

DYONIPOS: Proactive Support of Knowledge Processes

Silke Weiß¹, Josef Makolm², and Doris Reisinger³

¹ Federal Ministry of Finance, Austria, silke.weiss@bmf.gv.at

² Federal Ministry of Finance, Austria, josef.makolm@bmf.gv.at

³ m2n consulting and development gmbh, Austria, reisinger@m2n.at

Abstract: The success of knowledge-intensive organizations depends significantly on the degree of knowledge availability, knowledge transparency, knowledge structuring, and knowledge up-to-dateness. The research project DYONIPOS (DYnamic Ontology based Integrated Process OptimiSation) meets these challenges: DYONIPOS sets up a context sensitive, intelligent and agile assistant based on the development of semantic and generic knowledge discovery technologies [6]. The assistant supports the knowledge workers just in time and automatically with the currently needed knowledge, without additional work and violation of knowledge workers privacy. Furthermore an individual and a global process- and knowledge base is built-on.

This article is structured as follows: Section 1 addresses the relation between the applied approach and the challenge in e-Government, summarizes the aims of the research project DYONIPOS and emphasizes the motivation. In Section 2 the semantic and knowledge discovery technologies used are presented. The article concludes with the presentation of the use-case project, showing current results of the project.

Keywords: Knowledge management, Knowledge work support, Semantic technologies, Research project DYONIPOS, Use-case, Public administration

1. The Research Project DYONIPOS

E-Government means not only the use of information technology to improve the exchange of service and information with citizens or businesses. E-Government also means the use of information technologies to improve internal information, data and service quality. Public administration work is knowledge work par excellence, because information of governmental organizations is widely scattered and civil servants are confronted with an overload of information. An improvement can be generated e.g. through regulation and support of knowledge processes, which ensures the supply of administration processes with knowledge. The knowledge of an organization can be classified into three kinds of information: public domain knowledge, partly available knowledge and tacit knowledge. These three kinds of knowledge are mostly produced or at least organized by the knowledge workers and are required to carry out their processes. The present knowledge is often unstructured, intransparent and far-scattered, e.g. public domain knowledge is available for all knowledge workers. This knowledge is often stored online on the intra-or internet and accessible through search engines or knowledge databases. Partly available knowledge is only accessible for individual or specific groups of knowledge workers. This kind of knowledge is often stored on the employee's PC or the organization's server. The third kind of knowledge – tacit knowledge – is in the minds of employees and therefore only available to the owner of the knowledge. For a knowledge worker it is very difficult and time-consuming to find the adequate knowledge in the existing overload of information, which supports the achievement of his work steps. Furthermore if tacit knowledge is needed, the receipt of required resources depends on chance. The research project DYONIPOS meets these challenges. Its aim is to provide personal, agile and proactive support for the knowledge worker by means of proactive, context-sensitive knowledge delivery on the one hand and by suggestion of next process steps on the other hand. Furthermore DYONIPOS creates and continuously updates an individual as well as an organizational knowledge base. This knowledge base makes the organization's growing knowledge available. DYONIPOS provides all kinds of knowledge that has been released for the organizational knowledge database, assumed that the user has the right to access this knowledge. Knowledge not released to the organizational knowledge base is just available in a knowledge workers individual knowledge base. This knowledge is still only available for his owner. Furthermore DYONIPOS identifies tacit knowledge; the identified knowledge will be supplied through transfer of the name of the information owner. For ensuring data protection only the name of persons which are stored in a "whitelist" will be supplied. Only DYONIPOS users are stored in the "whitelist". Another aim is to support the process engineer through information about the recorded ad hoc processes, e.g. visualization of the workflow and process, landscape visualization of similar tasks and sub processes. A task is part of a set of actions which accomplish a job, problem or assignment. The process engineer can

use this information for improving standard processes or compiling statistics [6]. The DYONIPOS research project started on January, 2, 2006, and it will be completed by the end of the first quarter of 2008. The DYONIPOS research consortium consists of m2n consulting and development gmbh¹, Know-Center Graz², the Institute for Information Systems and Computer Media (IICM) of the Graz University of Technology³ as well as HP Austria⁴. Together they are developing the prototype DYONIPOS. The DYONIPOS research project is financed by the "semantic systems" program within FIT-IT, an Austrian research program provided by the Federal Ministry of Transport, Innovation and Technology (BMVIT)⁵. The proposal of the DYONIPOS project was awarded as the best proposal of the regarding call. In order to ensure the applicability of DYONIPOS, a parallel use-case project is carried out by the Austrian Federal Ministry of Finance (BMF).

2. Knowledge Discovery and Semantic Technologies

DYONIPOS is based on automatic and semiautomatic knowledge management methods and technologies e.g. knowledge discovery, semantic systems, knowledge flow analysis, and process visualization. The semantic technologies qualify as enabling technology to handle structured as well as unstructured parts and data of knowledge intensive processes [5]. DYONIPOS begins with the recording of all interactions between the user and his computer; these are the so called "events", e.g. mouse clicks or key strokes. Several events belonging to a logical unit are grouped together into event-blocks by using predefined rules. Similar event blocks form semantic sets are assigned to knowledge worker's tasks. The assignment initially has to be done by the user, but after the training phase in which DYONIPOS learns the classification features, a task will automatically be detected. The events, event-blocks and tasks are represented and stored with semantic technologies (RDF) [2].

2.1. Functionality of DYONIPOS

The discovery of work patterns and the delivery of relevant information just in time are the major functions of DYONIPOS [10]. To provide these functions

¹ <http://www.m2n.at>

² <http://en.know-center.at/>

³ <http://www.iicm.tu-graz.ac.at/rootcollection?timestamp=1188552118992>

⁴ <http://welcome.hp.com/country/uk/en/welcome.html>

⁵ <http://www.bmvit.gv.at/en/index.htm>

DYONIPPOS captures the knowledge work, discovers the tasks and processes and supports the knowledge worker with information.

The first challenge is the observation of knowledge worker's interactions with and reactions to the system and existing application data. This data is the so called low level sensor data on application and operating system level [5]. The second challenge is to develop adequate techniques to discover the work patterns and to support the users automatically with the information they need. The third challenge is to detect how knowledge workers can be effectively supported [5].

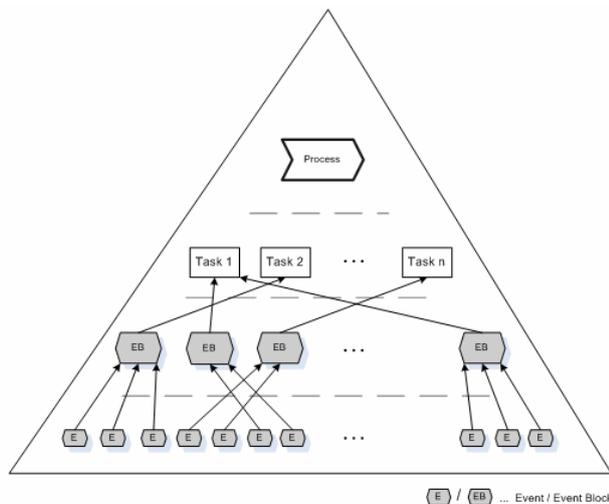


Fig 1: The semantic pyramid of DYONIPPOS: all activities e.g. events, event blocks, tasks and process are modelled semantically and stored by RDF.

The semantic pyramid mapped in figure 1 displays the four different layers of the semantically modelled and stored data. The first layer is the event layer, the second layer is the event block layer, the third is the task layer and the layer at the top of the pyramid is the so called process layer. Only one single user executes an event while in the execution of processes different users can be involved. Each layer of the pyramid provides a different representation of the data regarding the semantic quality [5]. The semantic quality is a measure of accordance of the information document in the model with the mapped domain.

Below the specific steps required to capture the workers patterns are described. For capturing the workers patterns a java tool called DYONIPPOS Task Recognizer has been implemented [5].

From user and system interactions to events...

DYONIPPOS starts with the recording of all events. Different sensors of the context observer module observe all interactions of the user with the desktop environment. DYONIPPOS uses a key logger program to record and log all recognized events [2]. The observed events e.g. keystrokes, mouse inputs or system events are

stored in the so called event log. This monitored data are the base for discovering the work patterns.

...from events to event blocks...

The next step is to reduce the immense quantity of data and to assign events to event blocks by filtering and by relation analysis. Through filtering analysis it is possible to select the irrelevant data – e.g. mouse movements – from the relevant data. Mapping of a set of events to an event block is possible by relation analysis. Currently generic based rules, application based rules and web browser based rules are applied for mapping events to event blocks [5]. Generic rules are based on windows title, process, and application name. A reason for assignment of events to the same event block is the same windows title, process and application name. Application dependent events like reading and forwarding of an e-mail result in the same event block through appliance of the application based rules. Identification of similarity between URL's is a sign for combination of the relevant events to event blocks. The implementation of further rules for mapping events to event blocks is easily extensible. An example for an event block is e.g. "start a file".

...from event blocks to tasks...

The utilized methods to learn task assignments are k-nearest neighbor classification, Support Vector Machines based on graph kernels (compare to 2.3) and the possibility of the classifier to learn task assignments from the user. Methods of extracting tasks are clustering based on similarity among content and structural features and the scatter/gather approach.

...from tasks to processes

The transition from tasks to processes emerges by combination of individual tasks which were conducted by a number of different knowledge workers. Similar methods as described in the part above will be used to learn the assignment from tasks to processes. To extract processes the application of the ProM Framework will be utilized (compare [5]).

2.2. Architecture of the DYONIPOS Task Recognizer

The DYONIPOS Task Recognizer consists of the context observer module, the task-recognizer module and the KnowMiner framework, which uses Apache Lucene, an open source java library. This tool supports the KnowMiner Framework by creation of the resource index. The context observer module is a C# program [7] and consists of different sensors. These sensors are able to send the observed low level operation system and application events based on XML in a specific event stream format to the context observer module. The server transfers the event streams to the task recognizer module. This module has the following purpose:

- Identification of tasks
- Revelation of information needs

- Discovery of used resources.

The KnowMiner Framework supports the mechanism of the task identification by classification and clustering algorithms. In addition the KnowMiner Framework is responsible for the resource indexing. The discovered information needs and resources, the observed events and identified event blocks, tasks and processes are semantically modelled, related and stored in the personal knowledge base, by RDF triple store (compare to [5]).

2.3. Classification of Task Instances via Graph Kernels

The objective of this chapter is to explain which techniques are applied for categorization of task instances to predefined task models and process instances to process models. The utilized techniques are Support Vector Machines. This technique allows the definition of the algorithm's view in data by selecting the kernel [7]. Starting with the event blocks which consist of a sum of events stored in a RDF repository represented as a graph. Classification of event blocks to tasks is possible through determination of similarity based on common content, but this method is not enough to combine related task instances to task models. For example a user is writing a document which includes knowledge about two different scientific fields. Content information is not enough to classify the corresponding task instances into the same task, because the content of the task instances is different. Additional structural information is required. Structural information is information about the type of an event block e.g. which resources have been used or which person is involved. The type of different entity blocks provides valuable information for data mining [7]. The similarity of event blocks is determined by using graph kernels [1] and kernel methods for graphs [9] which allow processing of content and structure [7]. The assignment of event blocks to a knowledge worker task is performed by applying Support Vector Machines [13].

2.4. The Semantic Technology

DYONIPOS is a modern information system which supports the users by proactive delivery of contextual information (resources) while the knowledge workers are doing their daily work. The application of ontologies is useful in such a system, because they ensure interoperability and the development of "new" knowledge. Furthermore, ontologies can be used for the unambiguous description of information resources. As a consequence, RDF is a key technology of DYONIPOS. Resource Description Framework (RDF) is an ontology language, a formal lan-

guage used to encode ontologies. An ontology represented in RDF language consists of:

- classes of entities (concepts) within a domain,
- properties of classes (relationships between the concepts),
- individuals belonging to classes and
- constraints on classes and properties [14].

An RDF-Statement or RDF-Triple is based on:

- the subject (resources),
- the predicate (property), and
- the object (value).

The subject identifies the described object, the predicate defines the data in the object, and the object represents the actual value [14]. All events, event-blocks and tasks described in section 2.1 are represented and stored by RDF-Triples [3]. This means all data extracted from metadata (e.g. from integrated applications), documents, presentations, e-mails etc. will be saved in a structured manner. For example, the DYONIPOS ontology consists of the concepts "Person", "Organization", "Document" and "Topic". An example for a concrete "Person" may be the employee John Q. Public. John Q. Public works at the Federal Ministry of Finance and has written some articles about semantic technologies. The circumstance described above results in the following concrete classifications: John Q. Public is an object of the concept "Person", the Federal Ministry of Finance is an object of the concept "Organization", all written articles are objects of the concept "Document", and the identified "Topic" is semantic technology. The following relationships exist between the objects: John Q. Public is employed by the Federal Ministry of Finance, John Q. Public is the author of some articles; John Q. Public deals with the topic of semantic technology. Further conclusions drawn are the following: John Q. Public is an expert in the topic of semantic technology and the Federal Ministry of Finance deals with the topic of semantic technologies. The newly learned knowledge – e.g. that John Q. Public is an expert in the topic of semantic technology – is a recognized resource of DYONIPOS. It should be mentioned that for privacy reasons only knowledge concerning persons who are registered DYONIPOS participants will be stored and supplied.

3. Use-Case Project

Parallel to the research project DYONIPOS the use-case project DYONIPOS is implemented in the Directorate General for Information Technology (DG-IT) of the Federal Ministry of Finance, Austria. Administration work is knowledge work

par excellence, because the flood of information is immense and the existing knowledge is widely scattered. The knowledge workers of the administration need the following additional know-how to tackle the daily work:

- Where is the relevant information stored
- How can this information be found
- How relevant is the delivered information

The challenge is to provide the administrative employees automatically with the information they need. Consequently the above mentioned additional know-how is made available by DYONIPOS. The relevant information is stored in the internet and intranet, the server of the organisation or the fixed disc of each knowledge worker. DYONIPOS finds the information through detection of mouse clicks and key-board inputs. DYONIPOS assigns this information to a proper task and delivers the needed knowledge automatically in terms of documents, internet sites or DYONIPOS links the knowledge worker directly to the documents of the applications (ELAK). DYONIPOS delivers the information ordered by relevance.

Other objectives are to support the employees of the DG-IT without creation of additional work by means of knowledge management and to ensure the privacy of the knowledge workers. DYONIPOS supports this challenge. Due to this fact DYONIPOS will be used for efficiently and effectively supporting the daily work of the individual employees in the DG-IT. The DYONIPOS task recognizer supports the employees with the necessary knowledge which is produced by semantic cross-linking of the relevant information from the existing repositories and processes. Additionally, DYONIPOS independently develops new relations between the sources of knowledge. This explains, for example, that the DYONIPOS task recognizer at the one hand supports the user by visualization of existing documents, files or websites etc. and displays the new generated information such as the name of the person who has the specific know-how. The ministry respectively fifteen employees support the research consortium. Together they work on the realization of the research results and they ensure the transformation of current scientific results to an easily useable software solution. The staff of the ministry shares its domain specific know-how with the research consortium, by supporting the development of DYONIPOS base technologies.

Initial interviews with employees have been carried out to obtain an impression of the kind of work and how this work is done. The results of these interviews provide the information which sensors should be developed and which events the sensors should observe. The researchers found out that the employees work especially with the following standard applications: Microsoft Office tools, Internet Explorer and the e-mail system Novell GroupWise. That is why a first research step was to develop sensors to observe events of these applications. In addition to the observation of these standard applications the final DYONIPOS prototype should record all electronic artifacts from the electronic record management system (ELAK), the file-system on the servers, the mail-system as

well as the specific application KOMPASS, a system to administrate persons, resources and authorizations.

The implementation of the use-case is structured in three evaluation phases. These tests serve as basis to support the improvement of the functions of DYONIPOS e.g. by a continuous refinement of the rules to assign events to event blocks. In the first test phase of the pilot software ten key-users took part to support the work of the researchers. The test occurred over a period of five weeks from April to May 2007. Main objective of the first test was to gather detailed information on the key users. The preliminary data collection has been included the user input and the work content. Further objectives are to test and evaluate the recording and analysis module of the DYONIPOS proactive assistant. In addition to the test and the evaluation key users had also the possibility to express concrete requests concerning the functionality and graphical user interface of DYONIPOS. Therefore the key users had the chance to take part actively in the design process of the system. At this stage the prototype DYONIPOS was stored and implemented on the local hard disc of each personal computer of the participating key users. A central storage on the server has not been carried out. The employees were introduced in the software handling and to do a manual assignment of event blocks to tasks. The collected information will serve as training basis for the DYONIPOS Task recognizer. Furthermore the key users had the possibility to evaluate the functions and to document suggestions for improvement. The results of this first test was stored in log files and documented in test protocols and questionnaires.

3.1. The Application DYONIPOS

Figure 2 shows the graphical user interface of the DYONIPOS task recognizer window. Different flags allow the navigation between the different supplied resources and functionalities. At the screenshot the flag “Übersicht” is opened. At this flag all recorded events (e.g. mouse clicks and movements as well as keyboard input and system interactions) and event blocks are displayed at the top. Central the so called “InformationNeeds” are mapped. Only after detecting of events DYONIPOS starts with the identification of information needs and suggests next tasks, resources and experts. All identified resources were displayed in the bottom field. These are resources of the individual as well as of the organizational knowledge base. Furthermore the key user has the possibility to search actively for information in the iteratively generated resource repository by using the search field. This search field is displayed in the screenshot at the bottom next to the magnifying glass. At the flag “Experte” the name of the experts can be displayed or the specific topic, which the organization deals with. At other flags the adjustment of the DYONIPOS functionalities can be carried out, e.g. the deletion of knowledge, which is stored at the organizational knowledge base.

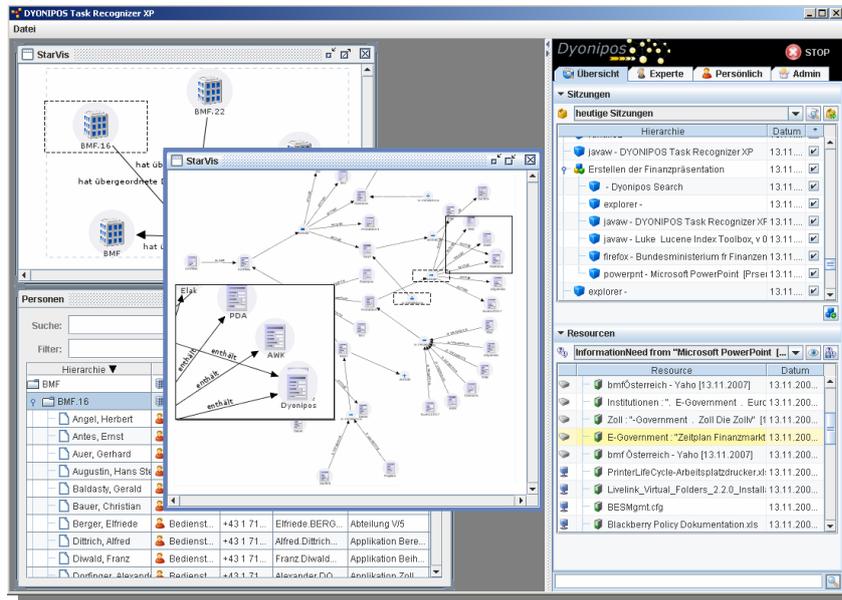


Fig. 2: Screenshot of the DYONIPOS Task Recognizer

3.2. The Results of the First Test Phase

Through evaluation of the logfiles, questionnaires and the test protocols it was possible to derive the following information and operating figures about the key user and about the DYONIPOS Task Recognizer: Typical activities which a key user implements are: project work, description and modelling of process, writing of protocols, compiling of statistics, participation in discussions, executive functions, searching in the Internet, providing of presentations, etc.

- A key-user performs the following activities without PC support: mental work, meetings, telephone calls and face to face discussions with other employees.
- A key-user uses the following tools to do the daily work: text processing tools, spreadsheet programs, presentation programs, web browsers and file explorer, e-mail and calendar, the electronic records management system (ELAK), different SAP systems, specific application programs, database and data mining tools, different information retrieval systems (Google or other special search applications) and occasionally image editing tools.

- A key-user often interrupts its activities because of: telephone calls, e-mails, individual spontaneous requests of other employees but also planned meetings.
- A key-user usually uses the following information sources: the Internet (basically Google), the documents stored on the hard disc on his own PC, the server-side data collection, the existing documents on the mailbox, paper based documents, meetings as well as vocational and private social networks.
- A key-user uses the following programs and IT applications to search for information: Internet search machines (Google and Wikipedia), SQL navigator, Windows search on PC and server, as well as the search function of the e-mail system.
- A key-user implements approximately 50% up to 90% of his/hers complete work by computer support (mean value: 74.4%).
- A key-user executes 25% to 90% of his/hers communication via computer (mean value: 60.6%).
- A key-user executes at least one task and maximal 15 tasks per day by computer (mean value: 7.8%).
- A key-user works parallel at least at 2 tasks up to 8 tasks (mean value: 5)
- A key-user daily spends between 2.1% and 60% of his/hers working time by accessing information (mean value: 16.1%).

A basic result of the evaluation is that key-users always work on several tasks at the same time. This information represents a challenge for DYONIPOS, because it is an objective of DYONIPOS to provide just-in-time information based on the context. Furthermore we found out that a key-user use different searching tools and search in very heterogeneous sources. An objective of DYONIPOS is to support the work of the user by proactive and context-sensitive information delivery. DYONIPOS searches for information in different repositories to support the user. DYONIPOS implements the function of a searching tool and creates cross-links between the context of different repositories to deliver existing information and new generated information. Using DYONIPOS the knowledge workers receive transparency over the existing sources of information. DYONIPOS gives additionally references about the relevance of the found search results, which includes all currently available information. The parallel implementation of the funded research project and the use-case project makes it possible to exchange ideas between research and practice constantly; this is useful for both projects. Furthermore the inclusion of all stakeholders [8] such as researchers, users, IT experts and also the staff council in the development process assures that the results of the research project DYONIPOS can and will be transformed optimally and in real-time into a practical application.

3.3. The Second Test Phase

The second test phase was started in Februar 2008. A fundamentally improved version of the prototype DYONIPOS which establishes an organizational knowledge base, with new functionalities and which also includes artifacts stored on server as well as electronic records is proved now. In the second test the former manual assignment of event blocks to tasks work automatically. The key-user just observes this assignment, by doing corrections of wrongly assigned event blocks and by confirmations of correctly assigned event blocks.

The second test has shown the assurance that it is not necessary to identify processes. It is enough to know on which task a knowledge worker works. Furthermore the agility of the system needs to be adequate. If the system is too agile, too much information will be found. This costs time and processing power. But if the system is too less agile, DYONIPOS finds too less ressources. The tuning of the system is a necessity.

3.4. Next Steps

Finally the third test phase starts approximately in April 2008. In this testing phase key-users will test the whole functionalities of DYONIPOS. This test will take 75 days. This third test will be closed with an evaluation and documentation of the use-case results in a final project report.

In 2008 the whole DG-IT or rather all 180 employees will take part on a final test of the prototype DYONIPOS. A Final evaluation will be done after one year of practical experience. Documentation and evaluation of this final test provides the basis for the decision whether DYONIPOS will be used in the DG-IT further on. In the same way the decision will be made whether DYONIPOS should be advanced to commercial tool or not.

References

1. Gärtner, T. A survey of kernels for structural data. SIKGKDD Explor. Newsl., ACM Press, 5, 49/58, 2003.
2. Kröll, M., Rath, A., Granitzer, M., Lindstaedt, S., Tochtermann, K. (2006). Contextual Retrieval in Knowledge Intensive Business Environments. GI-Workshop Information Retrieval 2006.
3. Kröll, M., Rath A., Weber, N., Lindstaedt, S., Granizer, M. (2007). Task Instance Classification via Graph Kernels, Mining and Learning with Graphs (MLG 07). Florenz, Italy.
4. Maier, R. (2005). Modeling Knowledge Work for the Design of Knowledge Infrastructures. Journal of Universal Computer Science 4, 11, 429-451.

5. Rath, A. (2007). A Low-Level Based Task And Process Support Approach for Knowledge-Intensive Business Environments. In Proceedings of the 5th International Conference on Enterprise Information System Doctoral Consortium (DCEIS 2007), 35-42. Madeira, Portugal.
6. Rath, A., Kröll, M., Andrews, K., Lindstaedt, S., Granitzer, M., Tochtermann, K. (2006). Synergizing Standard and Ad-Hoc Processes. Proceedings of 6th International Conference on Practical Aspects of Knowledge Management, LNCS Springer, Vienna (Austria).
7. Rath A., Kröll, M., Lindstaedt, S., Granitzer, M. (2007). Low-Level Event Relationship Discovery for Knowledge Work Support. 4th Conference on Professional Knowledge Management, ProKW2007 Productive Knowledge Work: Management and Technological Challenges. Potsdam, Germany, GITO-Verlag, Berlin.
8. Makolm, J., Orthofer, G. (2007). Holistic Approach, Stakeholder Integration and Trans-organizational Processes: Success Factors of FinanzOnline. In: E-Taxation: State & Perspectives, E-Government in the Field of Taxation: Scientific Basis, Implementation Strategies, Good Practice Examples, p.389-402, ISBN 978-3-85499-191-5, Schriftenreihe Informatik 21.
9. Schölkopf, B., Smola, A.J. (2000). Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond. MIT Media Laboratory.
10. Tochtermann, K., Reisinger, D., Granitzer, M., Lindstaedt, S. (2006). Integrating Ad Hoc Processes and Standard Processes in Public Administrations. Knowledge transfer across Europe: 4th Eastern European eGov Days and 5th eGov Days. OCG Schriftenreihe Band 203, Wien.
11. Van der Aalst, Wil. M. P., De Michelis, G., Ellis, C.A. (1998). Workflow management: Net-based concepts, models, techniques and tools (wfm'98). In Computing Science Report 98/7.
12. Van der Aalst, Wil. M. P., Weske, M., Grünbauer, D. Case handling: a new paradigm for business process support. Data Knowl. Eng. 53(2), page 129-162, 2005.
13. Vapnik, V.N. (1995). The nature of statistical learning theory. Springer Verlag New York, Inc.
14. W3C webpage: <http://www.w3.org/TR/owl-ref/#acknowledge>, Accessed 28 January 2008.