Computing and Informatics, Vol. 28, 2009, 393–398

ACQUIRING, ORGANISING AND PRESENTING INFORMATION AND KNOWLEDGE ON THE WEB

Pavol Návrat

Institute of Informatics and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava
Ilkovičova 3, 842 16 Bratislava 4, Slovakia
e-mail: navrat@fiit.stuba.sk

Ján Paralič

Department of Cybernetics and Artificial Intelligence
Faculty of Electrical Engineering and Informatics
Technical University in Košice
Letná 9, 042 00 Košice, Slovakia
e-mail: jan.paralic@tuke.sk

Mathematics Subject Classification 2000: 68T30, 68T05, 68T10, 68Nxx

The research topics related to all aspects of the web deserve increasing attention in the last decade, so it is quite natural that the respected journal Computing and Informatics provides appropriate space for presentation of the latest research results in this area; see e.g. papers published in previous volumes of this journal on web information retrieval [8], semantic web [3, 5, 11] or more generally on web information processing [4, 9, 12]. Therefore, it is indeed appropriate for Computing and Informatics to provide its space for such an exciting theme of research in a more compact form.

However, this issue is special also in a way that deserves to be explained in more detail. Unlike most special issues that consist of a selection of papers from a conference that are usually extended and possibly amended, or of papers submitted in response to a specific call, all the articles included in this issue emanate from a single research project. The call was intentionally targeted at topics that were among the main themes of the NAZOU research and development project, which has been

quite unique in Slovakia. Researchers from four different institutions, including two universities, one research institute and one software company combined their efforts in this three year project with a budget at least modestly comparable to European standards.

All of the institutions share a general mission of contributing to the progress in various fields of informatics and information technologies, but at the same time all of them are quite different. The researchers are from Faculty of Informatics and Information Technology of the Slovak University of Technology in Bratislava – the main contractor of the project; from Institute of Informatics of the Slovak Academy of Sciences; from Institute of Informatics of the University of Pavol Jozef Šafárik in Košice; and from Softec, Ltd., a private enterprise in the IT sector. Thus there has been a genuine blend of academia and industry on board. The leading partner has been the first faculty in Slovakia that was established focusing completely on a broad range of fields in informatics and information technologies. It is embedded within a university of technology. But there has been also an Institute of Informatics embedded within a more classical university, and also an Institute of the Slovak Academy of Sciences. It has been also very important to have also one of the leading companies in the software industry in the country within the project.

The NAZOU project (the acronym stemming from the Slovak title Nástroje pre Získavanie, Organizovanie a Udržovanie znalostí v prostredí heterogénnych zdrojov, i.e. Tools for Acquiring, Organising and Presenting Information and Knowledge in an Environment of Heterogeneous Information Sources) was 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 aimed at developing methods and tools, based on researching principles and models for acquiring, organising and presenting information and knowledge in heterogeneous information environment. Motivation for this research comes from the nature of the current World Wide Web, strengthened by the expectations from it for the future.

Motivation for this kind of research is similar as it has been in 2004. It is still true, as it could be expected, that the amount of accessible information and knowledge grows at an unprecedented rate. Its extent, quality and accessibility change also due to the world-wide use of the Internet, as well as newly appearing technologies and applications bringing new ways of internet usage, having impact on the whole society and its particular subgroups. The Internet and its services (e.g. World Wide Web, electronic mail, various web services [2], etc.) can be used as a very appropriate environment for the research of new ways of knowledge acquisition from heterogeneous sources, as well as efficient organisation, validation, evaluation and maintenance of actual knowledge. Knowledge in one of the possible interpretations originates in information, which in turn can be understood as data interpreted in a certain context. The Internet forms a distributed environment for the heterogeneous sources of information and data. The distribution of information is important from the point of view of information accessibility, while heterogeneity is the key feature of the documents presented on the Internet.

When looking for certain information we are often overwhelmed by a huge amount of data of various kind and quality. Many search tools provide too extensive and partly even irrelevant answers to user queries. On the other side, these tools are not able to provide information which is available 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, which focus on a certain area of interest only. Models, methods and approaches that have been designed, implemented and evaluated within the NAZOU project targeted both directions. Underlying theoretical models are general enough to support universal search, indexing and organising of semantically enhanced information. On the other side, selected application domains have been used in order to experiment with personalisation and other methods in their full detail on real data of realistically large extent.

It is becoming clear that approaching to the stage when virtually everything is on the Web will 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 [1]. This may require devising new models of information (and knowledge) representation. In this context, the project was developing models and approaches that basically contribute to the concept of Semantic Web. The core of the business, so to speak, remains the search. However, to retrieve the right information requires identifying what is right for a particular user. Recommending is one possible way to narrow the set of answers. Annotating improves quality of the retrieved information. Personalising the presentation improves the overall outcome. As a sample domain for experimentations, job offers have been used in the project. This domain is sufficiently complex, broad and structured. Recommendation, personalisation and annotation are in the centre of interest in major part of the articles presented in this issue of the journal.

The project has produced many specific results. It is not possible to present all the results in a single journal issue. Anyway, most of them have been presented at various conferences or published in journals. Preliminary results were presented and extensively discussed at specifically targeted workshops. Papers from these workshops have been published in two volumes of proceedings [6, 7].

The six articles included in this issue, while being representative as far as the scientific merit of the project is concerned, cannot represent the whole spectrum of research results that have been achieved in the project. The sole criterion for inclusion in this issue has been positive recommendations of independent reviewers appointed according to the standard procedure of this journal. In response to our call, we received eight submissions. Based on outcomes of the reviewing process, we decided to accept the following six articles.

Barla et al. describe a user modelling approach based on automated acquisition of user behaviour and its successive rule-based evaluation and transformation into

an ontological user model. They developed an enhanced faceted browser, which provides personalised navigation support and recommendation. One of the problems with personalised content presentation or navigation is how to compare ontological concepts. Andrejko and Bieliková propose a method for instance comparison. By comparing properties of documents that attracted a particular users interest, they attempt to discover information on that users interests. They also developed a concept comparer software tool called ConCom and performed extensive experiments in domains of job offers and scientific publications. The third article dealing with user properties written by Horváth proposes a formal model of user preference learning specifically for content-based recommender systems. He works with three kinds of learning tasks: the exact, the order preserving and the iterative user preference one. He developed a learning algorithm for his model of induction of generalized annotated programs.

Gurský et al. concentrate on an interesting problem connected to searching top-k objects when there are generally several users having different preferences. They propose a new combination of back end data maintenance system and a middleware top-k optimization heuristics resulting in interesting speedups and significant disk access savings when compared to other state-of-the art approaches. In another article by Gurský, Horváth et al. the authors describe a system for user preference web search which employs their method for searching top-k objects. In addition, the system also contains a web information extraction part, a text search engine, and a user interface for querying and evaluation of the search results. The model of user preferences is represented by fuzzy sets and fuzzy logic and its learning is performed by means of fuzzy inductive logic programming.

Laclavík et al. studied feasibility of automated annotation of web documents. Their method is based on regular expression patterns. They developed Ontea, a platform for automated semantic annotation or semantic tagging. Moreover, the authors also present an approach how their method can be scaled up to large pattern based annotation and evaluate the experimental results achieved.

We hope this selection of papers will give the reader a sufficiently broad insight into results of the project. Of course, the researchers have published much more papers dealing with research issues of methods for acquisition of data and offers, methods for analysis and organisation of data and offers, methods for presentation of job offers and user modelling, including wrapper learning, probabilistic clustering, adaptive faceted browsing, lemmatization, off-line and on-line annotation.

We wish to thank all the reviewers for all their efforts and valuable feedback to the authors. We especially thank all the authors for their devotion to the project and their enthusiasm in preparing these papers. And, finally, we express our hope that you, dear reader, will enjoy this issue and will find here interesting information and knowledge for you.

REFERENCES

[1] AKT, Advanced Knowledge Technologies, `http://www.aktors.org/akt/`.

[2] Habala, O.—Paralič, M.—Bartaloš, P.—Rozinajová, V.: Semantically-Aided Data-Aware Service Workflow Composition. In: Lecture Notes in Computer Science, Vol. 5404, Heidelberg, Germany, Springer, 2009, ISSN 0302-9743, pp. 317–328.

[3] Laclavík, M.—Balogh, Z.—Babík, M.—Hluchý, L.: AgentOwl: Semantic Knowledge Model and Agent Architecture. Computing and Informatics, Vol. 25, 2006, No. 5, pp. 421–439.

[4] Machová, K.—Bednár, P.—Mach, M.: Various Approaches to Web Information Processing. Computing and Informatics, Vol. 26, 2007, No. 3, pp. 301–327.

[5] Matušíková, K.—Bieliková, M.: Social Navigation for Semantic Web Applications Using Space Maps. Computing and Informatics, Vol. 26, 2007, No. 3, pp. 281–299.

[6] Návrat, P.—Bartoš, P.—Bieliková, M.—Hluchý, L.—Vojtáš, P. (Eds.): Tools for Acquisition, Organization and Presenting of Information and Knowledge. Proceedings. Research Project Workshop, Bystrá Dolina, Low Tatras, Slovakia, September 2006, 256 pp., ISBN 80-227-2468-8.

[7] Návrat, P.—Bartoš, P.—Bieliková, M.—Hluchý, L.—Vojtáš (Eds.): Tools for Acquisition, Organization and Presenting of Information and Knowledge. Proceedings, Research Project Workshop, Poľana, Slovakia, September 2007, 256 pp., ISBN 978-80-227-2716-7.

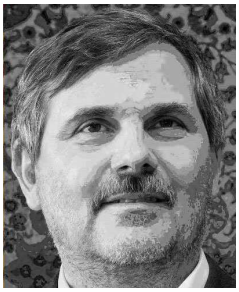
[8] Návrat, P.—Kovaľ, R.: Intelligent Support for Information Retrieval of Web Documents. Computing and Informatics, Vol. 21, 2002, No. 5, pp. 509–528.

[9] Svátek, V.—Vacura, M.—Ten Teije, A.: Modelling Web Service Composition for Deductive Web Mining. Computing and Informatics, Vol. 26, 2007, No. 3, pp. 255–279.

[10] Tvarožek, M.: Personalized Navigation in the Semantic Web. In: Lecture Notes in Computer Science, Vol. 4018, Berlin Heidelberg, Springer-Verlag, 2006, ISSN 0302-9743, pp. 467–471.

[11] Wang, P.—Xu, B.: Debugging Ontology Mappings: A Static Approach. Computing and Informatics, Vol. 27, 2008, No. 1, pp. 21–36.

[12] Alípio, P.—Neves, J.—Carvalho, P.: An Ontology for Network Services. Computing and Informatics, Vol. 26, 2007, No. 5, pp. 543–561.



Pavol Návrat received his Ing. (Master) degree cum laude in 1975, and his Ph. D. degree in computing machinery in 1984, both from Slovak University of Technology. He is currently a Professor of informatics at the Slovak University of Technology and serves as the Director of the Institute of Informatics and Software Engineering. During his career, he was also with other universities overseas. His research interests include related areas from software engineering, artificial intelligence, and information systems. He published numerous research articles and the following books: Programming in Lisp, Microcomputers, Programs, People, and textbooks Functional and Logic Programming, and Artificial Intelligence. He co-edited and co-authored the monographs: Knowledge-Based Software Engineering (Amsterdam, IOS-Press, 1998), Advances in Databases and Information Systems (Heidelberg, LNCS Springer, 2002) and SOFSEM 2008 (Heidelberg, LNCS Springer, 2008). He was editor of a special issue on Knowledge Based Software Engineering of the journal Informatica in 2001. He is a Fellow of the IET and a Senior Member of the IEEE and its Computer Society. He is also a member of the ACM, Association for Advancement of Artificial Intelligence, Slovak Society for Computer Science and Slovak Artificial Intelligence Society. He serves on the Technical Committee 12 Artificial Intelligence of IFIP as the representative of Slovakia.

Ján Paralič received his Master degree in 1992 and his Ph. D. degree in 1998, both from the Technical University in Košice. He is currently an Associate Professor at the Department of Cybernetics and Informatics, Technical University in Košice and the Head of the Centre for Information Technologies at the same university. He (co-)authored two books, (co-)edited 12 proceedings from various international workshops and conferences and published more than 80 scientific papers. His research interests currently include knowledge discovery, text mining, semantic technologies, and knowledge management. He was editor of a special issue of this journal devoted to knowledge technologies and related topics in 2007 (Vol. 26, No. 3). He is a member of the ACM, IEEE, Slovak Society for Computer Science and Slovak Artificial Intelligence Society.