SCIENTIFIC PUBLICATIONS OF THE STATE UNIVERSITY OF NOVI PAZAR
SER. A: APPL. MATH. INFORM. AND MECH. vol. 5, 2 2013, 111-123.

An Approach to Governance and Policy Making Architectural Framework

D. Randjelović, E. Kajan, Ć. Dolićanin

Abstract: This paper proposes a novel approach to facilitate and foster e-government optimization and automation through the use of advanced information retrieval methods and techniques, and advanced Web technologies, as well. The approach suggested in this work aims to enable interactive processes that are simple, effective, and based on the user's needs and capabilities, rather than the government's organizational structure or government business models. It should create the opportunity to evaluate and eliminate redundant or unnecessary steps and processes as well as to reduce costs and cycle times by transitioning from the processes mainly based on human-related work to automated and more intelligent processes.

Keywords: clustering, e-governance, text similarity, Web 2.0 similarity, Web 2.0

1 Introduction

Nowadays e-Government is faced with society requirements to provide user-friendly and efficient services for citizens and businesses. This challenge is of particular importance in developing countries where e-Government is still in an infant phase, but also of wider importance having in mind globalization, migration, financial crisis, climate changes and other natural hazards that may happen either locally or across the globe. Regardless of the subjects of interest and reasons why people ask government for services they have rights to be precisely and accurately informed with respect to their constitutional laws (e.g. respecting minorities, people with disabilities, gender, age, etc.). The increasing number of data, regulations and policies and their updates, as pointed out in [21] is resulting by a huge number of semi-structured documents with potentially similar content given in different formats, terminology and context, including multilingual problem of such documents. The situation in Serbia is very similar. Following European Union directives, the Serbian parliament every year issues a number of new legal documents; the Government regulates many things by decrees, etc. Multilingual issues are also presented, and to have things worse, there is a

Manuscript received February 5, 2013. ; accepted May 30, 2013.

D.Randjelović is with Police Academy Belgrade, Serbia; E. Kajan and Ć. Doličanin are with State University of Novi Pazar, Novi Pazar Serbia

significant number of obsolete regulations, know in public as ”the forest of decrees”, that should be cut, but no one had it yet. Charalabidis et al. [5] emphasized another paradox that comes from an exponential increase of information available for use in socio-economic environment and governments that are failed to make sense to these data and take appropriate actions, despite many open government initiatives. Advanced e-Government and public services systems are intended to help citizens in making decisions between those are [20]: 1. Citizenry demand to participate and collaborate in local and state governments by means of suggestions and complaints submissions and they are very disturbed if are not satisfied with received answers and if they haven't received answers or they are not under theirs volition. 2. Extraordinary situations were, are, and will be parts of our lives, regardless if they appear in local, wider environments or planet-wide. They may appear in various areas of social, biological, economic life, etc. Examples include but not limited to: fires, floods, storms, earthquakes, various medical and biological risks such as urgent assistance and medicine, epidemic or terrorist's attack by biological weapons, than financial crisis, market unstable, etc. All of the above have their own sources and consequences, but they have several things in common. All are dangerous and people and community must be able to foresee such situations and must be prepared to act on time and in proper ways. To take care and act about aforementioned situations is under responsibility of the local or state government. In many situation they are exchanged information with other authorities either using direct communication and/or via some common global body, for example World Health Organization, in case of well-known bird flu. Since the early days of computers various government and scientific organizations collected relevant data, analyzed them and tried to predict what may happen and how to protect particular community. Thinking about that now, we have in mind a lot of available technological enablers. They are, but not limited, to: Internet and Web that lead to build the efficient and cheaper government systems (known as E-Government), power of new hardware technologies that are able to store terabytes of data (regardless of their nature, pure text, figures, tables, pictures, code, etc.) exchange them via very reliable and high speed channels, and make complex calculations over that data either on source or destination computer, etc. The rest of the paper is organized as follows. Section two gives background on e-government research and objectives of proposed framework. Section three is dedicated to technological enablers, especially Web 2.0 that will be used in development and deployment of the architectural components, while section four is dedicated for framework description. The paper is ending with concluding remarks, including future research.

2 Background and objectives of the framework

In 2003 OECD defined E-Government as ”better government” that enables higher quality services, greater engagement with citizens and better policy outcomes [18]. Since then many research studies have done in order to define what ”better” means. Ndou [17] emphasized paradigm shifts in public service delivery between bureaucratic- and E- Government

and opportunities that E-Government may bring to society such as improvement of the quality of decision making, as well. These shifts may be done by different steps starting with bilboarding or partial delivery of services to an interactive democracy [26]. The achievement of interactive democracy with vertical integration of local and statewide institutions and horizontal integration across different functions is the main objective of the framework proposed. A vision of such an objective has been recognized by four stage model for fully functional E- Government [13], but existing E-Government solutions in developing countries are far away from that. The project aims to research and develop prototype solutions for Web applications that allow automated interaction with users in terms of answering the questions, based on existing knowledge base and unstructured data in the form of available documents. There are three groups of users: citizens, government, and business whose communications with system are shown in Figure. 1.

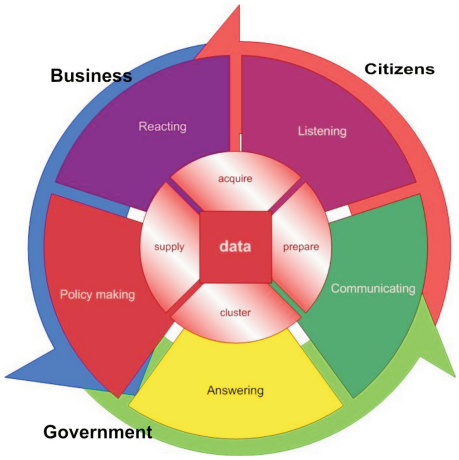

Fig. 1. Framework objectives

The idea behind objectives is to have accurate data that are acquired from different sources, prepare, cluster and supply them to citizens or business in interactive communication listening user's demands and reacting on time. The project aims to meet the needs of the following target groups:

" Citizens, especially young people and minorities in EU that originated from Western Balkan countries whose natural languages are formerly known as Serbo-Croatian language. " The government and its ministries (Ministry of Economy, Ministry of e-Government, Ministry for EU Integration, Ministry of Internal Affairs, city and local governments); " Specialized government departments (Department for protection against natural disasters, Department of Health Protection, Data Protection Department); " The institutions of the European Union; " Universities and research institutes; " Small and medium enterprises (SMEs); " Professional association of economic and social actors (Chamber of Commerce, Economic Association, The Law Society, Medical society, The Society for Standardization, Epidemiological Institute, etc.); " NGOs. In order to meet these objectives, we have extended current research project [9], [15], [23] that is concentrated on questions and an-

swers clustering system only for a particular Serbian government institution, to an EU wide project for e-governance and policy framework based on Web 2.0 technologies, that will support various e-Government applications intended to meet needs of aforementioned target groups.

3 Challenges and opportunities of human networks

Over a decade we are facing up with rapid growth of new Web technologies that in turn bring new opportunities for development and deployment power applications for various purposes. These technologies are usualy summarized as Web 2.0 and beyond referenig to human, service, semantic, intelligent networks and networks of things [10]. Human networks rely on so-cold Web 2.0 technologies. These include blogging, crowdsourcing, folksonomies, Mashups, social networks, RSS feeds, etc. The power of Web 2.0 does not rely on its software tools and techniques, but on data that Web 2.0 may bring within [14]. Key Web 2.0 technological enablers here that will be used to update system with accurate data are social networks, crowdsourcing and Data Mashup. Term crowdsourcing has been coined by Howe in June 2006 [7]. It may be defined as an open call that a company or institution broadcast to network of people (a large group, by default) in order to find the best way to accomplish some task. Following that idea Daren Brabham summarized crowdsourcing idea to a model that leverages ingenuity and aggregates talent in solving problems with reduced time and costs formerly needed for it [4]. The main problem in crowdsourcing is that there exist so many solutions but there are no harmonized concepts and interfaces for the integration of all this activities. This opens a broad field for future research. Kajan and Dorloff [10] summarized that into several categories: technical, sociological, quality, governance and ethical. If we concentrate on technical challenges only, the standardization of crowdsourcing protocols, security and governance are of particular importance, and modeling human computation in existing business processes, as well. For later, an ongoing research on weaving social processes into business process modeling might be of importance for wider research community [11]. In that project authors defined actors (tasks, people and machines) that participate in business process execution and their potential social behavior during the execution of business process. In addition they defined and developed a system, based on BPMN 2.0 (Business Process Modeling Notation) standard that allows business process designers to model business processes according to social behavior of actors involved. Security issues are not goal of this paper, but it will cover the framework described in section 4. The challenges and issues of security of such a framework are discussed in [19]. Another Web 2.0 technology of particular importance for the approach given in this paper is Data Mashup [6], [29]. It relies on the SOA (Service Oriented Architecture) paradigm and offers a new class of data integration ”on the fly”, even in real time. In addition Mashup is used not only for data integration, but also for ”glue” these data into entirely new application. There are yet open obstacles and research issues that should be overcome in Mashup. These analyzed in many papers, but summarization here is

based on two concise based on facts from practice [3], [8]. The main obstacle is behind the nature of data sources, i.e. information often resides in a mix of relational database, applications, legacy systems, and to have it worse these data are not mutually accurate. The latest is one of the main obstacles in data integration and application interoperability, in general [8]. The other obstacles are either procedural or programming. From the procedural point of view an average domain expert is almost unable to handle all ICT (Information Communication Technologies) procedures and understand them in order to build new and valuable information or application for the manifold existing sources. The similar circumstances are valid for an average Web programmer familiar with basic client-side programming, e.g. html, css, xml and some of scripting languages and with some server-side programming e.g. php or Python, etc. Dealing with Mashup programming requires much deeper programming expertise including workflows, Web services, AJAX, semantic technologies, etc. The Web 2.0 is yet a network platform that offers a lot of information exchange and collaborations possibilities via connected ICT systems and devices. These possibilities and power of data that may be acquired by human networks may be and should be further combined with power of key features of other advanced Web technologies. Examples include, but not limited, to research on identifying and locating people in case of disaster such as earthquake using Data Mashup from various sensors [1], [2], incorporation of crowdsourcing in an emergency response system [22], analyzing microblogs posted via Twitter in real-time hazard situation occurred in the US during spring of 2009 [27].

4 Architectural framework overview

The idea behind architectural design of desired framework is based on three key design principles: distributed, plural and collaborative, but at the same time allowing government institution to aggregate and use all collected information in data warehouse, mixed them with existing documents, and harnessing a collective knowledge. A general architectural overview of framework is given in Figure 2. It consists of four modules: (1) Data acquisition/supply; (2) Communication with users; (3) Document preparation and (4) Clustering that all together makes DSS (Decision Support System) capable to support citizens, businesses and government people in making decisions, either in normal life and business or in hazard situations. DSS described here will be used as mechanism for helping decision processes in situations that may be classified into two basic groups: a. Decisions would be used for reasoning based on known cases (e.g. in applications that support suggestions and complaints from citizens) b. Decisions would be carried out using an inference from the known groups of algorithms of machine learning (e.g. in all applications where choices from available options are exist). Both aforementioned cases share the common scenario based on questions issued by citizens and answers provided by system. Questions and answers are accumulated over time into some kind of knowledge base. Interactions between citizen and systems take place through three phases: (1) system offers set of keywords; (2) system offers set of questions that are closely related with chosen keyword and the possi-

bility for citizen to put new question; and (3) system is either answers the question, or finds the most similar question/answer pair or makes a new one.

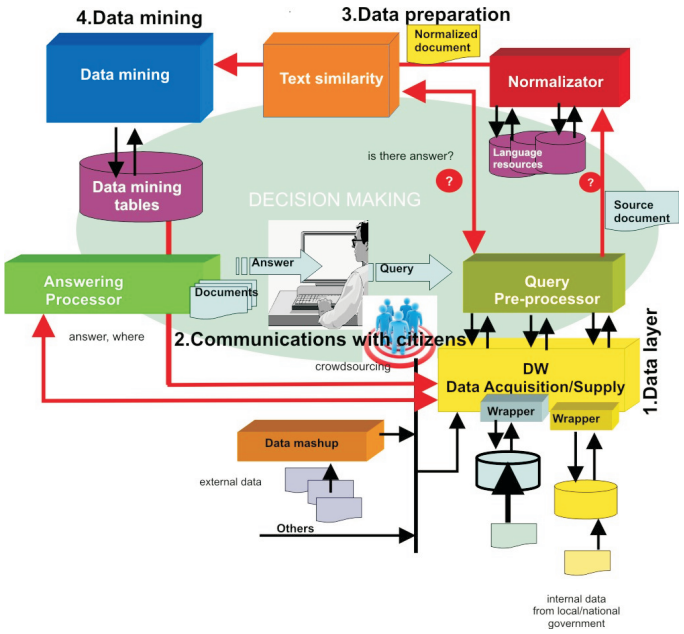

Fig. 2. Architecture of the framework.

4.1 Data acquisition/supply and communication with citizens

Written text is a basic way to store human knowledge. Today the information and documents have multimodal character, but the most of the content is in text form. In this DSS, DW (Data Warehouse) is intended to collect data from different sources including but not limited to, local and state government data, frequently asked questions, (internal data), and external data such as crowdsourced information, data provided by Mashup, etc. The role of DW here is manifold. It allows aggregation of all data that may help citizens and business in making decisions, as well as to guide government in policy making. It also allows switching from automatic to semi-automatic discovering of needed data. Questions are issued by citizens and proceed to DSS via Query Pre-processor as shown in Figure. 2. If there is an answer that already exists in DW, it returns to citizen who asked it via Answering Processor. In order to improve flexibility, there are different communication scenarios implemented in the system. Regardless of which one is used, the interaction starts when the citizen chooses the category of interest and either selects one of offered questions (list of frequently asked questions), or enters the new one. Both of cases include finding of already answered question; in the first case this is performed manually (by citizen) while in another case the system tries to find the question which is similar 'enough' to the citizen's one. The

most desirable situation is when such a question is found and the system sends its answer as a response. If the similarity condition is not satisfied, the system continues with searching of documents. It this case, similarity is measured by comparing documents' metadata such as titles (table of content), bookmarks, index tables, key words and abstracts with the citizen's question. If such a document is found, the system responds with it. Citizens can express their satisfaction with the responses by giving marks. This data are used by the system to enforce the 'importance' of answers or documents within the clusters. The clusters are formed for improving the system performances: grouping of answers or documents based on similarity reduces the time and improves the quality of response. There is another option if the similarity condition is not satisfied in any case, or the citizen is not satisfied with the any response. Than the question is forwarded to the subject matter expert and default response is sent to the citizen (e.g. the answer will be sent after the certain time). After the new answer is created, the system connects it to the forwarded question and determines which clusters the question belongs to.

4.2 Data preparation

In data preparation module two functions play key roles: text normalization and text similarity measurement. In order to provide some background for novice readers on such issues, the following few paragraphs explain key issues and challenges here, followed by specific issues on Serbian language. Text encoding is primarily related to the structural (paragraphs, sentences, headings, etc.) and analytic (grammatical categories, syntactic categories, etc.) styles of the written text. Text documents that have no any additional information are called the unmarked texts. By contrast, the marked text contains a variety of structural types (title, paragraph, sentence, etc.) and linguistic (grammatical category, syntactic structures, etc.) information. Many features of the text would be difficult to determine and processed using a computer without such explicit information (i.e. labels). Applicability of the text is substantially increased if it is annotated. The main types of marking the text are: segmentation of the sentence, tokenization, part of speech tagging, lemmatization, parsing, and semantic tagging. The basic assumption is there three types of words that carry semantic value of a sentence. These are nouns (e.g. sto (a table)), verbs (e.g. govoriti (to speak)) and adjectives (e.g. plavo (blue)). Other words of minor importance (e.g. contributions, suggestions, conjunctions, exclamations, etc.) are known as stop words. Clustering, information retrieval and contextual responding represent the main background system functions in the services that e - government offers to the citizens. Those functions can be performed automatically, by using different algorithms and methods. Regardless of the text size, the information content is just a part of the text body. The sentences contains lot of non - information content such as pronouns, prepositions, adjectives, adverbs, phrases, connection words and other language- and grammar- dependent elements and forms. Furthermore, the documents might be presented in different formats (e.g. html, doc, pdf, ps, etc.). All aforementioned lexical and formatting text elements represent the information noise which disturbs the processes such as clustering or similarity measuring. Dealing with

language and grammar issues in a particular language, in this case Serbian language, and various macros that may be embedded in documents is not subject of this paper. In fact, identifying and cleaning all those elements, and taking care about grammar rules of the language is a challenging task, especially in the language like Serbian where seven cases and a number of verbal tenses are in use, as well as various consonant alternations, genitive plurals, aspectual pairs of verbs, conjugation, transitive and intransitive forms, etc. Language resources are the sources of language texts, which consist of corpus and dictionaries in the digital form, i.e. in the form of electronic text. Typical document-centric representative documents are linguistic resources. Morphological electronic dictionary of modern Serbian language is one important resource. This is not a dictionary to read but is intended for automatic processing of text and is significant for the linguistic and lexicographic research because it provides real-time accurate data. We have developed a php routine that parses the document sentences and words, normalizing document and cleaning it from stop words. Such a normalized document is going further to text similarity module where automatic keyword extraction plays important role. Keywords are index terms that contain most important information. Automatic keyword extraction is the task to identify a small set of words, key-phrases or keywords from a document that can describe the meaning of document. Keyword extraction is considered as core technology of all automatic processing for text materials. By extracting appropriate keywords, we can choose easily which document to read or learn the relation among documents. Existing methods about automatic keyword extraction can be divided into four categories [12]. Simple statistic approach includes word frequency, word co-occurrence, tfidf (term frequency, inverse document frequency), etc. The linguistic approach uses linguistic features such as lexical or syntactic characteristics, whilst machine learning approach uses extracted keywords from training documents in order to learn a model and then applies it to find keywords in new document. The other methods are usually combination of aforementioned approaches with additional helpful features such as position. The idea of measuring position weight comes behind the fact that words in different positions carry different entropy. If the same words appear in introduction and conclusion of a document, for example, it is more likely that these words are more important than the others in document [25]. It uses three important elements: paragraph weight, sentence weight, word weight. This approach required documents to be marked, in xml, for example. Another way is to use an algorithm for indexing based on tfidf measure, which extracts keywords frequently appear in a document, but don't appear frequently in the remainder of the corpus [16]. In this case, keywords extraction requires to having corpus with similar text, and it does not work for single document since there are no other documents to compare keywords to algorithms, so it will choose keywords based on term frequency. In our system we use aggregated set of keywords for a domain, made by predefined set of them and than updated during the system exploitation. Comparing new document with all documents that already exist in corpus may be too long. As such, a keyword extraction without a corpus with the same kind of documents is very useful. Two terms in a sentence are considered to co-occur once. We can obtain frequent terms by counting term frequencies. Dealing with such language and grammar issues, regardless of

which particular language they belong to is a complex task; identifying and removing of all these elements and taking care about grammar rules in order to find generic form of the particular word or complete sentence, represent some of the challenges in clustering of text portions in the language like Serbian. Moreover, there can be several cases and a number of tenses, as well as various consonant alternations, genitive plurals, aspectual pairs of verbs, conjugation, transitive and intransitive forms. All that is needed to be considered in order to prepare text portion before clustering.

4.3 Clustering

A hybrid solution for multilayered document clustering based on Fuzzy C - mean algorithm (FCM) is proposed in the solution [23]. The clusters are formed by predefined set of keywords from the domain vocabulary or by aggregated set of keywords as explained in section 4.2. The citizens' questions and formal documents represent the objects of clustering process. They belong to the separate layers due to their difference from each other. The most obvious difference is in size: the questions' length is usual one or more sentences while the documents usually consist of several pages. They also differ in structuring: the documents usually have hierarchical structure while the questions are unstructured - simple flat sequence of words (sentences). The last one, but not the least important is the difference in meaning: questions represent input values for the system while the documents represent its possible outputs. Because of these three main differences, questions and documents (hereinafter referred as text portions) are separated from each other. On the other hand, text portions are clustered in the same way. They belong to the clusters according to similarity of their body (words, phrases) with the clusters' key terms. Due to lexical and formatting text elements they consist of and represent the information noise which disturbs the clustering processes, text portions are not processed in original form. They pass through text preparation module first. The clusters formed by K - means algorithm [24] are fully separated from each other. It means that there are no mutual intersections and the text portions which would bee clustered belong to only one cluster by default. In contrast, FCM algorithm enables that certain text can belong to more than one cluster. Therefore, FCM is recognized as more flexible solution which enables more generic approach to deal with the different language issues. Eq. (1) represents modified FCM clustering function used in the proposed solution. It includes fuzzy membership function m_{ij} of particular text portion x_i as a measure of its belonging to the particular cluster. In such a formal expression, there are K clusters and N text portions which mean that every particular question or document xi has K membership functions m_{ij} .

$$f_{fcm} = \sum_{k=1}^{N} \sum_{j=1}^{K} m_{ij} \cdot (x_i) \tag{1}$$

Multiplicative nature of FCM function expresses the fact that every text portion belongs to every cluster in some degree. According to fuzzy set theory, used membership function

is normalized in range 0 to 1. If text portion xi has $m_{ij} = 0$, it practically means that xi does not belongs to j cluster at all. If text portion contains r key terms (e.g. it is typical for documents) than it belongs to r clusters in some degree defined by membership function. In the proposed solution, the membership function $m_{t,q}$ (where t represents clusters' key term and q represents the text portion) is calculated (2) by using both of term frequency - $tf_{t,q}$ and inverse document frequency - $idf_{t,q}$ [28]. The term frequency $tf_{t,q}$ represents the internal characteristic of the certain text portion - how many occurrences $f_(t,q)$ of specified term t it consists. In the other hand, inverse document frequency $idf_{t,c}$ represents outside property - it is a ratio of number of text portions which contain specified term t $N_{q,t,c}$ and full number of text portions in the whole corps $c(N_{q,c})$. The product of these two functions, commonly called tfidf is used in the calculation of membership function.

$$m_{tq} = k \cdot tf_{tq} \cdot idf_{tc} = k \cdot \log(f_{tq} + 1) \cdot \log(\frac{N_{qc}}{N_{q,t,c}}) \tag{2}$$

The membership function is normalized in range from zero to one due to using logarithm functions. Value of $f_{t,q}$ is incremented by one in order to avoid logarithm of zero. Finally, coefficient k represents correction factor which provides the better dispersion of the values of $m_{t,q}$ in the range. This coefficient is calculated as reciprocal of tfidf function maximum. After the initial clustering the system is ready for use. Every new question will be processed in the same manner: firstly the system determinates which clusters the question belongs to; after that it measures the similarity of the new question with the answered questions in these clusters in order to find the most appropriate answer(s) or document(s) for responding.

5 Conclusions and future work

The information content which is held in the e - government systems are represented in different forms. Documents, citizen questions and answers, provided by subject matter experts, represent three main forms. The advanced searching engine, which is enabled to provide high quality of response, is based on grouping and filtering of the content. There are semantically relationships between questions, answers and documents formed for this purpose. Moreover, they are vertically layered by content type and horizontally clustered by similarity with key terms. Hybrid solution represented in paper is mainly focused on providing conditions for advanced responding to the citizens' requests. The content is clustered by using modified FCM algorithm that provides flexibility of the system. The boundaries between clusters do not exist and pieces of information can belong to more than one cluster. This way the citizen's question can be similar to more than one question or document and system is able to deliver more than one result in response. Moreover, the relations between questions, answers and documents are changeable and depend on the citizens' satisfaction with the response. One question can be related to more than one answer and vice versa. The priority in response is determined by the relation strength. The future work will be

directed on development of multilingual support. Finding a way how to use already defined domain vocabulary (e.g. written in English) for automatic translation into target language by using existing services represent one of the goals. One of the challenges is how to use existing domain vocabulary for generating the new one by using the third language (e.g. English can be used as a language for common representation). Using of multimedia content in citizen services also represents one of the goals. In addition, future research will also focus on a user needs analysis to identify what needs to be done and to develop a technical model architecture specification that is based on clouds, Web.2.0, social networks, sensors and Internet of Things.

6 Acknowledgments

This paper has been supported by the Ministry of Education, Science and Technological Development Republic of Serbia from the project III44007.

References

[1] N. BESSIS, E. ASIMAKOPOULOU and F. XHAFA, *A next generation emerging technologies roadmap for enabling collective computational intelligence in disaster management*, Int. J. Space-Based and Situated Computing, Vol. 1, (2011), 76-85.

[2] N. BESSIS et al., *Next Generation Technology Victim Location and Low Level Assessment Framework for Occupational Disasters Caused by Natural Hazards*, Int. J. Space-Based and Situated Computing, Vol. 2, (2011), 45-53.

[3] A. BOUGUETTAYA et al., *End-to-End Service Support for Mashup*, IEEE Transactions on Services Computing, Vol. 3, (2010), 250-263.

[4] C. D. BRABHAM, *Crowdsourcing as a Model for Problem Solving: An Introduction and Cases*, The International Journal of Research into New Media Technologies, Vol. 14, (2008), 75-90.

[5] Y. CHARALABIDIS, *ICT for Governance and Policy Modelling: Research Challenges and Future Prospects in Europe*, In:Proceedings of 45th Hawaii International Conference on System Sciences, Hawaii, 2012, pp. 2472-2481.

[6] G. DI LORENZO , H. HACID and H. Y. PAIK, *Data Integration in Mashups*, SIGMOD Record, Vol. 38, 2009, 59-66.

[7] J. HOWE , *he Rise of Crowdsourcing*, Wired, Vol. 14, 2006.

[8] E. KAJAN, *Interoperability issues of business processes- Key issues and technological drivers" in Encyclopedia of e-business development and management in the global economy, In Lee, I. (Ed.) pp. 908-917*,IGI Global, Hershey, PA, 2010.

[9] E. KAJAN et al., *Normalization of text documents in Serbian language for efficient searching in E-Government systems, in Serbian*, In:Proceedings of ETRAN International Conference, Zlatibor, 2012, pp. 1-4.

[10] E. KAJAN, F. D. DORLOFF and I. BEDINI, *Handbook of Research on E-Business Standards and Protocols: Data and Advanced Web Technologies*,IGI Global, Hershey, PA, 2012.

bibitemKajanetal E. KAJAN, et al., *Social Approach to Design Business Processes", submitted for publication*,IGI Global, Hershey, PA, 2013.

[11] K. JASMEEN and G. VISHAL, *Effective Approaches for Extraction of Keywords*, International Journal of Computer Science Issues, 7(6), (2010), 144-148.

[12] L. KAREN and L. JUNGWOO, *Developing fully functional E-Government: A four Stage Model*, Government Information Quarterly, Vol. 18, No 2, (2001), 122-236.

[13] L. IN, *Overview of Emerging Web 2.0-Based Business Models and Web 2.0 Applications in Business: An Ecological Perspective*, International Journal of E-Business Research, Vol. 7, (2011), 1-16.

[14] U. MAROVAC , E. KAJAN and G. SHIMIC, *A solution of semantic clustering of text documents,Book of Abstracts*, In:Proceedings of CPMMI International Conference, Novi Pazar, 2012, pp. 59.

[15] M. YUTAKA and I. MITSURU, *Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information*,International Journal on Artificial Intelligence Tools, Vol. 13, (2004), 157-169.

[16] V. NDOU , *E-Government for Democratic Countries: Opportunities and Challenges*, The Electronic Journal of Information Systems in Developing Countries, Vol. 18,(2004), 1-24.

[17] OECD, *The e-Government Imperative, e-Government Studies*, 2003.

[18] M. RANDJELOVIC , D. RANDJELOVIC and M. SUKNOVIC,, *An Approach to Intelligent Web-Based Multi-attribute Decision Support System*, In:Proceedings of 19. Telfor International Conference , Belgrade, (2011), pp. 1514-1518.

[19] D. RANDJELOVIC , M. BAJAGIC and B. CAREVIC,, *Internet as a function of terrorism*, In:Proceedings of International Conference Internet as a function of terrorism, crime prevention and European integration with emphasis on Cyber crime, Banja Luka, (2012), pp. 317-330.

[20] K. RAO and S. DAY, *Decision Support for E-Governance: A Text Mining Approach*,International Journal of Managing Information Technology, Vol. 3, (2011), 73-91.

[21] J. ROGSATIDIUS et al., *Towards Real-time Emergency Response using Crowd Supported Analysis of Social Media*,In:Proceedings of CHI 2011 International Conference , Vancouver, (2011).

[22] G. SHIMIC et al., *An Approach to Document Clustering using Hybrid Method*,In:Proceedings of IADIS e-Society Conference, Berlin,(2011), 153-159.

[23] S. VIEWEG et al., *Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness*,Proc. of CHI 2010, Atlanta, GA, US,(2010), 1079-1088.

[24] C. WARTENAand BRUSSEE, *Topic Detection by Clustering Keywords*,In:Proceedings of Proc. of DEXA'08 International Conference ,(2008), 54-58.

[25] D. WEST, *E-Government and the Transformation of Service Delivery and Citizen Attitudes*,Public Administration Review, Vol. 64,(2004), 15-27.

[26] H. C. WU et al., *Interpreting TF-IDF term weights as making relevance decisions*,ACM Transactions on Information Systems, Vol. 26,(2008), Article 13.

bibitem Yu J. YU et al., *nderstanding Mashup Development*,IEEE Internet Computing,(2008), 44-52.