ADVANCED KNOWLEDGE TECHNOLOGIES FOR E-GOVERNMENT

Ján PARALIČ, Tomáš SABOL, Marián MACH
Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, tel. 055/602 4128, E-mail: Jan.Paralic@tuke.sk

SUMMARY

There are a growing number of e-Government portals and solutions available today. But what the users lack in particular is a customized assistance – help that meets the individual situation and competence [15]. In this paper, a system called Webocrat will be presented as an attempt to shift e-Government portals toward this direction, providing various knowledge technologies as its core components [13].

The Webocrat system applies a knowledge-based approach [6]. Information of all kinds produced by various modules is linked to a shared ontology representing an application domain. Such ontology serves as a mean for structuring and organizing available information resulting in improved search capability and contents presentation. Moreover, this approach is utilized also e.g. for support of personalization, when each user may define her/his area of interest by means of concepts from the knowledge model and be automatically notified about new information in the system relevant to her/his profile.

Finally, knowledge discovery techniques [11] applied on text information available within the portal are used for automatic annotation of new sources.

Keywords: *e-Government, knowledge technologies, ontology, knowledge discovery in texts*

1. INDRODUCTION

Knowledge can be simply defined as actionable information [14]. That means that (only) relevant information available in the right place, at the right time, in the right context, and in the right way can be considered as knowledge.

The knowledge life cycle defined in [10] hinges on the distinction between *tacit* and *explicit* knowledge. Explicit knowledge is a formal one and can be found in documents of an organization: reports, manuals, correspondence (internal and external), patents, pictures, tables (e.g. Excel sheets), images, video and sound recordings, software etc. Tacit knowledge is personal knowledge given by individual experience (and hidden in peoples' minds) [14].

A considerable amount of explicit knowledge is scattered throughout various documents within public and governmental organizations and people minds working there. In many cases the possibility to efficiently access (retrieve) and reuse this knowledge is limited [4]. As a result of this, most knowledge is not sufficiently exploited, shared and subsequently forgotten in relatively short time after it has been introduced to, invented/discovered within the organization. Therefore, in the approaching information society, it is vitally important for knowledge-intensive organizations as public and governmental institutions to make the best use of information gathered from various information resources inside the organizations and from external sources like the Internet. On the other hand, tacit knowledge of authors of the documents' provides important context to them, which cannot be effectively intercepted.

Knowledge management [14] generally deals with several activities relevant in knowledge life cycle [1]: identification, acquisition, development, dissemination (sharing), use and preservation of organization's knowledge. Our approach to knowledge management in the e-Government context supports most of the activities mentioned above. Based on this approach, a Web-based system Webocrat[1] [13] has been designed and implemented. It is being now tested on pilot applications at Wolverhampton (UK) and in Kosice (Slovakia). Firstly, it provides tools for capturing and updating of tacit knowledge connected with particular explicit knowledge inside documents. This is possible due to ontology model, which is used for representation of organization's domain knowledge. Ontology with syntax and semantic rules provides the 'language' by which Webocrat(-like) system can interact at the *knowledge level* [9].

Use of ontology enables to define concepts and relations representing knowledge about a particular document in domain specific terms. In order to express the contents of a document explicitly, it is necessary to annotate it, i.e. create links between the document and relevant parts of a domain model, i.e. links to those elements of the domain model, which are relevant to the contents of the document. Model elements can be also used for intelligent search and retrieval of relevant documents.

Existence of a knowledge model (ontology) in the center of the system is the key difference to approaches followed by other 5th FP IST projects like EDEN (they are strong in use of natural language processing techniques supporting communication between citizens and public administrations) or DEMOS (very elaborated approach focused on on-line consultation).

[1] EC funded project IST-1999-20364 Webocracy (Web Technologies Supporting Direct Participation in Democratic Processes)

The rest of this paper is organized as follows. Section 2 describes the functional overview of the Webocrat system. Section 3 presents three examples of knowledge technologies exploitation within the Webocrat system. Finally, section 4 provides a summary of the whole paper.

2. WEBOCRAT SYSTEM FUNCTIONAL OVERVIEW

From the point of view of functionality of the Webocrat system it is possible to break down the system into several parts and/or modules [13]. They can be represented in a layered sandwich-like structure, which is depicted in Fig. 1.

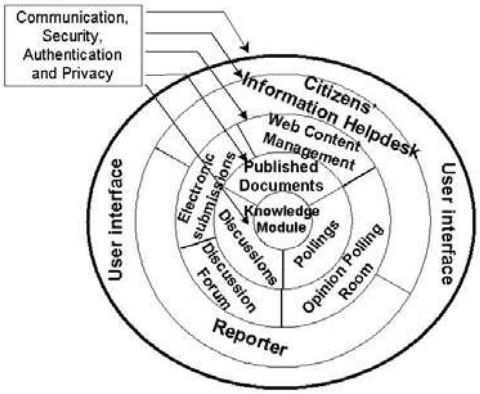

Fig. 1 Webocrat system structure from the system's functionality point of view

2.1 First layer

A Knowledge Model module occupies the central part of this structure. This system component contains a conceptual model of a domain. The purpose of this component is to index all information stored in the system in order to describe the context of this information (in terms of domain specific concepts). The central position symbolizes that the knowledge model is the core (heart) of the system – all parts of the system use this module in order to deal with information stored in the system (both for organizing this information and accessing it).

2.2 Second layer

Information stored within the system has the form of documents of different types. Since three main document types will be processed by the system, the document space can be divided into three subspaces – publishing space, discussion space, and opinion polling space. These areas contain published documents, users' contributions to discussions on different topics of interest, and records of users' opinions about different issues, respectively.

2.3 Third layer

Since each document subspace expects different way of manipulating with documents, three system's modules are dedicated to them. Web Content Management module (WCM) offers means to manage the publishing space. It enables to prepare documents in order to be published (e.g. to link them to elements of a domain model), to publish them, and to access them after they are published. It is also possible to publish interesting web links, i.e. commented URL references to other interesting and relevant sources. These are also indexed and therefore are searchable as other, internal sources.

Special subpart of the WCM module is responsible for publishing of tenders (TDM). There can be published all necessary information related to an open tender, as well as publishing of results of a closed tender.

Discussion space is managed by Discussion Forum module (DF). The module enables users to contribute to discussions they are interested in and/or to read contributions submitted by other users. There is possible to set a discussion forum as a moderated one, or not.

Electronic Submissions module (ES) enables users to access special part of the document space comprising their formal or informal submissions to local authority.

Opinion Polling Room module (OPR) represents a tool for performing opinion polling on different topics. Users can express their opinions in the form of polling.

2.4 Fourth layer

In order to navigate among information stored in the system in an easy and effective way, this layer is focused on retrieving relevant information from the system in various ways. Two modules represent it, each enabling easy access to the stored information in a different way. Citizens' Information Helpdesk module (CIH) is dedicated to search. It represents a search engine based on the indexing and linking (to knowledge model) of stored documents. Its purpose is to find all those documents, which match user's requirements expressed in the form of a query defined by means of a free text, or a query composed from concepts from knowledge model.

The other module performing information retrieval is the Reporter module (REP). This module is dedicated to providing information of two types. The first type represents information in an aggregated form. It enables to define and generate different reports concerning information stored in the system. The other type is focused on providing particular documents – but unlike the CIH module it is oriented on off-line mode of operation. It monitors content of the document space on behalf of the user and selects new information that the user may be interested in.

2.5 Fifth layer

The upper layer of the presented functional structure of the system is represented by a user interface. It integrates functionality of all the modules accessible to a particular user into one coherent portal and provides access to all functions of the system in a uniform way (see Fig. 2 as an example).

The user interface can be easily adjusted as required by actual public administration. Good examples are 3 different Webocrat pilot applications sites (Wolverhampton[2], UK, Košice – Furča[3] and Košice – Dargovských hrdinov[4], Slovakia).

Fig. 2 Webocrat user interface at the local authority Košice – Dargovských hrdinov

In order for the system to be able to provide required functionality in a real setting, several security issues must be solved. This is the aim of the Communication, Security, Authentication and Privacy module (CSAP) [5].

Technical achievements comprise also a system designed to provide automatic routing of messages from citizens to the appropriate person within the public administration (ES module) and tools for easy access to public administration information and to competitive tendering (WCM module).

3. KNOWLEDGE TECHNOLOGIES APPLIED

3.1 Knowledge model for concept-based retrieval

The main idea behind the whole Webocrat system is to associate information with a part of a knowledge model – ontology. In this way it is possible to annotate discussions, reports, tenders, polling or ordinary WWW pages (all those documents that are published, such as news, announcements, memos and other documents that could be interesting for users).

When pieces of information are submitted, they are annotated first, whether manually or semi-automatically (see subsection 3.3). After that they are prepared for intelligent retrieval. When accessing information, user can define his/her query consisting of words for full-text search or of terms (concepts) used in ontology.

Using concepts ensures that also hidden meaning will be discovered. User selects interesting concepts and asks for information related to them. Selected concepts are used for information retrieval. The decision about document relevance to the user query is based on a similarity between set of query concepts and a set of concepts, which are annotated to the document. When making this calculation of similarity, whole ontology structure (i.e. concepts connected with various relations) is taken into account. In such a way not only exact matching, but also partial matching is secured.

This task of document retrieval can be viewed as a classification task when the decision is made, whether the document is relevant for the user or not. With appropriate ontology which models domain well, use of this knowledge model can yield better results than full-text retrieval.

3.2 Knowledge based personalization

Since the system can contain a lot of information in different formats (published information, discussion contributions, etc.), it may not be easy to find exactly the information user is looking for. Therefore he/she has the possibility to create his/her profile in which he/she can define his/her interests and/or preferred way of interacting with the system.

When defining an area of interest, user selects elements from a knowledge model (or subparts of this model). In this way user declares that he/she is interested in topics defined by the selected part of the knowledge model.

The definition of user's area of interest enables alerting – user can be alerted, e.g. on a new opinion polling, or publishing of a new document, opening of a new discussion, etc. User has the possibility to set alerting policy in detail on which kind of information he/she wants to be alerted in what way (including extreme settings for no alerting or alerting on each event taking place in the system). The system compares each event (e.g. submission of a discussion contribution, publishing a document, etc.) to users' profiles. If result of this comparison is positive, i.e. the user may be interested in the event, then the user is alerted.

Alerting can have two basic forms. The first alternative is represented with notification using e-mail services. User can be notified on event-per-event basis, i.e. he/she receives an e-mail message for each event he/she is alerted on. Alternatively, it is possible to use an e-mail digest format – user receives e-mail message, which informs him/her about several events. The way of packaging several

[2] http://www.wolforum.org/
[3] http://www.tahanovce.sk/mutah/
[4] http://www.kosice-dh.sk/

alerts into one e-mail message depends on time intervals given by the user.

The other alternative is a 'personal newsletter'. This does not disturb user at unpredictable time – user simply can access his/her newsletter when he/she desires to be informed what is going on in the system. Moreover, he/she can access it from arbitrary gadget connected to the Internet. The personal newsletter has the form of a document published in the publishing space. This document is generated by the system and contains links to all those documents, which may be of interest for the user. Since the document is generated when user logs in, it can cover all information submitted and/or published since the last user's visit.

User registered in the system as an individual entity (i.e. not anonymous user) is provided with a personal access page ensuring him/her an individual access to the system. This page is built in an automatic way and can consist of several parts. Some of them can be general and the others are user-specific.

The former can serve as a starting point for browsing all published documents accessible to the user, all discussions he/she is allowed to participate in (in passive or active way), all running polls for which he/she is eligible, using search facilities of the system, read hot information, etc. The latter parts are devoted to user's personal newsletter, links to documents and discussions topics of which match the user's area of interest.

3.3 Knowledge discovery for annotation

Three different text data mining techniques (clustering/visualization, association rules and classification models) have been analyzed and its exploitation possibilities within the Webocracy project are described in more details in [11]. Clustering and association rules discovery are well suited as supporting tools for ontology management. Classification models are used for automatic documents annotation.

Retrieval accuracy depends on the quality of documents annotation. Text mining methods used for knowledge discovery in text collections can be very useful to guide user at annotating new document. Annotation of the new document is the classification task (text categorization task in information retrieval terminology) when we need to make decision which concept (concept represents category) is relevant to the content of the document.

The system must propose relevant concepts for new document in real time, so important requirement to used algorithm is execution time efficiency. User can add or delete some link between new document and concepts, and these changes can be immediately integrated into classifier. This requires ability of incremental learning. Relevance weighting of the concepts to the new document is better than simple binary decision. Concepts can be ordered by weight of the relevance to the new document and user can search for additional relevant concept according to this ordering.

Different types of classifiers for this task have been analysed in [2]. At the moment two classifiers are used within the Webocrat system. The Naive Bayes classifier works well, when there are already a satisfactory number of well-annotated documents presented in the system. For the starting period of a new e-Government portal driven by Webocrat, a rule-based classifier is applied. After some period of time, when there is enough documents for training the Naive Bayes classifier, this one is utilized.

4. CONCLUSION

In this paper, a system called Webocrat has been presented as an attempt to shift e-Government portals toward a customized assistance and knowledge-enhanced services. The Webocrat system applies a knowledge-based approach. The functional overview of the system has been presented as well as three examples of knowledge technologies exploitation within this system.

The focus is on using knowledge models to annotate and organize information within the system in order to retrieve this information according to the content. This approach is utilized also for support of personalization, when each user may define her/his area of interest by means of concepts from the knowledge model and be automatically notified about new information in the system relevant to her/his profile.

Finally, classification text mining techniques applied on textual information available within the e-Government portal are used for automatic annotation of new sources that are being published on the portal.

Acknowledgements

This work is done within the Webocracy project, which is supported by European Commission DG INFSO under the IST program, contract No. IST-1999-20364 and within the VEGA project 1/8131/01 "Knowledge Technologies for Information Acquisition and Retrieval" of Scientific Grant Agency of Ministry of Education of the Slovak Republic.

The content of this publication is the sole responsibility of the authors, and in no way represents the view of the European Commission or its services.

REFERENCES

[1] Abecker A., Bernardi A. Hinkelmann K. Kühn, O. & Sintek M.: Toward a Technology for Organizational Memories, IEEE Intelligent Systems, No. 13, 1998, pp. 40-48

[2] Bednar, P., Hudak, S.: Use of text categorization for document annotation. In Proc. of the Znalosti 2003 conference, Ostrava, Czech

Republic, February 2003, pp. 291-296 (in Slovak)
[3] Bocock, R., Cizmarik, T., Novacek E., Paralic J. and Thomson, P.: Specification of pilot applications and design of trials. Webocracy Technical Report R12.1, Technical University of Košice, 2001
[4] Borghoff U. M. & Pareschi R. (Eds.): Information Technology for Knowledge Management. Springer Verlag, 1998
[5] Dridi, F. and Pernul, G. and Unger, V.: Security for the electronic government. In Proceedings of the European Conference on E-Government, Trinity College, Dublin, Ireland, 2001, pp. 99-110
[6] Dzbor M., Paralic J. and Paralic, M.: Knowledge management in a distributed organization. In: Proc. of the 4th IEEE/IFIP International Conference BASYS'2000, Kluwer Academic Publishers, London, 2000, pp. 339-348
[7] Macej P., Novacek E., Thomson P., Paralic J. and Cizmarik T.: Conceptual Knowledge Models (Phase I). Webocracy Technical Report R12.2.1, Technical University of Košice, 2002
[8] Mach M., Sabol T., Dridi F., Palola I., Furdik K., Thomson P., Cizmarik T., Novacek E.: Edited annual report for publication. Webocracy Technical Report R1.1.3, Technical University of Košice, 2001
[9] Newell A.: The Knowledge Level. Artificial Intelligence, No. 18, 1982, pp. 87-127
[10] Nonaka I., Takeuchi H.: The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation. Oxford Univ. Press, 1995
[11] Paralic, J., Bednar, P.: Knowledge Discovery in Texts Supporting e-Democracy. In Proc. of the 6th IEEE International Conference on Intelligent Engineering Systems, INES 2002, Opatija, Croatia, May 2002, pp. 327-332
[12] Paralic, J., Sabol, T.: Implementation of e-Government Using Knowledge-Based System. In. Proc. of the 12th Int. Workshop on Database and Expert Systems Applications, (W06: 2nd Int. Workshop on Electronic Government), Munich, 2001, pp. 364-368
[13] Paralic, J., Sabol, T., Mach, M.: A System to support E-Democracy. Proc. of the First International Conference EGOV 2002, Aix-en-Provence, France, LNCS 2456, Electronic Government, R. Traunmuller, K. Lenk (Eds.), Springer Verlag, 2002
[14] Tiwana A.: The Knowledge Management Toolkit. Prentice Hall, 2000
[15] Traunmüller, R. & Wimmer, M.: Directions in E-Government: Processes, Portals, Knowledge. Proc. of the Int. Workshop "On the Way to Electronic Government" in Conjunction with DEXA (Munich, Germany), IEEE Computer Society Press, Los Alamitos, CA, 2001, pp. 313-317

BIOGRAPHY

Ján Paralič, M.S., Ph.D. is Assistant Professor at the Dept. of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics. Studied Technical Cybernetics at the Technical University of Kosice and Computer Science at the Technical University of Vienna. Research in artificial intelligence - knowledge discovery, knowledge management, constraint logic programming, scheduling (published more than 60 papers in these fields). Dr. Paralic has extensive experience in international projects. His teaching experience involves lectures and seminars in Knowledge Discovery, Knowledge Management, Scheduling and Logistics, AI Languages and Intelligent Information Systems.

Tomáš Sabol, M.S., Ph.D. is Associate Professor, working with the Faculty of Electrical Engineering and Informatics and Faculty of Economics. He studied Technical Cybernetics at the Technical University of Prague and Technical University of Kosice. He has undertaken research in problem solving, knowledge management, knowledge modelling, continuing education management, project management (published about 80 papers). In 1994-1997 he was a Vice-Rector of the TUK for international relations and scientific activities. Dr. Sabol has considerable experience in international project management and international activities. Currently, he is also coordinator of the IST-1999-20364 project "Webocracy".

Marián Mach, M.S., Ph.D. is Associate Professor at the Dept. of Cybernetics and Artificial Intelligence Faculty of Electrical Engineering and Informatics. Studied Technical Cybernetics at the Technical University of Kosice. Research in artificial intelligence – uncertain knowledge processing, logic programming, evolutionary computation, expert systems, knowledge engineering, machine learning, constraint satisfaction, and programming languages for artificial intelligence (published more than 50 papers in these fields). Author of the book on knowledge acquisition for knowledge-based systems. Dr. Mach has extensive experience with research and technical projects. He has considerable experience in building information and control systems for business and industry. He participated in several industrial projects as a system analyst and project manager.