consulted during the execution of the activity, and who to be Informed about the advance, obstacles, completion or other information related to the given activity.

This knowledge is the basis of the proposed outcome, namely to be able to present the knowledge items required by a person in a given role, or in a broader perspective, in a given position.

There are two additional modeling dimensions that play an important part in enriching process information:

Many organizations have a well-structured IT infrastructure map, and in a higher-level process model, IT architecture elements are assigned to the process model at activity level. Modeling tools incorporate sub-models of the company’s IT infrastructure. In this sub-model we define the major systems, tools or resources, which are going to play an active role in our processes, and associate these elements at the activity level of the process model.

Documents are also essential artifacts of business processes; different documents serving different roles are being created, transferred, and utilized as a source of knowledge and information. These documents have to be taken into account throughout the complete BPM lifecycle, and this way also incorporated to the process models.

### III.2 Mapping of knowledge elements

As a last step of capturing the inspected processes, an overall semantic annotation is necessary to identify and connect knowledge elements of the processes at activity level.

In other words, we supplement the models with every available, explicit knowledge items at activity level.

This action is carried out in three steps:

- Domain experts and practitioners provide direct, structured knowledge items at the level of activities;
- As a second layer, an accurate, thorough description of the activity is recorded which can be treated as unstructured information. The information contained in underlying, non-structured form most undergo a semantic transformation to identify the knowledge elements or concept groups.;
The third layer relies on related documentation: guidelines, official procedures, best-practices, related legislation, etc. Acquiring knowledge element information is the most challenging in this case, the process can be aided with text-mining techniques.

Identified knowledge items can already exist in domain ontologies, in this case the mapping can be automated. In many fields of business areas general ontologies are available. If this is the case, it allows a more thorough concept building, and also results in more standardized outcomes adaptable as generalized solutions or industry level best practices. If there is no available pre-existing domain knowledge repository, the domain ontology specific to the examined organizational conduct is created. In both cases the domain ontology will hold all the knowledge item nodes that appear in processes. As we shall see in the thesis, nodes of the domain ontology hold the knowledge item description, which are represented by the classes of the domain ontology. In our institute's domain ontology structure, the classes Basic Concept and Knowledge area are used, depending on the nature of the knowledge items general or particular nature respectively. In case a pre-existing domain ontology is available, it must be imported to the modeling environments knowledge base. Concerning the modeling implementation of the semantic annotation, the first level knowledge items can be directly placed in Adonis EPC process models as information objects. The level of granularity set forth in our initial process models needs to be preserved. It has to remain unchanged, since this granularity applies to all other modeling dimensions as well. As a heuristic rule, we can say that the semantic annotation must not alter the initial process structure, except in cases where the alteration derives from structural and not annotational grounds.

III.3 Multilateral process views – process coupling via semantic transformations

The resulting complex process models contain interconnected, multilateral information on the following areas of the recorded processes:

- process structure, process hierarchy
- organizational structure, roles and responsibilities at activity level
mapped explicit knowledge

- IT architecture
- document structure

In order to make use of this holistic process-space, we need to apply semantic transformations to the models. The goal is to provide a machine-readable representation for further utilization in the form of ontologies.

Since the complex process models hold both process knowledge and domain knowledge, we have to conduct these transformations respectively.

Process ontology instances can be created automatically by XSLT transition. The process model hierarchy is represented in OWL format, and the additional structure of interconnected elements can also be transferred following a semantic annotation scheme. As far as my literature research extended, I have found no industry standards expressing the full requirements of such a process structure annotation, but an ad-hoc processing of such a markup is possible (Gábor, Kő, Szabó, Ternai, & Varga, 2013).

The creation of domain ontology also holds several challenges. The above described first level structured knowledge can be easily transformed into OWL ontologies, but the underlying levels need further elaboration. We are striving to provide automatic ways to create ontology knowledge elements or concept groups by means of applying text-mining techniques, but some extent of domain expert knowledge seems to be inevitable for transforming unstructured knowledge from the recorded processes. The PROKEX project intends to develop a reference architecture satisfying some aspects of automatic processing based on the multilateral process knowledge extraction of my thesis.

### III.4 Further Research Questions and Future Directions

My research area is dedicated to the challenges of knowledge extraction from business processes. I analyzed the opportunities of knowledge extraction based on the literature, my research background and practical experiences. I am proposing a solution to extract, organize and preserve knowledge embedded in organizational processes to enrich the organizational knowledge base in a systematic and controlled way. My other research problem is how to organize the extracted knowledge, what are the appropriate ICT solutions, environment for it. I reviewed theoretical foundations of related fields, like business process management, semantic technology, semantic business process
management and ontologies. Ontologies play a key role in semantic business process management, because they provide the structure for organizational knowledge. Therefore I discussed their background detailed in the literature review section.

I have identified the requirements in the business process modeling level to be able to use a complex process model as a base of creating the links between the process models and the domain ontology.

The novelty of the solution is based on the connection between process model and corporate knowledge, where the process structure will be extended with the annotation for knowledge structure. The resulting process and domain ontology duplet enables a higher level of automation for IT implementation and a wider range of possibilities for machine-reasoning.

The research outcome is going to be tested in a reference architecture, where the main goal is to create a supporting infrastructure capable to conduct multi-lateral searches especially for the purpose to support employees to easily acquire their job role specific knowledge, but there are wider areas for application.

The resulting knowledge repository holds multilateral information specifically for the viewpoints of organizational stakeholders and IT systems. The proposed solution support employees to easily acquire their job role specific knowledge, support IT departments to efficiently answer the challenge of changes to be applied at different processes, and knowledge engineers to have a better insight into the organizations’ knowledge environment.
IV. REFERENCES


V. REFERENCES OF THE AUTHOR

2012 Dr. Katalin Ternai, Mátyás Török
*Business process modeling and implementation in collaborating environments.*
International Journal of E-Business Development. Published by: THE WORLD ACADEMIC PUBLISHING CO., LIMITED.
ISSN: 2225-7411. 2012, Vol.2 No.4

2011 Dr. Katalin Ternai, Mátyás Török
*Semantic modeling for automated workflow software generation – An open model*

2011 Ternai, K. Török, M.
*A new approach in the development of ontology based workflow architectures.*
IEEE Xplore Print INSPEC Accession Number: 12289954 Date of Current Version: 10 October 2011.
ISBN: 978-1-4577-0772-8

2011 Török, Leontaridis
*Ontology based workflow architecture implementation for SMEs - case study*
eChallenges e-2011 Conference IIMC, Conference Proceedings Paul Cunningham and Miriam Cunningham (Eds) IIMC International Information Management Corporation, 2011

2011 Katalin Ternai, Mátyás Török
*A New Approach in the Development of Ontology Based Workflow Architectures.*
ISBN: 978-3-943024-04-3