



## M-GOVERNMENT APPLICATION INTENDED TO SEARCH DOCUMENTS WRITTEN IN SERBIAN LANGUAGE

Dzenan Avdic<sup>1</sup>, Aldina Avdic<sup>1</sup>, Žaklina Spalević<sup>2</sup>, Ulfeta Marovac<sup>1</sup>, Adela Crnisanin<sup>1</sup>

<sup>1</sup>State University of Novi Pazar, Serbia

<sup>2</sup>Singidunum University, Serbia

### Abstract:

Rapid improvement of mobile communications and mobile computing technologies has opened new opportunities for mobile interaction and mobile business. The use of mobile technologies in the public sector represents an alternative channel of communication and public services, but more importantly, it makes the traditional e-Government personalized, localized, context-aware and closer to the citizens. This paper presents one solution of m-Government application designed for finding answers for citizens questions, asked in Serbian language. According to this, it describes how to overcome the problems caused by legislation and also by the specific features of the Serbian language and, as well as connection with existing e-Government application.

### Key words:

m-Government,  
e-Government,  
document search,  
Serbian language,  
legislation.

*This work was partially funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia by the projects III-44007 and TR-35026*

## INTRODUCTION

M-Government is an integral part of the e-Government, which expands the range of its capabilities. E-Government represents the usage of information and communication technologies in order to improve the activities of public sector organizations, and m-Government is used to make public information and management services available “anytime, anywhere” to citizens and employees in the public sector. One of the reasons of the m-government existence is the fact that mobile services are cheaper and more accessible [1]. M-Government is particularly suitable for countries those are in transition, where rates of Internet access are low, but the use of mobile phones is growing rapidly in urban and rural areas. Globally, the number of mobile phones has surpassed the number of fixed phone, as it is the case in Republic of Serbia.

A goal of m-Government is not to replace e-Government; it is used to complement e-Government function-

alities. E-Government uses information technology such as WAN, Internet, and mobile computing in order to put public sector jobs at the service of citizens. On the other hand, m-Government uses mobile and wireless technologies such as mobile phones, PDA devices with Internet connection for improving government services and its accessibility to people anywhere and anytime [2].

Although mobile applications are based on good ICT infrastructure, on using networking technology, database administration, procedures for recording transactions and so on, they represent the tip of an iceberg, because they serves as the final information channel to citizens, with a complex infrastructure behind.

The goal of this paper is a presentation of the application for m-Government with the purpose of automatic finding answers for citizens questions, related to the public sector, followed by a description of its architecture and highlights the benefits of the existence of the m-Government.



The rest of the paper is organized as follows. The second chapter presents the basic principles, advantages and disadvantages of the m-government. Next chapter describes legislation related to the e-Government in Republic of Serbia, and then the components of the architecture of the e-Government application, which functionalities are extended by the m-Government application are shown. Finally, the results, conclusions, and trends in future research are given at the last chapter.

## ADVANTAGES AND DISADVANTAGES OF THE M-GOVERNMENT

Though, in the previous chapter, m-Government is mentioned in terms of improving the functionalities of e-Government, despite numerous advantages, it carries certain restrictions. Most significant advantages and disadvantages of m-Government are given below [3].

### Advantages:

- ◆ Overcoming spatial and temporal restrictions - the main advantage of m-Government is ability to overcome the limitations of e-Governance, as it is available anywhere and anytime;
- ◆ Increasing the productivity of personnel in public institutions - m-Government reduces necessary time for the actions of employees, because important information are carried with them and they don't have to spend time on activities such as recording data on paper, then going to the database and entering the data, but the operations can be performed in real time;
- ◆ Improving of information and services availability - a consequence of the first advantages is the fact that the use of m-Government information may be submitted to the citizen no matter where he is;
- ◆ Reducing the cost of communication - this involves informing citizens using mobile communications, e-Voting and etc.

### Disadvantages:

- ◆ Unavailability to the poorer part of population
- ◆ the anonymity of users and the use of cells for entertainment - mobile phones are usually unregistered, which can lead to the data abusing, and since cells are used so far more for entertainment purposes, it is difficult to convince people that cells can be used for doing serious work related to the public sector;
- ◆ (In) security of e-Payments - the ability to intercept transactions;
- ◆ Information overloads - the existence of users who are always involved;
- ◆ A wide range of mobile platforms - it is necessary to design a variety of clients in order to cover a lot of platforms, which produces additional development costs.

## LEGAL REGULATION OF E-GOVERNMENT IN THE REPUBLIC OF SERBIA

The Strategy for e-Government Development in Serbia is contained in the Strategy for Development of Information Society in the Republic of Serbia to the 2020<sup>th</sup> year, in the National Strategy for Sustainable Development and the Strategy for e-Government Development in the Republic of Serbia for the period since 2009<sup>th</sup> to the 2013<sup>th</sup> year[4].

Strategy of Information Society of the Republic of Serbia to the 2020<sup>th</sup> that provides reform and modernization of public administration based on the widespread use of information and communication technologies is one of the key elements of the overall transition of the Republic of Serbia in the modern information society [5].

According to the National Strategy for Sustainable Development in the Republic of Serbia is necessary to support increasing the level of digital literacy among citizens and introducing a quality education in terms of information and communication technology is at the elementary school level. It should provide conditions for further increase in the number of Internet users and the accessibility of ICT to all, both physical and legal entities [6,7].

Strategy for Information Society Development in the Republic of Serbia for the period since 2009<sup>th</sup> to 2013<sup>th</sup> year, which is in complete coordination with the basic policies of the European Union in the field of e-Government, went from state regulation in this area and the results achieved in the implementation of the legislation [8].

Legislation of the Republic of Serbia adopted during the 90s was conditioned by level of technological development and standards in the field of information technology. The current level of technology of information and communication technologies, the development of modern methodologies, tools, orientation to the object oriented modeling, Web technologies, distributed systems, new operating systems and new generally accepted standards require re-engineering of existing regulations. In addition to technology, legislation passed during the 90s is outdated and from the perspective of current social demands for the implementation of the concept of e-government from the point of harmonizing our legislation with EU and international standards.

Laws of the Republic of Serbia, who passed in recent years and which are consistent with the European Union, specifically refers to the Law on Business Registration Law on Access to Information of Public Importance and the Law on Electronic Signatures and the relevant by-laws, comprise the essential elements concept of e-Government, such as the introduction of electronic signatures and digital certificates, the possibility of filing of the application of natural and legal persons (beneficiaries) electronically providing customer service via the Internet, the communication of users and authorities by e-mail, sanctioning negligent and malicious acts, etc.

Among the legal provisions of the Republic of Serbia, the most significant is the implementation of the Law on Electronic Signatures (adopted in December 2004).



In this way, citizens in 29 cities of the Republic of Serbia has enabled the post offices submit their application for the apparatus which produces an electronic signature. It is expected that it will soon be possible in other branches throughout Republic of Serbia. After completing and signing the form, the citizens in two or three days, on the home address post delivers certificate, card, card reader and software that must be installed on users computer. Separately, to prevent misuse, a personal identification number, called a pin, is sent, which is typed in software when signing of documents is done by electronic signature, such as e. documents created in Microsoft Office suite.

Serbian Post has an extensive infrastructure of about 1,200 automated and real time linked branches. In addition to the signatures, in the near future, in post offices, citizens will be able to get a birth certificate, citizenship certificate and other documents. This would still mean that, after the entrenched use of electronic signatures, basic, legal prerequisite to move from paper to electronic processes in all areas of life, is now filled. That means, in practice, the one of the basic preconditions for the further development of e-government in the Republic of Serbia is achieved [9].

## M-GOVERNMENT APPLICATION - SERVER SIDE

M-Government application should help the citizen to get the answer to his question at any time, quickly, via cell phone. There are three types of content in such system: questions, formal documents and expert answers. Questions and documents are grouped based on keywords. When user asks a new question, its similarity with existing questions and documents from the selected group is counted. Preparation of documents for processing, normalization, extraction of keywords, clustering documents and questions and finding answers to this question are features implemented on the server-side of the application, and represent specific components of the system. Figure 1 depicts working of such system, from first aspect, when administrator inserts documents into database, and second, when citizens ask questions. System components are described further in this section.

Preparation of the document is executed when a new document (or question) is inserted in the system. Before the contents of question or a new document are saved into the database, they have to be normalized. In order to normalize the document or question, it is necessary to prepare them for normalization and it is done by system component named Query Pre-processor. Preparation of documents includes: changing document format, removing redundant and informal character and structuring of documents by the rules corresponding to the next step - the normalization. Preparation of the document, in this case, includes the following steps: processing of documents in Cyrillic and Latin script, the processing of documents in HTML format, removing unnecessary characters (emojicons, slang ...), removal of stop words, switching between the ASCII, UTF8 and Latin script. Finally, the document after preparation should be in Latin script in UTF8 format.

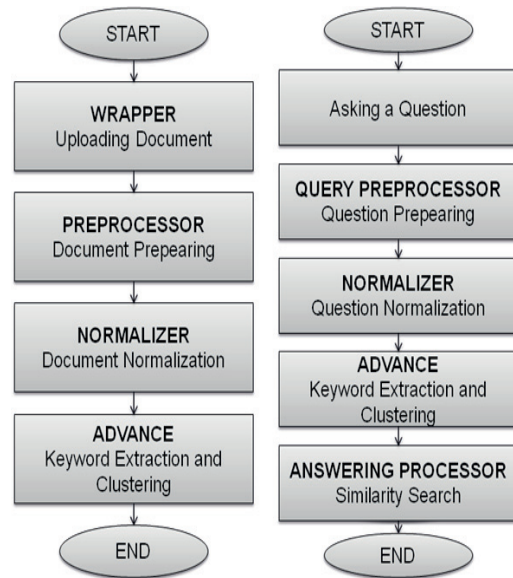


Fig. 1. E-Government server side: left) components used in inserting an document; right) components used in asking the question

Text normalization is the transformation of text into another form that is suitable for any type of computer processing. In our case it is a search. The purpose of the normalization of the word is elimination of redundant modification of words that do not make changes in the meaning of the word, i.e. it is the reduction of these modifications on the common, basic form. This transformation is performed by component named Normalizer. Normalization can be done by: lemmatization of word - elimination extensions and formative suffixes and reducing words to lemma, ripping the longest found suffix for certain type of the word, by separation of the first k letters of the word, by n-gram analysis of words and so on. Each of the above normalization has its advantages and disadvantages. Some requires specific lexical resources (morphological dictionary, extensions to form words [10], which can make processing harder, and the other does not solve complex derivation, prefixes and other grammatical peculiarities. We chose n-gram analysis, as it does not require lexical resources, and solves the problems of prefixes, suffixes and other grammatical peculiarities. It is a procedure that is applied to the text, and its result is to obtain a set of n-grams of a certain length. N-grams are obtained by moving the frame length N, which can start from positions 1 to  $m - n + 1$ , where m is the length of the string [11]. In this paper, the 4-gram analysis is show, as it gives the best results for Serbian. For example, word uraditi has these 4-grams (urad, radi, adit, diti). This would increase the number of data in the database, and therefore the processing time of the request. This type of n-gram analysis is, in some languages, more meaningful, because of finding phrases. This problem was solved in another way, by computing the matrix of mutual occurrence of words.

Grouping of documents is based on the similarity of contents. Each document is represented by the keywords that carry the meaning of the document. Extracting keywords and clustering documents is done by a component named ADVANSE. Automation of extracting keywords is the process of extracting a small set of words and phrases





that describes the content of the document. Extracting keywords in documents also requires the corpus of documents that have similar contents using algorithms for indexing  $tf * idf$ . In our case, the content of inserted document is unknown, and corpus cannot be used for keywords extraction. Details about extracting key n-grams, and keywords and phrases and grouping documents in clusters are given in papers [12, 13].

The component Answering Processor finds an answer for asked question. The procedure uses cosine similarity for finding similarity between documents, questions and answers, and it is precisely described in [14-16].

## M-GOVERNMENT CLIENT SIDE

The previous chapter describes the logic of the application that is installed on the server side of e-Government. Subset of its most important functionalities has been implemented to display on the mobile client that is designed for the Android Operating system. First reason for choosing Android operating system is its presence in our country, as well as the tendency of growth in the global market. Another reason is that it is an open-source operating system.

In our solution of e-Government, the user has the option of adding the document, search documents based on existing question or based on new one. Mobile application provides an opportunity to review answers to the existing (frequently asked) questions that are in the database, as well as the ability to answer a new question. Selection of a subset of the functionalities was made in order to facilitate the application in terms of memory, and having access to the features that are in urgent need.

Web application for e-Government can be accessed via a mobile phone browser, but it is not optimized for mobile devices, so user can find all the answers to this question, which would require a large amount of data to be transferred over the network. Besides increased cost for the user, this requires a slightly longer response time of the application. The client application provides an overview of the reduced number of responses, and display a large number of responses is done on request. This view is optimized for cell screen.

Figure 2 shows the entire e(m)-Government system. Server side designed for web application for e-Government can be used for m-Government. It is not necessary to design a new server side, but subset of the capabilities of the existing server-side is used. It is only necessary to add a new component that will filter a response before sending it to the mobile client. While the application of e-Government directly addresses the server side and gets unfiltered answers, mobile application addresses one part of the application server, a response is not received directly, but through the component Data Optimizer.

Requests to the server side are sent using the HTTP protocol. They go through all the components as in the case when the request was sent from Web client, and also through the Data Optimizer. This component restricts query results that depend on the number that the user specifies. The result is returned in JSON format, in order to facilitate the display on the application forms.

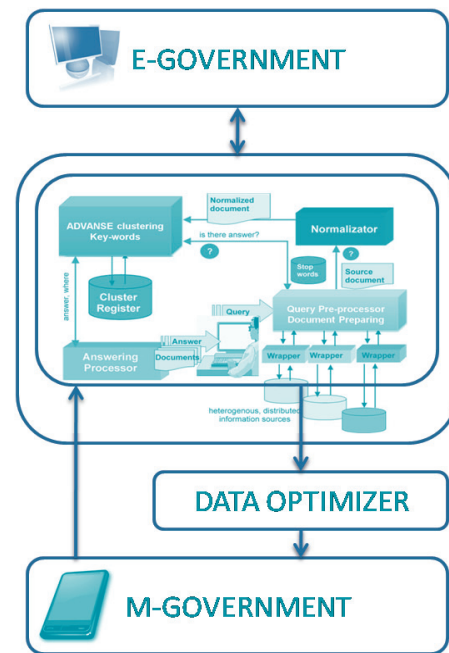


Fig. 2. E(m)-Government system architecture

In the application sliding menu, some of four options can be chosen: Početna (Home), Postojeće pitanje (Existing questions); Novo pitanje (New question); Česta pitanja (Usual questions) and Podešavanja (Settings). In the first tab there are news, and the final tab contains information about the application. Operation of the system is illustrated by tabs Postojeće pitanje (Existing questions) and Novo pitanje (New question). Due to the sensitive data of public services, the demo version was used repository which consists of several documents that represent previous scientific papers from authors regarding this paper, so the choice of questions is in this context. The appearance of tab Postojeće pitanje is shown in Figure 3 and Novo pitanje is shown on Figure 4 and the results of the elections of these options.

When the user selects an option Postojeće pitanje, it opens screen with ten questions, but at the bottom there is an option Učitaj još pitanja (Load more questions), so the user can choose in case that his question was not contained in the offered questions.

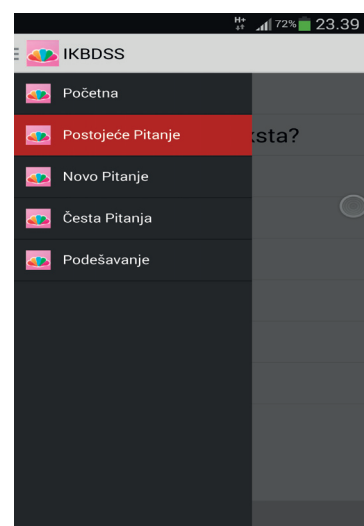


Fig. 3. Application options in sliding menu

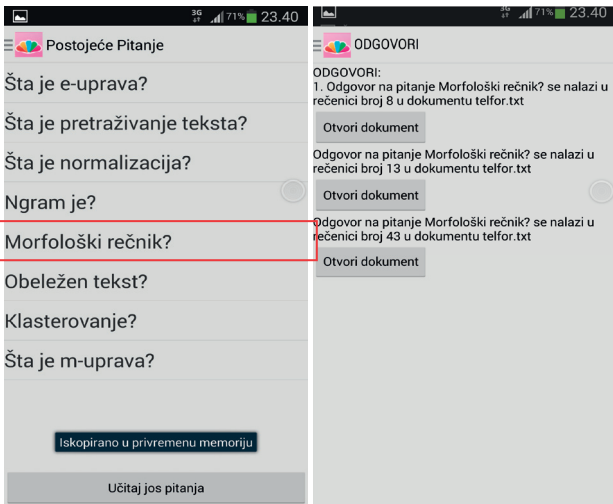


Fig. 4. Tab Postojeće pitanje (Existing questions)

After selecting the existing question, the application sends a request to the server side, which processes the request, and then returns the response to the user as a set of documents in which the answer is, as a collection sorted by degree of similarity with the question, i.e. the most relevant documents are sent first. If the number of responses exceeds ten, all of them will not be shown, but the top ten, while others can be downloaded on the user's request. Here there is an option for viewing and downloading the document.

If the user chooses not to ask the existing question, or determine that it is not there, by selecting a tab a Novo pitanje (New question) there is the possibility of entering new question, and the system will give a set of answers to the question, as in the previous case.

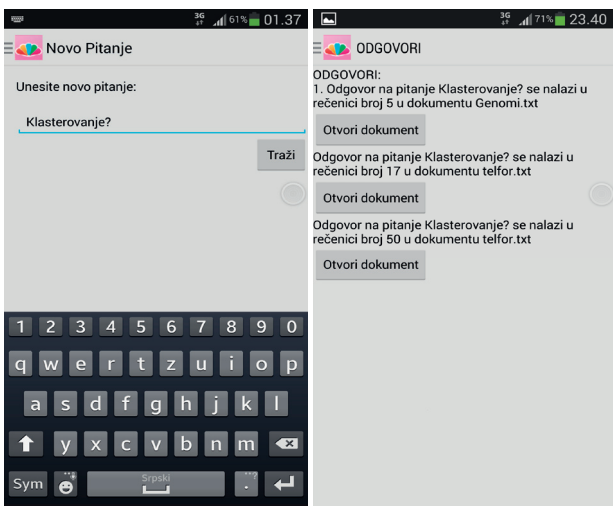


Fig. 5. Tab Novo pitanje (New question)

**CONCLUSION**

M-Government is not yet in an expansion of use in our country. This paper described how easily mobile client can be integrated into existing environment of e-Government, and the result is a comprehensive set of benefits, from ease of use, through better access to reduced communication costs.

In further study, authors will work on improve and optimize both the client and on the server side of the sys-

tem. Also, popularization of the system will be performed, and with the real repository of documents, and improving data security according to the legal regulation, the system will become a substitute to the public sector services in the country.

**REFERENCES**

- [1] S. Gang, „Transcending e-Government: a Case of Mobile Government in Beijing“, The First European Conference on Mobile Government, Brighton, July, 2005
- [2] G. Narayan, „Addressing the digital divide: E-Governance nad M-governance in a hub and spoke model“, EJISDC, vol.31, no.1, pp.1-14, 2007
- [3] P. Rossel, M. Finger, G. Misuraca, „“Mobile” e-Government Options: Between Technology-driven and User-centric“, Electronic Journal of e-Government, vol. 4, no. 2, p. 79-86, 2006
- [4] E-Europe+ 2003 (2001): A co-operative effort to implement the Information Society in Europe Action Plan, The Candidate Countries with the assistance of the European Commission.
- [5] Nacionalna strategija održivog razvoja, Službeni glasnik Republike Srbije, br. 57/2008.
- [6] Prlja, D. i Reljanović, M. (2010): Pravna informatika, Službeni glasnik, Beograd, str. 116-117.
- [7] Schneider, G. (2010): Electronic Commerce, Boston, Thomson Course Technology, pp. 229-239.
- [8] Strategija razvoja elektronske uprave u Republici Srbiji za period od 2009. do 2013. godine, Službeni glasnik Republike Srbije, br. 83/2009 i 5/2010.
- [9] Strategija razvoja informacionog društva u Republici Srbiji do 2020. godine, Službeni glasnik Republike Srbije, br. 51/2010.
- [10] E.Kajan, A.Pljasković, A.Crnišaniin, „Normalizacija tekstualnih dokumenata na srpskom jeziku u cilju efikasnijeg pretraživanja u sistemima e-uprave“, Etran, Zlatibor, jun, 2012
- [11] U. Marovac, A. Pljasković, A. Crnišaniin N-gram analiza tekstualnih dokumenata na srpskom jeziku, Telfor, Beograd, novembar, 2012
- [12] D. Subotić, N. Forbes, „Serbo-Croatian language – Grammar“, Oxford :The Clarendon press, str.25-31, 61-64, 101-113
- [13] M. Mansur, N. UzZaman , M. Khan “Analysis of n-gram based text categorization for Bangla in a newspaper corpus” , ICCIT 2006, Dhaka, Bangladesh, decembar 2006.
- [14] U. Marovac, E. Kajan, G. Šimić, ” A solution of semantic clustering of text documents”,CPPMI 2012, Novi Pazar, jun 2012
- [15] G. Šimić, E. Kajan, Z. Jeremić, D. Randelović. (2012): An Approach to Document Clustering using Hybrid Method, IADIS e-Society Conference, Berlin, March, 10-13, 2012, pp. 153-159
- [16] A. Crnišaniin, A. Pljasković, U. Marovac, E. Kajan. (2013): One solution of searching text documents in Serbian language, ICIST 2013, Kopaonik, March, pp. 28-23.