Pavol Návrat
Pavol Bartoš
Mária Bieliková
Ladislav Hluchý
Peter Vojtáš (Eds.)

Tools for Acquisition, Organisation and Presenting of Information and Knowledge

Research Project Workshop
Bystrá dolina, Nízke Tatry, Slovakia,
September 29-30, 2006, Proceedings

Slovak University of Technology in Bratislava
Institute of Informatics, Slovak Academy of Sciences
Pavol Jozef Šafárik University in Košice
Softec, Ltd.

Proceedings in Informatics and Information Technologies

Tools for Acquisition, Organisation and Presenting of Information and Knowledge

Editors

Pavol Návrat
Institute of Informatics and Software Engineering, Faculty of Informatics and Information Technologies, Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia

Pavol Bartoš
Softec, Ltd.
Kutuzovova 23, 831 03 Bratislava, Slovakia

Mária Bieliková
Institute of Informatics and Software Engineering, Faculty of Informatics and Information Technologies, Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia

Ladislav Hluchý
Institute of Informatics, Slovak Academy of Sciences
Dúbravská cesta 9, 845 07 Bratislava, Slovakia

Peter Vojtáš
Institute of Computer Science
Faculty of Science, University of P.J. Šafárik
Jesenná 5, 040 01 Košice, Slovakia

CR Subject Classification (1998): D.2, H.1, H.3, H.5, I.2, I.7, K.6

ISBN 80-227-2468-8

Published by Vydavateľstvo STU, Vazovova 5, Bratislava, Slovakia

Executive Editor: Tomáš Horváth, UPJŠ Košice
Copy Editor: Róbert Novotný, UPJŠ Košice
Cover Designer: Peter Kaminský

Visit project web site: http://nazou.fiit.stuba.sk
Project acronym: NAZOU (Nástroje pre získavanie, organizovanie a udržovanie znalostí v prostredí heterogénnych zdrojov, Tools for Acquiring, Organising and Presenting Information and Knowledge in an Environment of Heterogeneous Information Sources)

Printed in Košice

Preface

The project NAZOU (Nástroje pre Získavanie, Organizovanie a Udržovanie znalostí v prostredí heterogénnych zdrojov, Tools for Acquiring, Organising and Presenting Information and Knowledge in an Environment of Heterogeneous Information Sources) is funded by the Slovak State Programme of Research and Development Establishing of Information Society. The collaboration among four research partners in this project began in late autumn of 2004. It aims at developing methods and tools, based on researching principles and models, for "acquiring, organising and presenting information and knowledge in an environment of heterogeneous information sources". Motivation for this research comes from the nature of the current World Wide Web, strengthened by its prospects for the future.

Amount of accessible information and knowledge grows at an unprecedented pace. Its extent, quality and accessibility changes also due to the world-wide use of the Internet. The Internet and its services (e.g. World Wide Web or electronic mail) can be used as a very appropriate environment for the research of new ways of knowledge acquisition from the heterogeneous sources, organizing, validation, evaluation and maintenance of actual knowledge. Knowledge originates in information, which in turn is just data interpreted in a certain context. Internet forms a distributed environment for the heterogeneous sources of information. The distribution of information is important from the point of view of information accessibility, while heterogeneity is the key feature of the documents represented and presented on the Internet.

When looking for certain information we are often overwhelmed by a huge amount of data of various kind and quality. For instance, in the case of e-mail, an effective filtering of messages would be very helpful. Many search tools provide too extensive and irrelevant answers to user queries. On the other side, these tools are not able to provide information, which is on the Internet, but is represented in a form that is difficult to process. In principle, search tools and services can be devised as universal – their main concern is to search, index and organise potentially all the sources found on the Internet, or as specialised tools and services – they focus on a certain area of interest and present such information.

It is becoming clear that approaching to the stage when virtually "everything is on the Web" may not imply automatically that all the information becomes accessible or retrievable. Here is the role of new methods that would help people retrieve "the right information in the right form at the right time", as Nigel Shadbolt writes in his introductory words to the 2005 Volume of Selected Papers of the AKT project. This may require devising new models of information (and knowledge) representation, and it is this context that we see Semantic Web as one of the possible approaches.

The research on a theme like this one is challenging and exciting. The theme is very complex and its various social, legal and ethical impacts make it a multi-

disciplinary one. Researchers cannot investigate all the issues, but they should at least be aware of the more important ones. For example, the above formulation does not imply that hypothesizing that "everything is on the Web", even if it is obviously not meant too literally, means our accepting that it is necessarily a desirable state (c.f. e.g. the notion of privacy protection). Similarly, the concept of filtering should raise serious concerns.

In the project, we concentrate on a broad range of topics as characterised by its title. We, the researchers, come from four institutions, all of them sharing a general mission of contributing to the progress in various fields of Informatics and Information Technologies, but at the same time all of them different. The researchers are from Faculty of Informatics and Information Technology of the Slovak University of Technology in Bratislava – the main contractor of project, then from Institute of Informatics of the Slovak Academy of Science, from Institute of Informatics of the Pavol Jozef Šafárik University, Košice and from Softec, Ltd., a private enterprise in the IT sector. Thus we have a genuine blend of academia and industry on board. We have among us the first Faculty in Slovakia that was established solely and completely for a broad range of fields in Informatics and Information Technologies. It is embedded with a university of technology. But we have among us also an Institute of Informatics embedded within a more "classical" university, and also an Institute embedded with the Slovak Academy of Sciences. I find it very important to have with us also one of the leading companies in the software industry in the country.

The collaboration among four different and also geographically distant institutions poses managerial challenges that can be aided greatly by the recent information and communication technologies, but that have benefited even more from the spirit of team work among the partners. I wish to thank for it to all the researchers, but specially to Prof Peter Vojtáš, Dr Ladislav Hluchý, Ing Pavol Bartoš, who are project leaders at the respective institutions mentioned above, and last but not least to my deputy Prof Mária Bieliková.

The project is set for three years. This volume contains original papers describing our first results in investigating and developing methods and tools, achieved during the first half of the project. The papers were reviewed by the editors and presented at the NAZOU workshop 2006, which took part within the ITAT 2006 Information Technologies – Applications and Theory Conference in chalet Kosodrevina in Bystrá dolina, Slovakia. It should be noted, however, that these are by no means the only papers we have published on the research. We have been preparing contributions to journals or international conferences, but the papers in this volume constitute a collection that is uniquely comprehensive.

The volume contains 31 contributions. It starts with an introductory paper that describes the overall aims, structure and principal approaches of the project. Then, the remaining papers can be grouped roughly into four parts.

- Methods for acquisition of data and offers
- Methods for analysis and organization of data and offers
- Methods for presentation of job offers and user modelling
- Advances in software technology

Among them, the second theme has been focused in this phase of our research most extensively, resulting in the highest number of papers, concerned with methods and tools for ontological data organizing and analysing, clustering, indexing, categorising and searching, and text processing. Across all the parts, we have also papers dealing with attempts to form "chains" of tools as important steps towards creating a prototype application.

I wish to thank all my fellow editors, but most of all to all the authors for their devotion to the project and their enthusiasm in preparing these papers. To readers, I wish you enjoy our first volume that shows what we have been able to achieve already in the second year of our project.

August 10, 2006

Pavol Návrat
project leader

Acquiring, Organising and Presenting Information and Knowledge in an Environment of Heterogeneous Information Sources

Pavol Návrat[1], Pavol Bartoš[2], Mária Bieliková[1], Ladislav Hluchý[3], and Peter Vojtáš[4]

[1] Institute of Informatics and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia
{navrat,bielik}@fiit.stuba.sk

[2] Softec, Ltd.
Kutuzovova 23, SK-831 03 Bratislava, Slovakia
palo.bartos@softec.sk

[3] Institute of Informatics, Slovak Academy of Sciences,
Dúbravská cesta 9, SK-845 07 Bratislava, Slovakia
hluchy.ui@savba.sk

[4] Institute of Computer Science, Faculty of Science, University of P.J. Safarik,
Jesenná 5, SK-040 01 Košice, Slovakia
Peter.Vojtas@mff.cuni.cz

Abstract. This paper describes aims, progress and some results of a research conducted in a project that is aimed at devising ways of processing of information and knowledge in a heterogeneous environment, in particular at acquiring, organising and presenting information and knowledge from the Web. Important part of the project are pilot applications. Their main purpose is to test the devised methods and tools as extensively and realistically as possible. From the two pilot applications that are planned, we have already developed the first one, devoted to offers and requests on the job market. It is designed as a framework that allows incorporating most of the tools. They show usefulness in supporting the user when looking for new job positions, when mediating services connected with employment or when selecting an applicant for a job for various job positions.

1 Introduction

The problem of acquiring and presenting information and knowledge from the Web is a topic of intensive research interest nowadays. Many research groups at various places attempt to tackle it from different perspectives. Here we can

mention for instance the project AKTORS supported by the British government (www.aktors.org), projects supported by European Union, i.e. On-To-Knowledge (www.ontoknowledge.org), REWERSE (rewerse.net), Knowledge Web (knowledgeweb.semanticweb.org). Project SIMILE (simile.mit.edu) is a result of cooperation of a consortium built from W3C, MIT Libraries and MIT Computer Science and Artificial Intelligence Laboratory. There are also other projects on the topic, some of them going on in Slovakia.

When talking about research activities concerning this area in Slovakia, we must admit that they are rather fragmentized. It has been the project "Tools for knowledge acquisition, organization and maintenance in the environment of heterogeneous information sources" funded by governmental program of research and development "Building information society" that finally provided supportive conditions for forming a group of several dozens of researchers, mostly young ones, which started to cooperate very intensively. These researchers are from Slovak University of Technology, Faculty of Informatics and Information Technology – the main contractor of project, then from Institute of Informatics, Slovak Academy of Science, from University of Pavol Jozef Šafárik, Košice and from Softec, Ltd., a private enterprise in the IT sector. The project is significant from two points of view:

- research level as the topic is very up-to-date, and
- level of integration of research activities in Slovakia, facilitating cooperation of a relatively large group of researchers.

2 Project Overview

2.1 Research areas

Coming out from the main research objective of the project, which is improvement of providing actual and relevant information from the Web, the project is focused to the investigation of new ways of information and knowledge processing in the environment of heterogeneous sources, especially with imperfect and vague information. In the coincidence with the described goal, the research is oriented towards these areas:

- models of heterogeneous environment (uncertainty, systems for modelling imperfect information, models of application domain, user models, context models, navigation models, metadata and ontologies, multilanguage approach, multiagent systems,
- knowledge acquisition (information recommendation, obtaining of user or environment model, special languages for flexible query, ontology creation),
- knowledge organization (ontologies, various inductive methods, clustering, indexing, small world networks),
- knowledge presentation (adaptive navigation, adaptive content presentation).

2.2 Application domain

Results of our research are being verified on pilot applications with job market as an application area. This allows handling with information about job market much more effectively – when looking for new job positions, when mediating services connected with employment or when selecting an applicant for a job for various job positions.

Social importance of this application is undoubtedly high and there is no need to argue about it. It is a typical problem, and its nature is heterogeneous in more dimensions. The sources, which contain an offer are different – ranging from global enterprises to small local companies. The way of presentation will also vary. In the European Union we have more than a dozen of official languages, and there are naturally many more languages that are used on the Web. Job offers in these languages can be just as useful as the ones written in official languages. Other dimensions of heterogeneity are for instance profession, region etc. Professions may have different customs, ways of expressing, idioms and patterns of describing what is offered and what is sought. Similarly, natural cultural differences among regions can influence the meaning of offers and requests.

In the "jungle" of many data on the Web, one has to search, make accessible and present to the user that information, which best suit his/her need or preferences. Informations provided must be up to date, and this aspect should be watched by an agent working in the background. Finally, the most important aspect of the whole project is the degree of improvement of the access to the information and knowledge about job market, which helps the user find the most suitable job for him. The main task of the pilot application is to search effectively information and knowledge about job offers and to mediate them to the potential applicants for these positions, and thus:

- improving the process of selection of suitable job for the applicants by more effective processing of job offers,
- increasing chances of the applicant to find the job, which best suits his requirements and possibilities by following and adapting the retrieved information, also by providing the sources which were hidden so far,
- enabling employers to find appropriate job applicants which fulfil their requirements.

3 From the data on the Web to information and knowledge for the user

It is quite difficult to describe all aspects of the project in a few paragraphs (reports on overall progress at various preliminary stages were published c.f. [17, 18, 19]). Some view could be provided by describing the designed software tools and a transformation from data acquired from the Web to information and knowledge presented to the user performed by the tools. The tools realise a sequence of data acquisition and processing, so that they operate on various levels of semantics understanding of individual sources. The sequence follows in successive steps

form acquiring documents, which presumable contain job offers from Internet, through identifying the documents in which offers are incorporated, offers extraction, their analyzing and organizing up to their personalized presentation to the user (see Figure 1). This could be characterized as transformation of the part of the Web to the Semantic Web, where existing documents are transformed to such a representation, which augments at the highest level presented information by semantic concepts utilised with advantage by their automated processing [7].

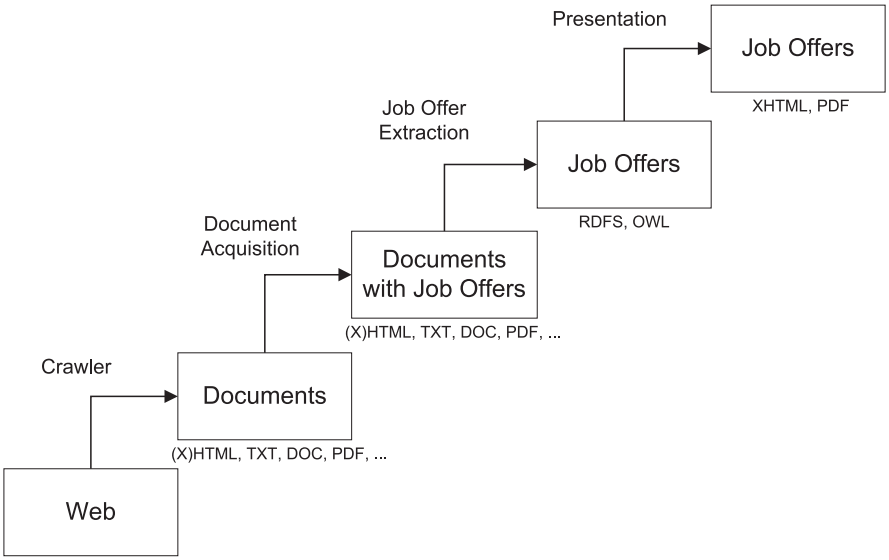

Fig. 1. Transformation data from the Web to presented information.

Design of software components for tools was influenced also by the fact that software systems in general and especially those concerned with hypermedia are developed without special attention to organized reuse, which is one of the main conditions of automating software development. Ad hoc reuse has not approved to be useful in practical applications. Development for reuse, which is the basis for domain-oriented approaches is inevitable [21, 25]. These approaches focus their attention to the whole domain of application, which must be appropriately constrained. Architecture and components for the whole family of software systems in the given domain create outcome of this development process, and the specific software systems are being developed in this way. This area is also the subject of research in our project.

Individual tools are designed employing two-layer architecture. Each tool is split into *domain independent layer* and *domain dependent layer* (see Figure 2). Domain independent layer includes an implementation of particular method for information acquisition, processing or presentation. Domain specific layer pre-

pares inputs for the tools and interprets outputs of the tools in particular application domain. Each tool defines in domain specific layer specific wrapper. Domain independent part does not understand the wrapper, the wrapper posses requests to this layer instead.

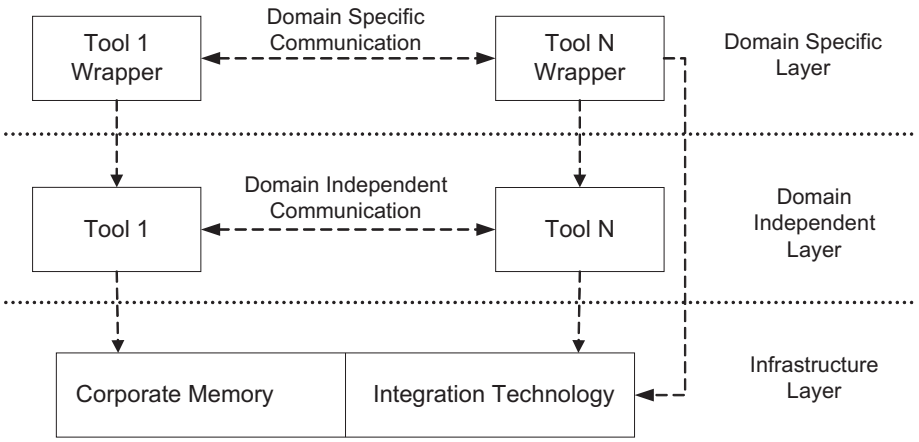

Fig. 2. Two-layer software tools architecture.

Each tool is designed in such a way that it can work autonomously to solve the specific problem. The tools are loosely coupled through data stored in the corporate memory. Corporate memory holds and manages data and knowledge created and processed by tools [5].

Corporate memory is accessible for other components using relevant client. The core of the corporate memory is running as XML-RPC server and other components can call relevant client method via XML-RPC. Corporate memory is organised into three layers (Figure 3):

- *physical layer* (file system, database system, and ontological models),
- *manipulation layer* (access to the stored data and information) and
- *interaction layer*.

Semantic part of the corporate memory is responsible for providing user interfaces for querying and manipulating the corporate memory content enriched by semantics as well as providing the physical back-end for persistent and transient storage of the semantics. The semantic model of the Web content is represented using an ontology (the OWL DL language). The corporate memory semantic part has two parts: the core interface (transparent access to the underlying knowledge repositories and reasoners) and the OntoClient interface (defines the possible interactions between the components of the system and the semantic part of the corporate memory).

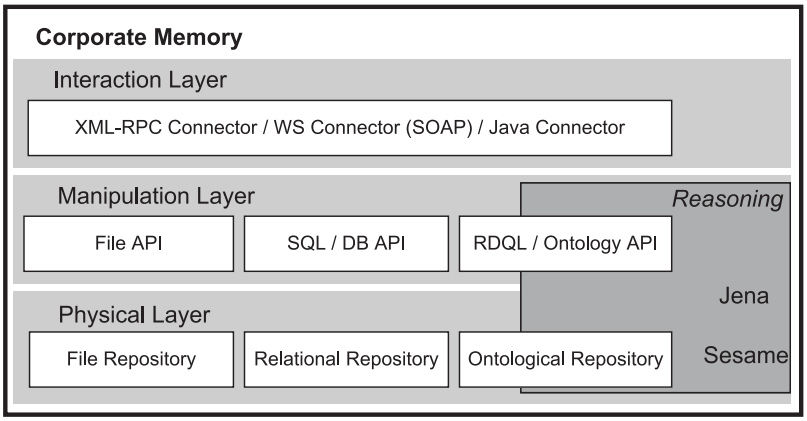

Fig. 3. Corporate memory structure.

Ontology is an important aspect to provide a semantic framework to describe application domain. Although ontology enables processing knowledge and data, the most important role of the ontology is in defining sharing meaning, emergence and discovery of gaps and for improving tacit knowledge transfer.

The relational database management part of the corporate memory is designed with virtualization concept making actual database system and database connection object transparent to the client applications.

The part of corporate memory dedicated to file management provides a way for manipulating the file storage using unified application interface, making actual physical file storage transparent to the user or application. In the current implementation, corporate memory's file storage is realized as a directory subtree of a file system directory tree. File management part of the corporate memory consist of core operations implementation and the client toolkit. Client toolkit can be configured to access the corporate memory's file storage through local Java API with XML-RPC call or through Web Service interface, which is realized by OGSA-DAI framework.

4 Proposed methods

As mentioned above, we have supposed following sequence of data processing when designing our methods and tools for knowledge acquisition, organization and presentation:

- primary data on the Web,
- acquired documents,
- documents, in which relevant data with respect to application domain (in our case job offers) are included,
- job offers extracted from documents with job offers identified and presented for a specific user or group of users.

The intention was to design methods (realized as software tools) in such way, that they would correspond with presented ideal sequence of data processing.

Proposed methods are divided into three groups:

1. methods for data and job offers acquisition from the Web,
2. methods for analysis, organizing and maintenance data and job offers,
3. methods for personalized presentation of job offers including methods for developing user models.

The research is oriented towards support of organizing of data and knowledge and their categorization by means of ontologies. Methods for knowledge acquisition are based on data mining, knowledge discovery, case-based reasoning, rule-based reasoning, and some other appropriate approaches.

In the course of research and design of presentation methods we have focused our attention on enrichment of information space of heterogeneous sources by elements of adaptation to user and/or environment (in context of the Semantic Web). We pay our attention especially to adaptive content presentation and adaptive navigation in hyperspace [2]. Our intention is to present the user personalized information, i.e. information that is for him relevant and in the way, which fits best to his demands.

Following sections give summarization of proposed methods. The methods are realised by software tools (see Figure 4 for a view on tools currently being developed). To fulfil the project aim the tools need to be integrated. Currently there exist several chains of directly collaborating tools. We mention here several examples:

- chain for job offers acquisition,
- chain for semantic annotation,
- chain of the tools for ranking offers,
- chain of the tools for user modelling development,
- chain of the presentation tools.

Moreover, some tools are in a position of serving for several other tools, e.g., the DocConverter tool serves for several tools which process text documents, indexing tools serve for tools aimed at searching and clustering, related words identification is used by several tools for offers analysing as well as concept comparing. Presentation tools are supposed to use the results of analysis performed by other tools (e.g., clustering or ranking results).

4.1 Methods for acquisition of data and offers

- a method for the retrieval query relevant resources on the Internet realised by the RIDAR tool (Relevant Internet Data Resource Identification),
- a method for downloading web documents realised by the WebCrawler tool [9],
- a method for the estimation of relevancy of web documents according particular application domain realised by the ERID tool (Estimate Relevance for Internet Documents),

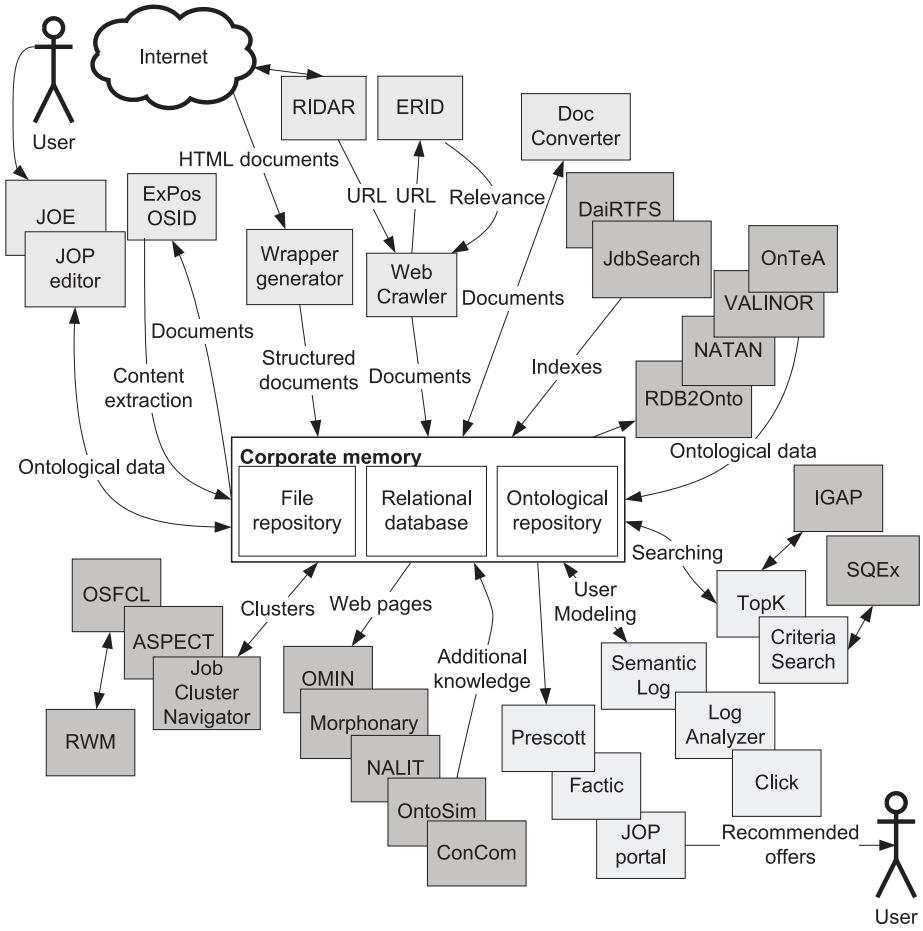

Fig. 4. Overview of tools.

- a method for the extraction of individual offers realised by the OSID tool (Offer Separation for Internet Documents),
- a method for the conversion documents to the plain text format realised by the DocConverter tool (Document Conversion),
- a method for distributed manual annotation of texts containing job offers realised by the JOE tool (Job Offer Editor),
- a method for job offer extraction from the web page realised by the ExPoS tool (Job Offer Extraction from the Web Page),
- a method for building data acquisition portals based on ontological representation using CRUD pattern (Create-Retrieve-Update-Delete) realised by the JOP tool (Job Offer Portal) [3],
- a method for the extraction of data in generated structured output from wrapped web pages realised by the Wrapper generation tool [22].

4.2 Methods for analysis, organization and maintenance of data and offers

Ontological data organizing and analysing

- a method for the annotation of text documents according domain ontology realised by the OnTeA tool (Ontology-Based Text Annotation) [15],
- a method for determination of similarity between ontology concepts and instances realised by the OntoSim tool (Ontology-based Similarity Determination),
- a method for comparing domain ontology concepts realised by the ConCom tool (Concept Comparer) [1],
- a method for creating and managing ontologies realised by the tool NATAN,
- a method for a method for mapping relational database data to ontology individuals realised by the RDB2Onto tool.

Clustering

- a method for providing an integrated view of job offers, which are visualized as network of companies and job positions realised by the Job Cluster Navigator tool [8],
- a method for finding out a hierarchy of clusters of domain words using fuzzy conceptual clustering realised by the OSFCL tool (One-Sided Fuzzy Concept Lattice Clusterer) [14],
- a method for probabilistic documents clustering realised by the Aspect tool [20].

Indexing

- a method for fulltext indexing of text documents and for retrieval realised by the tool DaiRTFS (Data Access and Integration Rich FullText Search),
- a method for full text indexing and querying over document collection with a support of relational database realised by the JDBSearch tool [16].

Categorizing and searching

- a method for finding user dependent top-k items realised by the Top-K aggregator tool [10],
- a method for rule induction for monotone graded classification of job offers to classes of appropriateness realised by the IGAP tool (Induction of Generalized Annotated Programs) [12, 13],
- a method for determining important markup elements in web pages by their visual importance realised by the Omin tool (Valuating Elements of Web Pages)
- a method for search of offers with criteria realised by the CriteriaSearch tool,
- a method for searching concepts within the ontology using heuristic inferences and logical deductions realised by the SQEx tool (Semantic Query Expansion),

- a method for searching words related to the given word realised by the RWM tool (Related Words Manager),
- a method for the normalization of values represented in various ranges realised by the VALINOR tool (Values Normalization).

Text processing

- a method for the identification of natural language in which the documents are written realised by the NALIT tool (Natural Language Identification Tool) [24],
- a method for determining all flexions of a word in Slovak language realised by the Morphonary tool.

4.3 Methods for presentation of offers and user modelling

- a method for logging user actions on the client side realised by the Click tool [1],
- a method for estimation of user characteristics from user activity logs realised by the Log Analyzer tool [4],
- a method for effective logging and integration of semantic events from various sources realised by the Semantic Log tool [1],
- a method for effective browsing of large structured information spaces realised by the FACTIC tool (Faceted Semantic Browser) [23],
- a frame for presentation of information represented by an ontology realised by the Prescott tool (Presentation-Cocoon-Ontology) [6].

5 Conclusions

One of the main outputs of the project was the Web portal, which serves for verifying research results. This is valuable especially in experimenting with new ways of knowledge processing in environment of heterogeneous information sources. Problem domain is the area of job market. This portal has been created by means of tools for knowledge acquisition, organization and maintenance. In this way we have practically approved methods and techniques designed within the research project

The main advantages of our approach can be summarized as follows:

- use of ontology (reasoning, document similarity),
- active search of information about job opportunities (automatic detection of job offers, checking company portals),
- concentrating job offers from heterogeneous environment,
- robustness from the point of view query formulation.

Original design of software architecture for intelligent knowledge management as well as original approach to component development for tools that acquire, process and maintain information are also important achievements of this

project. These tools are in various development phase, some are in stage of development, some in stage of specification, others in stage of design and some of them already in stage of verification. Their integration into software architecture is gradual. We develop in parallel the experimental ontological base of job offers assuming that this ontology will be usable in other similarly oriented projects as a testbase for experimental evaluation of various methods for analyzing and organizing data.

Acknowledgements

This work was partially supported by the Slovak State Programme of Research and Development "Establishing of Information Society" under the contract No. 1025/04.

The authors would like to thank all active researchers in the project NAZOU (Nástroje pre Získavanie, Organizovanie a Udržovanie znalostí v prostredí heterogénnych zdrojov, Tools for Acquiring, Organising and Presenting Information and Knowledge in an Environment of Heterogeneous Information Sources) for their excellent work.

References

1. Andrejko, A., Barla, M., Bieliková, M., Tvarožek, M. Tools for user characteristics acquisition. In *Proc. of Datakon'06*, 2006. Accepted.
2. Andrejko, A., Barla, M., Bieliková, M. Ontology-based user modeling for web-based information systems. In *15th Int. Conf. on Information Systems Development, ISD'06*, Budapest, Hungary, Springer. Accepted. (2006)
3. Barla, M., Bartalos, P., Sivák, P., Szobi, K., Tvarožek, M., Filkorn, R. Ontology as an Information Base for Domain Oriented Portal Solutions. In *15th Int. Conf. on Information Systems Development, ISD'06*, Budapest, Hungary, Springer. Accepted.
4. Barla, M. Interception of user's interests on the web. In V. Wade, H. Ashman, and B. Smyth, editors, *4th Int. Conf. on Adaptive Hypermedia and Adaptive Web-Based Systems, AH'06*, Dublin, Ireland, 2006. Springer, LNCS 4018, pp. 435–439.
5. Ciglan, M., Babik, M., Laclavik, M., Budinska, I., Hluchy, L. Corporate memory: A framework for supporting tools for acquisition, organization and maintenance of information and knowledge. In *9th Int. Conf. on Inf. Systems Implementation and Modelling, ISIM'06*, Prerov, Czech Republic, 2006, pp. 185–192.
6. Bieliková, M., Grlický, V., Kuruc, J. Framework for information presentation represented by ontology. In P. Vojtáš (Ed.), *Proc. of Workshop on Theory and Practice of IT, ITAT'05*, Račkova dolina, 2005, pp. 325–334.
7. Filkorn, R., Návrat, P. Feature-based Filtering in Semantic Web. In B. Thalheim and G. Fiedler(Eds.), *Emerging Database Research in East Europe, Proc. of the Pre-Conference Workshop of VLDB'03*, 2003, pp. 46–50.
8. Frivolt, Gy., Bieliková, M. An Approach for Community Cutting. In V. Svátek, V. Snášel (Eds.), *Proc. of the 1st Int. Workshop on Representation and Analysis of Web Space, RAWS'05*, Prague, Czech Republic, 2005, pp. 49–54.

9. Gatial, E., Balogh, Z., Laclavík, M., Ciglan, M., Hluchý, L. Focused web crawling mechanism based on page relevance. In P. Vojtáš (Ed.), *Proc. of Workshop on Theory and Practice of IT, ITAT'05*, Račkova dolina, 2005, pp. 41–46.
10. Gurský, P., Horváth, T. Dynamic search of relevant information. In L. Popelinsky (Ed.), *Proc. of 4th Annual Conf. Znalosti'05*, Vysoke Tatry, Slovakia, 2005, pp. 194–201.
11. Halanová, Z., Návrat, P., Rozinajová, V. A Tool for Searching the Semantic Web for Supplies Matching Demands. In B. Rachev, A. Smrikarov (Eds.), *Proc. of CompSysTech'06*, Veliko Turnovo, Bulgaria, 2006.
12. Horváth, T., Krajči, S., Lencses, R., Vojtáš, P. An ILP model for a graded classification problem. *J. Kybernetika*, 40 (2004) pp. 317–332.
13. Horváth T., Vojtáš P. Ordinal Classification with Monotonicity Constraints. In P. Perner (Ed.), *ICDM 2006*, LNAI 4065, Springer, 2006, pp. 217-–225.
14. Krajči, S. Cluster based efficient generation of fuzzy concepts. Neural Network World 13,5 (2003) 521-530.
15. Laclavík, M., Seleng, M., Gatial, E., Balogh, Z., Hluchý, L. Ontology based Text Annotation OnTeA. In Y.Kiyoki et.al. (Eds.), *Proc. of 16th European-Japanese Conf. on Information Modelling and Knowledge Bases, EJC'06*, Trojanovice, Czech Republic, 2006, pp. 280–284.
16. Lencses, R. Indexing for Information Retrieval System supported with Relational Database. In P. Vojtáš et al. (Eds.), *Sofsem 2005*, Slovakia, 2005, pp. 81–90.
17. Návrat, P., Bieliková, M., Rozinajová, V. Methods and Tools for Acquiring and Presenting Information and Knowledge in the Web. In B. Rachev, A. Smrikarov (Eds.), *Proc. of CompSysTech'05*, Varna, Bulgaria, 2005.
18. Návrat, P., Bieliková, M., Rozinajová, V. Acquiring, Organising and Presenting Information and Knowledge from the Web. In B. Rachev, A. Smrikarov (Eds.), *Proc. of CompSysTech'06*, Veliko Turnovo, Bulgaria, 2006.
19. Návrat, P., Bieliková, M. Tools for Acquiring, Organising and Presenting Knowledge in Heterogeneous Environment of Information Sources. In J. Paralič, J. Dvorský, M. Krátký (Eds.), *Proc. of 5th Annual Conf. Znalosti'06*, 2006, pp. 237–242.
20. Polčicová, G., Tiňo, P. Making sense of sparse rating data in collaborative filtering via topographic organization of user preference patterns. In *Neural Networks*, Elsevier, 17 (2004) 1183–1199.
21. Smolárová, M., Návrat, P. Software Reuse: Principles, Patterns, Prospects. *Journal of Computing and Information Technology*, 5 (1), (1997) 33–48.
22. Sýkora, P., Janžo, A., Kasan, P., Jemala, M., Berta, I., Szöcs, V. Automatized Information Retrieval from Heterogenous Web Sources. In M. Bieliková (Ed.), *Proc. of IIT.SRC 2006: Student Research Conference*, Bratislava, Slovakia, 2006, pp. 137—144.
23. Tvarožek, M. Personalized navigation in the semantic web. In V. Wade, H. Ashman, and B. Smyth, editors, *4th Int. Conf. on Adaptive Hypermedia and Adaptive Web-Based Systems, AH'06*, Dublin, Ireland, Springer, LNCS 4018, 2006, pp. 467–471.
24. Vojtek, P. Natural Language Identification in the World Wide Web. In M. Bieliková (Ed.), *Proc. of IIT.SRC 2006: Student Research Conference*, Bratislava, Slovakia, 2006, pp. 153—159.
25. Vranić, V. Multi-Paradigm Design with Feature Modeling. *ComSIS (Computer Science and Information Systems)*, 2 (1), (2005) 79–102.