SEMANTIC-BASED GROUPWARE SYSTEM FOR COLLABORATIVE SUPPORT IN PUBLIC ADMINISTRATION PROCESSES

Peter BUTKA*, Gabriel LUKÁČ**, Marián MACH**, Tomáš SABOL*
*Faculty of Economics, Technical University of Košice, Letná 9, 042 00 Košice,
e-mail: peter.butka@tuke.sk, gabriel.lukac@tuke.sk, tomas.sabol@tuke.sk
**Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, e-mail: marian.mach@tuke.sk

ABSTRACT

In this article a Semantic-based Groupware System (GWS) implemented within the SAKE project (Semantic Agile Knowledge-based E-government) is presented. The overall objective of the SAKE project is to specify, develop and deploy a holistic framework and supporting tools for an agile knowledge-based e-government that will be sufficiently flexible to adapt to changing and diverse environments and needs. Public administration processes are knowledge-intensive, dependant on valid legislation, and experts in specific domains are needed for their management. The core of the whole SAKE system is represented by an integrated knowledge space unifying different perspectives and interpretations of existing knowledge resources. It enables to assign metadata to each knowledge object, allowing thus more sophisticated retrieval. The main objective of the Technical University of Kosice research team was to provide a semantic-based groupware system supporting public servants in sharing their knowledge using the collaborative software as well as to help them in managing and creation of expert working groups, which are frequently used in governmental internal administration processes. One of the main contributions to the semantic enhancement of the system, described in this paper, is based on a method of discussion forums analysis and calculation of users' ranking, based on the actual annotated discussion topic.

Keywords: *groupware, collaboration, semantic technologies, discussion analysis, knowledge-based systems, e-government*

1. INTRODUCTION

At present times, public administration (PA) has to adapt to frequent changes in legislation and outer environment, while within the process a large amount of new information has to be absorbed by the organisation in a very short time. Public administration processes are influenced by external changes (mainly political, legislative and economic) as well as internal changes (e. g. organizational restructuring).

A change in one activity in an administrative process may imply changes in other parts of the same process or system. Therefore, there is an urgent need for resolving these changes in a systematic manner, ensuring overall consistency.

Furthermore, these changes impose the need of updating the knowledge needed to administer the given process. These changes are even more frequent in the case of the EU new member states, since their full integration heavily depends on the possibility to adapt their public administrations to the existing EU regulations in a very short period of time.

The existing approaches to knowledge management in e-government focus mainly on the efficient management of a particular, isolated knowledge resource and on supporting only message-based communication between public administrators (see Fig.1). However, the demands for knowledge-based e-government are much higher [1]:

- Firstly, the existing approaches do not take into account the increased granularity of informational resources and the manifold semantic differences in dealing with these resources.
- Secondly, due to the complexity of decision making processes, effective knowledge management requires creation of a supportive, collaborative culture while eliminating traditional rivalries.
- Thirdly, the usage of existing knowledge resources is indeed a valid aspiration, but for realizing a learning e-government it is crucial to create *new* knowledge resources.
- Finally, ad hoc management of the changes in e-government systems might work in the short term, but to avoid unnecessary complexity and failures in the long run, this change management must be done in a systematic way.

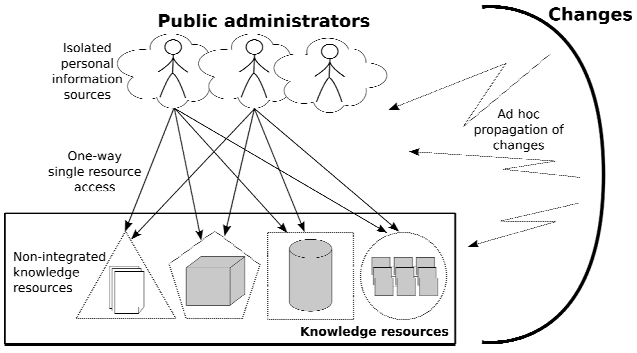

Fig. 1 As-Is situation in public administrations

SAKE (Semantic Agile Knowledge-based e-Government, http://www.sake-project.org) is a three-year IST Project (STREP) co-funded by the European Union, which started in March 2006. Project consortium consists of 11 partners from five countries (Germany, Greece, Hungary, Poland, and Slovakia). The overall objective of SAKE is to specify, develop and deploy a holistic framework and supporting tools for an agile knowledge-based e-government that will be sufficiently flexible to

adapt to changing and diverse environments and needs. The whole SAKE approach will provide tools and methodologies to address these problems. More specifically, SAKE intends to provide (see Fig. 2, [1]):

1. *Integrated knowledge space* instead of a set of isolated and heterogeneous knowledge resources.
2. *Collaborative working environment* instead of a single person decision making process.
3. *Attention (change) management system* instead of ad-hoc management of changes.
4. A platform for *proactive delivery of knowledge* (instead of an one-way knowledge access) that enables creation of an adaptable knowledge sharing environment through learning from the collaboration between public servants and their interaction with the knowledge repository and supporting in this way full empowerment of public servants.

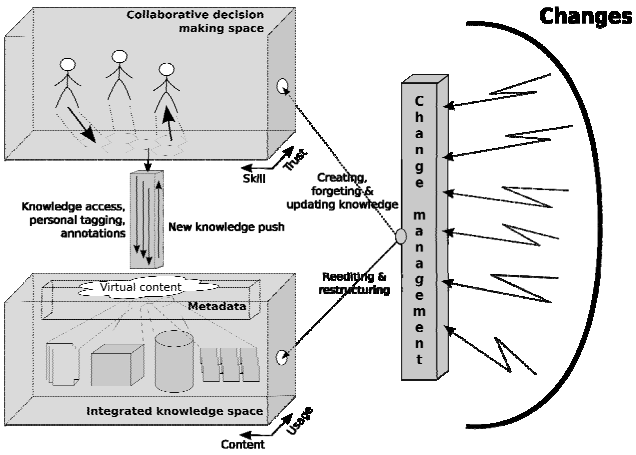

Fig. 2 To-Be situation enabled by SAKE

In the aforementioned way, SAKE provides a framework for an agile knowledge-based e-government, fulfilling even "unpredictable" knowledge needs of public servants in order to ensure high and homogeneous quality of the decision making process, especially in a frequently changing environment.

In the second section of this article the overall SAKE architecture will be described. In the next three sections details regarding the groupware component and main semantic enhancements and features provided by this component (especially method for discussion analysis) will be presented. The discussion analysis leads to ranking of users regarding their previous communication activities within the forums. It combines user's feedback and filtering of threads (according to annotation of messages in threads) and therefore provides 'argumentation-like' support solution leading to feedback-based, topic-sensitive ranking of discussion forums users. Results of this ranking can be subsequently used for a decision whom (from an available pool of experts and public servants) to invite for a discussion of a new case. At the end of the paper related work section and conclusions are provided.

2. SAKE OVERALL ARCHITECTURE

Based on the state-of-the-art analysis of back-office processes, with the intention to apply semantic technologies, we have identified three main individual technological components and a unifying conceptual framework based on a set of ontologies (see Fig.3):

- Semantic-based change (attention) management component;
- Semantic-based content management component;
- Semantic-based Groupware component;
- Conceptual framework for semantic-enabled agile knowledge-based e-government.

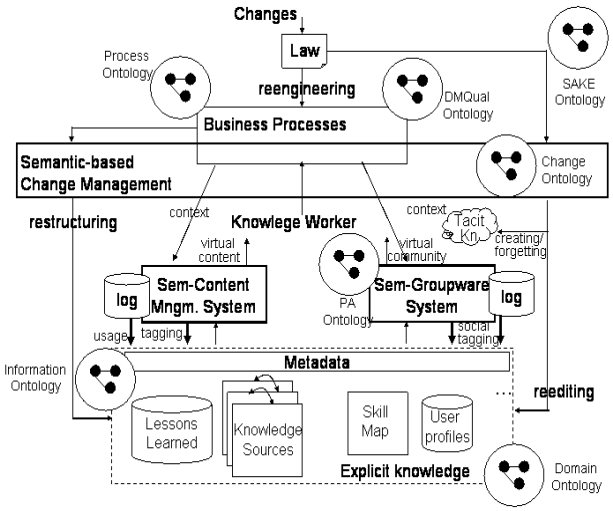

Fig. 3 Overall architecture of the SAKE system

Semantic-based attention (change) management system (AMS) ensures high quality of the knowledge update (reediting) process by developing a change management process enabling consistent propagation of changes to all involved knowledge stakeholders in order to ensure quality of the decision making process, formal and explicit modelling of changes in the PA regulations and their relations to corresponding artefacts in the form of Change Ontology (which will serve as a backbone of the change management approach), and developing methods and tools for verification of the existing knowledge repository in order to make it easier to understand and cheaper to manage, without any loss of information content.

Semantic-based content management system (CMS) enables efficient provision of knowledge in the context of a PA process by semi-automatic population of the Information Ontology. It develops methods and tools for ontology-based tagging, methods and tools for realizing context-aware searching for virtual content, and also methods and tools for an editorial process, to satisfy the knowledge items evaluation requirements.

Semantic-based groupware system (GWS) supports more efficient knowledge sharing by developing methods and tools for ontology-based tagging of the interaction between public servants, methods and tools enabling building of a community of practice from the interaction log and their specific vocabularies, methods and tools for collaborative knowledge pushing and searching for experts.

SAKE also develops a *conceptual framework* for a semantic-enabled agile knowledge-based e-government that will comprise an analysis of the knowledge

infrastructure and knowledge sources in e-government. The main ontology is the SAKE ontology that gives a global view on the dependencies between entities involved in an e-government decision making process. It serves as a merging point between particular ontologies and it is important integration element for integrated SAKE platform (known as Common Knowledge Space – CKS). The dependencies between these ontologies are shown on Fig.4. Several ontologies have been developed for the SAKE, based on the purpose of these ontologies:

- *Domain ontology* – models the terminology used in the e-government domain;
- *Information ontology* – models different kinds of information sources including their structure, access and format properties;
- *Change ontology* – is related to Information ontology, which is edited for structuring of changes of the information sources;
- *Process/Profile ontology* – models how an administrative process works and what it is about;
- *PA (Public Administrators) ontology* – semantic models of users (public servants), their roles and skills;
- *DMQual ontology* – quality model of the decision making process in PA (including guidelines for estimating the quality of the decision making processes based on the user and quality model).

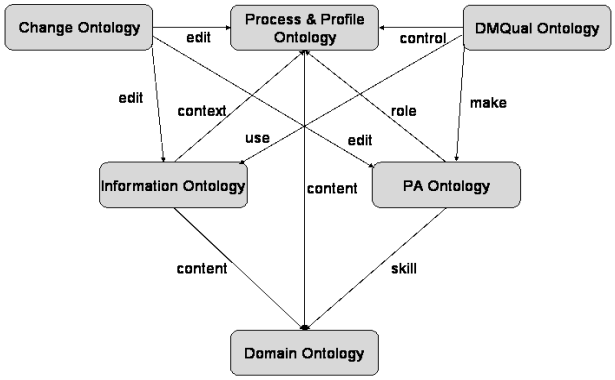

Fig. 4 The SAKE ontology

Another important integration element was identified after the first year of the project, which is (from functional point of view) the core technological element of the business process management inside AMS – *Workflow management system* (WfMS). WfMS is fully developed and implemented using jBPM (for more information – http://www.jboss.com/products/jbpm). All other SAKE subsystems are tightly integrated using this workflow-based business management, strongly supporting business context management and sharing of the actual context. The subsystems log all the user-actions semantically to provide a data-source for the AMS that perform analyses on these data and pushes new knowledge back to the users.

The SAKE solution (from the implementation perspective) consists of (at least) three individual components: 1) AMS – Attention (Change) Management System; 2) CMS – Content Management System; and 3) GWS – Groupware System. These three components are supplemented by a more integration-related, CKS (ontology integration subsystem) and Workflow Management System (WfMS). The platform is an integrated platform acting as a shell, where different kinds of problems may be introduced. User access to the platform is provided by a single point of access to the system. Personalized access is required, where user interface (and the offered actions) depends on the roles.

The graphical user interface (GUI) components are decoupled from business logic. Thus the AMS, CMS and GWS components need to offer an interface. The Portlet technology (JSR 168) is used, which allows components to access the desired business logic. The use of Portlet technology allows a better modularization; it facilitates development of the user interface in a distributed team. Ideally, portlets for the CMS, GWS and AMS can be put together in a "plug-and play" fashion thanks to the well defined interfaces of the Portlet specification. The CKS consists of ontologies described above.

To avoid problems with integration of multiple components, it was decided to use a common development platform based on the use of J2EE/EJB for the implementation of AMS, CMS, GWS business logic (JBoss used as an Application Server), Portlet technology (JbossPortal used as portlet container – http://www.jboss.org/jbossportal/) for the implementation of the user interfaces provided by AMS, CMS, and GWS. The SAKE main interface is 'Portal' which contains 'Portlets'. The KAON2 (http://kaon2.semanticweb.org/) system with its API (wrapped in an EJB) is used for ontology management and reasoning.

3. GROUPWARE SYSTEM

3.1. Conceptual architecture and components

The structure of the SAKE groupware system (GWS) can be viewed as consisting of four basic layers with several components in the particular layers (Fig. 5) [2].

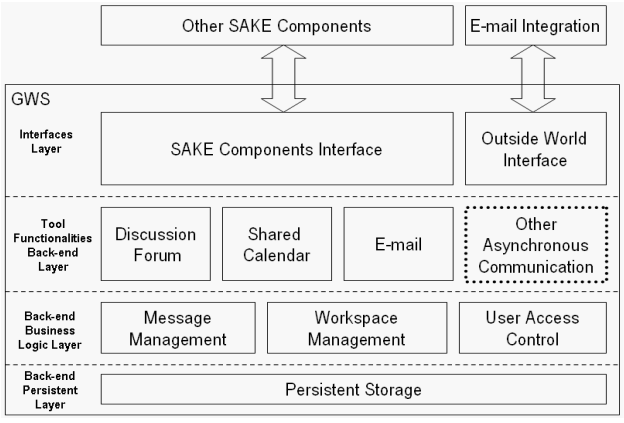

Fig. 5 High-level conceptual architecture of GWS

The *Back-end Persistence Layer* – this is the back-end storage necessary for supporting GWS actual work and functionalities in storing different types of data. It is based on the persistent layer of an integrated open-source

groupware project. The module *Persistent storage* has two basic supported functions – persistent layer for data used by the original open-source product and possible place for storing of different types of actual data related to enhanced functionality of the tool. As a persistent middleware layer the Hibernate framework is used.

The *Back-end Business Logic Layer* – the logic behind provided functionalities of the groupware component is covered by this layer, mainly management of a shared workspace, user access control to resources inside component and additionally needed management of message-based communication resources (processing of stored forums etc.). The *User Access Control* is responsible for managing and administration of the user access to groupware modules and tools. Access rights are then combined using access rights from the main interface and/or A/A (Authentication/Authorization module) and the PA ontology. The *Workspace Management* is used in a more intuitive way, where workspace will be the space for collecting people, files, discussions, calendar events important according to the defined case (business context) to support their group-work during the business process. The *Message Management* is a specialized back-end support component for pre-processing of message-based groupware data sources like discussion forums.

The *Tool Functionalities Back-end Layer* – is the layer responsible for back-end support of different tools functionalities provided to the users. Different types of communication are provided in this layer like discussion forums, e-mail or other tools for communication as well as shared calendar or file sharing, provided by the corresponding modules.

The *Interfaces Layer* – is mainly responsible for integration with other SAKE components. This means interactions with the main integrated User Interface, CMS, AMS and ontology repository as well as integration with external tools, e.g. e-mail server. It covers connections to services provided by other components as well as services provided by the GWS to other parts of the whole integrated system.

3.2. Implementation details

Implementation of the Groupware System (GWS) in SAKE is based on the open source solution Coefficient (http://coefficient.sourceforge.net/). This system has been selected after a detailed analysis. The Coefficient provides several collaboration functionalities. It is a scalable project collaboration platform that can run in J2EE and web application containers. The Coefficient uses the term *projects,* but in other contexts, projects are also called *groups, workspaces,* or *share spaces.*

The Coefficient functionality is driven by modules. Some modules, such as the project module are core of the system. Others augment the behaviour of the system. The *Project Module* serves as a workspace where individuals can collaborate. The project can be adorned with modules that enhance the functionality of the project. The Coefficient currently deploys with several add-on modules. The *Mail Forum Module* implements a discussion forum based on own framework (called Dithaka).

The GWS component in SAKE extends these functionalities with an integrated approach including semantic logging, which provides semantic information to pieces of the content; and integration with all other components of the system. Specific GWS functionalities are extended and enhanced regarding user, implementation, semantic and integration requirements to provide relevant functions in the scope of SAKE.

One of the problems of the selection process was that we have not found any other proper solution supporting all the required functionalities. The most notable example is the shared calendar, which was implemented on the basis of an external open-source server solution (Jakarta Slide WebDAV server with iCal4j Java library supporting the iCal standard).

4. SEMANTICALLY ENHANCED FEATURES OF GROUPWARE SYSTEM

According to the usage of the Groupware System (GWS) in context of SAKE we have identified several aspects for which semantics and text-mining methods are interesting (within the scope of internal public administration processes, which are mainly reflected in SAKE). The role of GWS in SAKE can be (more practically) viewed from a different perspective:

- Management of shared workspace of users in specific process context - forums, shared calendar, mail;
- Providing GWS-related services to other components;
- Providing support for GWS-related activities in the workflow;
- Providing semantically-enhanced features in order to achieve re-use of knowledge;
- Capturing the user behaviour.

Several of these aspects are interconnected directly through the usage of semantic technologies, e.g. capture of user's behaviour in a semantic way leads to semantically enhanced feature known as semantic log. All such features within GWS are provided in next paragraphs [4].

Full support/usage of semantic context within GWS

It is one of the basic assumptions for semantic-based support in the whole integrated SAKE system. Context is defined according to the current user (precisely identified user with a defined role, e.g. in the case of a small local authority Mayor can be a role, Expert role is important for experts' discussions etc.) and actual state of the current process instance (precisely defined workflow-based process with several activities and tasks, current context knows actual state in one particular process for every moment, e.g. process is 'preparing a new law about dogs/pets in the municipality, current context is 'preparing a draft version of a law'). This context is then important for:

- Preparing information, which is shown to the user – the current context (user, process) is important for access rights to GWS functionalities and data;
- Logging context-specific information into integrated system (see next feature).

Semantic log of user's activities within GWS
All activities of users within GWS are semantically logged into ontologies. It means that all the operations, like creation, editing, deleting, accessing of discussion forums, discussion threads, messages, calendar events, and so on, are logged with their current context into the logging part of the Information Ontology). This information is used by the AMS for querying and reasoning in order to achieve proactive delivery of knowledge extracted from the log. This is one way how to fulfil the goal of the agile e-government knowledge-based system.

Annotation of discussion messages
Contributions of users in discussion forums (messages inside discussion threads) are annotated in two different ways. Firstly, the set of keywords is attached as metadata to every contribution to the discussion – there is one annotation to one contribution, keywords mean free text separated by commas. If the application domain has a corresponding fully understandable ontology with a good covering concepts' vocabulary, then also domain concepts could be used for the annotation. In the project pilot application (an internal process for preparing local regulations for the local authority) we cannot have only a specific domain ontology (regulation can be about anything, we can have only a top-like ontology about regulation-like facts, structure of process etc), free-text keyword-based annotation seems to be a reasonable approach. Secondly, the relevance feedback by users to discussion messages is provided. This means that every user can add a feedback to a message representing his/her opinion on the message's relevance to the current problem. It is based on a simple scaled annotation from negative to positive feedback. Then this information can be used in other features like extraction of potential experts and search for relevant information in discussion forums.

Metadata search
In order to get some benefits from the information provision to the user, metadata search for every subcomponent (functionality) is provided. It means that search is realized as a combination of metadata searches for different information resources like documents, discussion forums, threads, and messages, calendar events, and so on.

Extraction of potential experts for new case by ranking users according to previous discussions
The statistical method for discussion forums analysis (described in section 5.1 and introduced in [3]) has been improved by means of user's feedback (scale of feedback from negative to positive) and annotations of the messages. A combination of user's feedback and filtering of threads (according to annotation of messages in threads) provides 'argumentation'-like support solution, which leads to feedback-based topic-sensitive ranking of discussion forums users. This information can be used by someone who wants to invite new members to the process (e.g. experts) to help distinguish between several users. This improved method will be more deeply described in the following chapter as our main goal for semantic enhancement of groupware system in order to achieve collaboration-based support for building of 'Communities of practice' for specific cases.

5. DISCUSSION FORUM ANALYSIS AND RANKING OF USERS (EXPERTS)

Ranking of users within the GWS can be viewed as a voting procedure in the following way: Initialising a new discussion thread represents a desire of an author to increase his/her authority and to strengthen the author's position within the community discussion space. Responding to a contribution of another author represents voting of a respondent for the author of the contribution and increasing the authority of the author (the contribution of the author is worth for the respondent to react). The used statistical approach is based on application of previous discussion analysis algorithm, detailed description and experiments can be found in [3] (For completeness, short description is provided in section 5.1).

5.1. A statistical approach to discussion threads evaluation

A discussion thread consisting of one or more contributions is a basic structural unit of a discussion group. The thread represents a particular view on the process of introducing a topic by presenting an opinion of one person and developing the topic by adding different views of different people.

An authority of group members (represented as numerical weights) is changing during their participation in discussions within discussion threads. The threads are ordered using time (publication time of root contributions is considered) as an ordering criterion. Particular threads are processed sequentially and the authorities of authors are updated based on their participation within each given thread.

From the point of representation, a discussion thread is a tree-like structure nodes of which represent particular contributions while each arc represents a relationship among two contributions, one playing the role of an initiator and the second one the role of a respondent. If a contribution responses to another contribution, it is a respondent. If it attracts at least one response, then it is an initiator. It means that the root node of a tree (consisting of more than one node) representing a thread is an initiator, leaf nodes are respondents and intermediate nodes play the both roles – they are initiators and respondents at the same time.

In order to process a discussion thread and to update weights of participants, two-step iteration procedure is used:

- Calculating popularity of contributions;
- Updating weights of contributors.

The first iteration ensures that different contributions in the thread are treated differently. The difference is given not by the content of the contributions but by the response the contributions were able to attract. Therefore two contributions are different if they are responded by different numbers of responses or the responses were produced by authors having different authorities.

In order to calculate popularity of all contributions forming a thread, a bottom-up approach is employed. Popularity of pure respondents (contributions representing leaf nodes of the tree) is set to *1*. Popularity of each initiator is based on the popularity of its respondents and authority of authors of these responding contributions. It is calculated according to the following formula:

$$y_j = \sum_{i=1}^{k} w_i y_i \quad (1)$$

where k is the number of contributions reacting to the contribution in question, y_i - popularity of the i-th contribution, and w_i - the weight of the i-th contribution's author. Firstly, popularity of leaf nodes is calculated, popularity of the thread root is determined at the end.

In the second step, weights of authors contributing to the given thread are updated. The weight modification depends on how many contributions they have authored, how popular are their contributions as well as how many contributors were attracted by the given thread. The weight of an author is updated for each his/her contribution in the thread according to the following formula:

$$w_k^{n+1} = w_k^n + \frac{y_i}{y_R}\frac{n}{N} \quad (2)$$

where w_k is the weight of the k-th author, y_R is popularity of the root contribution of the thread, n is the number of different authors participating in the thread and N is the number of all group members. Since discussion threads have different sizes and the number of group members is also not constant, two normalisations are included in the used formula – popularity of the given contribution is related to the population of the root contribution and the number of involved authors is related to the current number of all group members.

Finally, when the thread evaluation process is finished, every forum contributor is equipped with a weight value that represents his authority within the discussion forums. Having the weight value available for every forum user, it is possible to sort contributors by their weight values and to find those with the highest authority.

5.2. Semantic extension to discussion threads evaluation

When using the previous method for discussion analysis, the relevance of reactions is not guaranteed to be involved in the process of evaluation. To consider this disadvantage, it was required to extend the method with a possibility to enable discussion forums users to enter their personal feedback on messages manually to inform the system about the message quality, to express their positive or negative reaction [4]. Then authority of users is updated according to users' reactions.

5.2.1. User feedback on messages

The purpose of the user feedback on messages is to provide a solution to cope with disadvantages emerging from the previously mentioned method. Particularly, there is a threat that authority values of forum participants contributing with low quality messages accumulate as they are being frequently commented by messages whose authors posses a higher rank – a higher value in the context of a discussion forum. This accumulation of the high popularity values of messages with obvious low content quality after such statistical computation may be interpreted as very competent discussion contributions.

While having these problems in mind, we have extended the *statistical approach to discussion threads evaluation* algorithm with a third (middle) step, so the algorithm has the following form:

- Calculating popularity of contributions for a given discussion thread;
- Calculating popularity of contribution for a given discussion thread based on received user feedback values;
- Updating weights of contributors.

The user feedback is collected for every message manually in the manner, that every discussion forum user possesses a possibility of one single evaluation of any message by selecting a value from a list of predefined choices (*0 – Bad, 1 – Rather Bad, 2 – Neutral, 3 – Rather Good, 4 - Good*) that represent his/her opinion about the concrete message.

On the basis of collected user feedbacks for particular messages, new message popularity values are calculated:

$$y_j' = 0{,}25\overline{f_j} y_j \quad (3)$$

where y_j' is a new popularity value of the contribution j, averaged f_j is an average value of the user feedbacks for the contribution j, y_j is a popularity value of an original contribution j.

By using this new proposed approach the original popularity value y_j of the message j is transformed to a new value from the interval $<0, y_j>$ according to the enriched user feedback value of the original post. If there is no user feedback value for contribution j, the average value of feedbacks is set to a neutral value of 2 and with this new setting is the new popularity computed. Omitting this step for all of the contributions without the user feedback can be perceived as very competent contributions, but it might not be true.

5.2.2. Discussion threads filtering

According to a need of expert selection for particular internal public administration processes (e.g. expert group creation important in the process of preparing of new local law), it makes sense to analyze only the discussion threads related to the topic of the current process. For that reason we have designed a method for filtering topic-related discussion threads that are used as an input for *extended statistical approach to discussion threads evaluation* algorithm described in the previous section.

Every user when creating a message contribution is necessarily requested to create an annotation by specifying keywords describing the content of the message. The message content together with its annotation is saved into index (Lucene engine used for the indexing). The filtering process is based on searching for particular keywords in the index, especially in message annotation fields.

Discussion threads, which were found by performing such search, are extracted into a new thread collection and used as an input for a context-specific discussion threads evaluation.

6. SAKE PILOTS AND TESTING

The SAKE project evaluation has been performed by three public administration authorities of different size and area of activities [5].

6.1. Pilot sites and selected processes

A Hungarian pilot carried out by the Ministry of Education with a support of the Corvinus University in Budapest deals with receiving and processing the changes in legislature and recommending actions resulting from these changes, especially for the preparation of a strategy for higher education in Hungary – the selected process is "Higher Education Portfolio Alignment with World of Labour Needs".

Similar philosophy has a Polish pilot (performed by the Town of Czestochowa Town Hall supported by the "Cities on Internet" Association). The application is focused on processing the changes in legislature and the consecutive identifying of fields influenced by the change according to material resources for education institutions and their needs for repairs and reconstructions – the selected process is "Management of educational institutions' material resources".

The Slovak pilot application (performed by the Košice – City Ward Sídlisko Ťahanovce supported by the Technical University of Košice) is focused on the process of adopting of global national or European acts into local authority environment – the selected process is "Making local legal regulations".

6.2. Testing of the groupware system

Implementation and testing follows a three-phase iterative process. Firstly, the basic functionality prototype was tested. This version had only basic functionality without semantic enhancement. Also, basic functionality prototype was used for the first testing on the pilot sites in order to update user's requirements as well as some implementation details and suggestions for next steps. Secondly, first iteration of semantically enhanced version has been developed also with the re-use of the first version of the discussion forums analysis algorithm (as described in chapter 5.1). According to the first evaluation of the basic functionality prototype and experiences with the first prototype, we have finalized development of the GWS by implementation of the second prototype. All required enhancements are implemented and are fully integrated with other subsystems of SAKE. Internal testing demonstrated full potential of the system and whole SAKE system was successfully adopted in the real testing environments.

7. RELATED WORK

During the last years several related projects to SAKE could be identified. According to knowledge modelling and use of semantics (for similar purposes as in SAKE), projects like ONTOGOV (http://www.ontogov.com/), QUALEG (http://www.qualeg.eupm.net) and TERREGOV (http://www.terregov.eupm.net) used similar approaches for modelling of e-Government domain and problems using semantic technologies based on the ontologies. Moreover, ONTOGOV approach is closely related to changing environments, modelling of changes and adopting mechanisms, with context awareness (modelled using semantic business process management) in mind. Integration aspects were analysed and implemented in projects like TERREGOV, SmartGov (http://smartgov.e-gov.gr) or (knowledge management oriented) DECOR (http://www.dfki.uni-kl.de/decor). The main difference according to mentioned projects is more intuitive and coherent level of semantic integration of (potentially very different) components within SAKE architecture.

From the collaboration tools in the e-government applications we should mentioned the project Webocracy (http://www.webocrat.sk), in which the knowledge-based system for public involvement was produced (with discussions and content management system). Similar groupware-like applications to SAKE (with use of semantics and support of internal public administration processes) were produced in TERREGOV or ICTE-PAN project [6]. Additionally, our solution provides systematic way for knowledge-based support and collaborative work within workflow steps (e.g. forum is used as input/output in workflow) as well as for work outside of the workflow steps, but still within context of the current process.

During the last several years some noticeable research approaches in the field of discussion analysis have appeared. From those, which are tightly related to our ranking approach it is necessary to mention the whole area of *opinion mining* - especially the work of [7], [8] categorizing the user contributions according to the sentiment classification. In the work of [9], which is conceptually very closely related to our approach, are discussion participants ranked according to the *diffusion of influence* based on the *Influence Diffusion Model* introduced in [10]. Moreover, our solution provides ranking of users based both on context and relevance annotations and is integrally implemented in workflow.

8. CONCLUSIONS

In this article we have presented the SAKE system groupware component within the knowledge-based supporting system for public administration and their (mostly) internal processes. The core of the system is represented by an integrated knowledge space unifying different perspectives and interpretations of knowledge resources. It enables to assign metadata to each knowledge object, allowing thus more sophisticated retrieval and use. A goal of the groupware system is to support sharing of knowledge using the collaborative software as well as to help user in managing and creation of expert groups, which are used in governmental internal administration processes. We have identified and described several implemented enhancements of standard groupware solutions. One of the main contributions to semantic enhancements of the system is based on the method of

discussion forums analysis. Combination of user's feedback and filtering of threads (according to annotation of messages in threads) provides 'argumentation'-like support solution which leads to a feedback-based topic-sensitive ranking of discussion forums users. This information can be used by a decision maker who needs to invite new members into the process (e.g. experts) and help him/her to distinguish between several users in order to achieve collaboration-based support for building of 'Communities of practice' for specific cases.

ACKNOWLEDGMENTS

The work presented in the paper was supported by the EC within the FP6 IST 027128 project "SAKE – Semantic-enabled Agile Knowledge-based E-government", by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic within the project No. 1/4074/07 "Methods for annotation, search, creation, and accessing knowledge employing metadata for semantic description of knowledge", and by the Slovak Research and Development Agency under the contract No. RPEU-0011-06 (project PoZnaŤ).

REFERENCES

[1] STOJANOVIC, N. – MENTZAS, G. – APOSTOLOU, D.: Semantic-enabled Agile Knowledge-based E-government, *Semantic Web meets eGovernment*, Stanford, USA, Mar. 2006.

[2] BUTKA, P. – HREŇO, J.: Architecture of Groupware System in SAKE, *6th International EGOV Conference: Proceedings of ongoing research, project contributions and workshops*, Regensburg, Germany, pp. 301–308, Sep.2007.

[3] MACH, M. – LUKÁČ, G.: A Dedicated Information Collection as an Interface to Newsgroup Discussions, *Proceeding of 18th International Conference on Information and Intelligent Systems*, *IIS 2007*, Varaždín, Croatia, pp. 163–169, Sep. 2007.

[4] BUTKA, P. – LUKÁČ, G.: SAKE Project – Second Prototype of Groupware System, *Proceedings of the 3rd Workshop on Intelligent and Knowledge oriented Technologies,* Smolenice, Slovakia, Nov. 2008.

[5] ZELINSKY, T. – SABOL, T. – HRENO, J. – BUTKA, P.: SAKE Project: Agile Change Management in Knoweldge-based eGovernment Back-office, *National and Regional Economics VI*, Herlany, Slovakia, pp. 163–169, Oct. 2006.

[6] LOUKIS, E. – KOKOLAKIS, S.: Computer Supported Collaboration in the Public Sector: The ICTE-PAN Project, *EGOV 2003*, LNCS 2739, pp.181–186, 2003.

[7] XIAOWEN, D. – BING, L. – PHILIP, S.Y.: A holistic lexicon-based approach to opinion mining, *Proceedings of the WSDM'08 conference,* New York, USA, pp.231–240, Feb. 2008.

[8] TURNEY, P.D. – LITTMAN, M.L.: Measuring praise and criticism: Inference of semantic orientation from association, *ACM Transactions on Information Systems*, Vol. 21, No. 4, pp. 315-346, 2003.

[9] MATSUMURA, N. – OHSAWA, Y. – ISHIZUKA, M.: Profiling Participants in Online-Community Based on Influence Diffusion Model, *Transactions of the Japanese Society for Artificial Intelligence*, Vol. 18, pp. 165-172, 2002.

[10] MATSUMURA, N. – OHSAWA, Y. – ISHIZUKA, M.: Influence diffusion model in text-based communication, *In Posters of the WWW conf., 2002*.

Received January 7, 2009, accepted May 5, 2009

BIOGRAPHIES

Peter Butka received his MSc degree in 2003 at the Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics at Technical University in Košice. He studied as a PhD student at the same department (text-mining and semantic technologies). Since 2006 he is working as a researcher at the Faculty of Economics and in the Centre of Information Technologies. His research interests include text mining, knowledge management, semantic technologies, applications in e-Government, and information retrieval.

Gabriel Lukáč graduated (MSc) in 2007 at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at Technical University in Košice. Currently he is PhD student at the same department (artificial intelligence with focus on text-mining techniques for discussion analysis). Since 2007 he is working as a researcher at the Faculty of Economics at the same university. His scientific research is focusing on the areas of text mining and information extraction, discussion analysis and semantic technologies.

Marián Mach graduated (MSc) in 1985 at the Department of Cybernetics and Artificial Intelligence at the Technical University in Košice. His PhD thesis on uncertainty processing in expert systems was defended in 1992. He is an associate professor at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at the Technical University of Košice. His scientific interests are knowledge management, data and web mining, classification of text documents, information retrieval, semantic technologies, and knowledge modelling.

Tomáš Sabol received his MSc degree in 1981 and PhD degree in 1986 in technical cybernetics at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at the Technical University in Košice. Currently he is a professor of Artificial intelligence, working for the Faculty of Economics and Faculty of Electrical Engineering and Informatics. His research interests include artificial intelligence, knowledge management, semantic technologies, e-Government, and project management.