
LOGICAL-ALGORITHMIC PROCESSING OF TEXT INFORMATION IN THE ELECTRONIC GOVERNMENT SYSTEM

**AkhramKhasanovichNishanov,
SaidrasulovSherzodNorboyo'g'li,
BabadjanovElmuradSatimbaevich**

DSc, Professor, Department of Software of Information Technologies, Tashkent University of Information Technologies named after Muhammad Al-Khwarizimi, Tashkent, Uzbekistan.

External PhD student of Mukhammad Al-Khorezmiy Tashkent University of Information Technologies.

, E-mail: sh.saidrasulov@mininnovation.uz.

Nukus branch of Tashkent University of Information Technologies, PhD of the Department of Information Security

ABSTRACT

The article examines the issue of automatic processing of text appeals of individuals and legal entities in the e-government system, proposing a logical mathematical and algorithmic support for the effective use of natural language to provide information services in public administration reform. The study proposes a model of processing text documents, which allows to classify the texts of written appeals of individuals and legal entities in automated systems of government agencies. The algorithm for solving the problem of determining the purpose of natural and legal appeals of individuals and legal entities, as well as the necessary bases for the automatic identification of the organization serving written appeals were expressed mathematically and the mathematical solution of turning appeals into needs and its search was studied.

Keywords: E-Government, document formalization, keyword database, classification and machine learning, intelligent search method.

Introduction

The most pressing issue of the e-government system is the virtual presentation of information on the legitimate interests of individuals and legal entities and the competent response to them in ensuring the transparency of public authorities and administration. Electronic representation of complaints, appeals or proposals of individuals and legal entities by the legislative and executive authorities means the automation of public services in fields where information and communication technologies are applied. At the same time, the number of e-partnerships is constantly growing.

Automatic analysis of electronic applications of individuals and legal entities and their division into classes certainly leads to the issue of textual information analysis in natural language. This depends on the types of semantic, syntactic, morphological and linguistic structure of natural languages and an advanced development of respective softwares and algorithms. Many scholars are working on analysing a natural language and classifying a document according to its title or the text which makes it up.

Analysis of research publications on formalisation of text documents

The work of P. Kozlov [1] reviews the stages of automated analysis and methods of formalization of text documents. The sorting method used by modifying the linguistic features of text documents is proposed as a result of the morphological and syntactic stage. He studies the stages of formalizing text documents by dividing them into classes as shown in Figure 1.

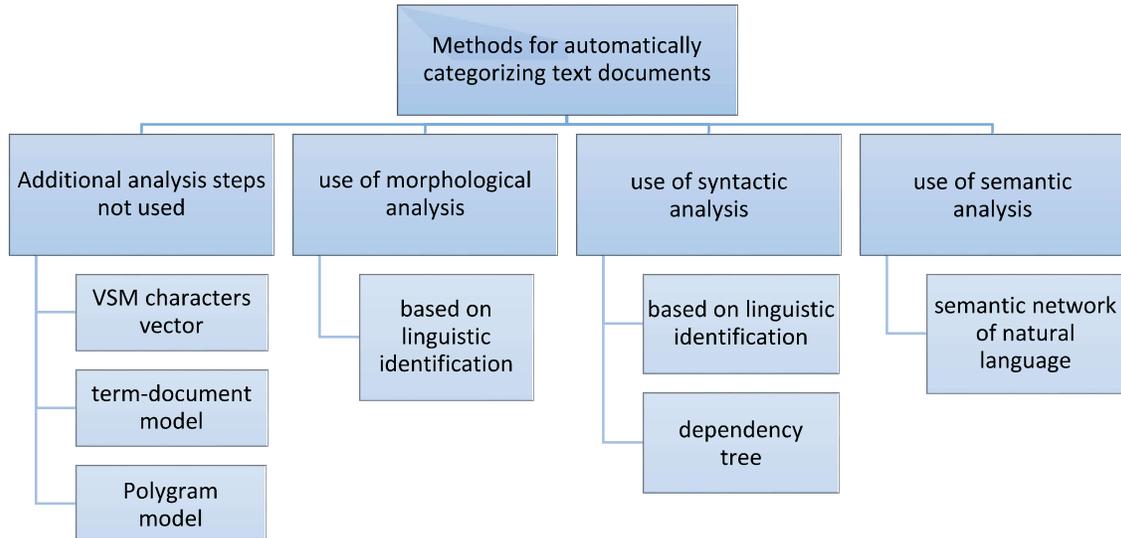


Figure 1. Methods of formalizing text documents at different stages of analysis

As a result of the study, a model of processing text documents which allows to classify the texts of written appeals of individuals and legal entities in automated systems of government agencies has been proposed.

A comparison of modern methods of solving the problem of classification of texts, the search for development trends in this field and the selection of the best algorithms for practical application can be seen in the work of T. Batura [2]. The article provides formalization of the task of classifying texts, and the solution of this problem is divided into the following four successive stages:

- pre-processing and indexing of documents;
- reducing the measurements of the feature space;
- constructing and training of classifiers with the help of machine learning;
- assessing the quality of classification.

Classification and machine learning

Nowadays, popular and modern approaches of classification are based on machine learning methods. This study examines the most common algorithms for constructing classifiers, experiments with them, and the results of these experiments, including available online scientific publications performed in 2011-2016 and materials published in reputable journals highly valued by the scientific community.

The article analyzes the effectiveness of various methods of classification on features such as accuracy, completeness, algorithm run time, the ability of the algorithm to work in incremental order, the amount of initial information in the classification, dependence on natural language. Most of the experiments considered were conducted on English texts. Overall, Russian researchers have been shown to be in need of a more in-depth study of the problem of automatic processing of textual information in order to compare different methods of classifying Russian texts.

The work co-authored by [3] focuses on solving the problem of constructing a classification mechanism using the decision tree method. In this work, the decision tree algorithm for classifying documents consisting of Uzbek texts is used to determine the values of attributes and increase the efficiency of the algorithm, as well as the results of experimental work. The experiments were performed by determining the values of the attributes (Information Gain, Gain Ratio, Gini index) of the decision tree algorithm for classifying documents consisting of Uzbek texts. The information was obtained from the official state source of the National News Agency of Uzbekistan. Six

hundred pieces of news belonging to six categories were obtained, and the classification accuracy in the software developed for the decision tree algorithm was 64%.

The method of intellectual search of data

In the work of A. Nishanov and E. Babadjanov [4-6], who proposed a unique method of intellectual searching for information from the information environment, a combination of searching for information from the information environment and analyzing the found data in a unique way, logically and physically close to the sought information the issue of presentation of the result is considered. The task is to determine the degree of closeness of textual information S to a given set of keywords A. (Fig. 2).

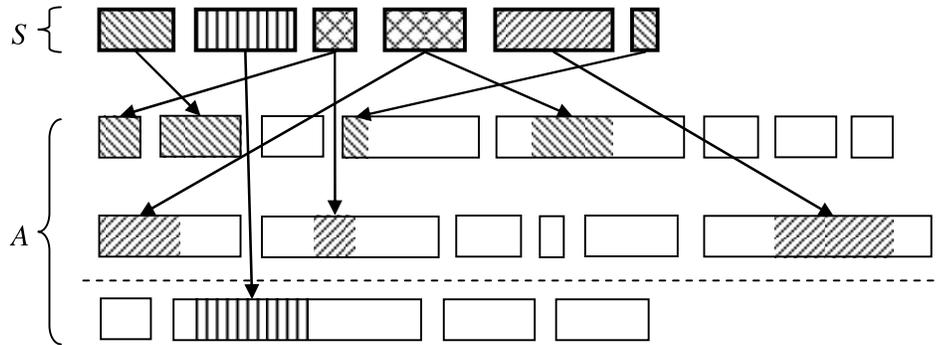


Figure 2. An example of a multi-criteria search task for texts using a keyword database

In solving the problem, the elements of set C are included in the metric that calculates the degree of similarity (in percent) to the elements of set A and the distances between the indices of the elements. The solution is to determine the distances and errors between the elements as a minimum, and the maximum proximity of the elements. To determine proximity, it is proposed to determine whether each element of the object S is related to the elements of each object in the set of objects A, or in other words, to declare a combination search and separate the results based on different criteria depending on index and weight. For a 7-step algorithm, the natural language structure does not matter. It was noted that the proposed algorithm can be used by information system operators by changing the selection criteria as required.

The task of satisfaction of citizens' appeals in the Uzbek language from the database directed to one field can be found in the work of E. Babadjanov. In this author's research, the problem of automatic translation of query texts into SQL using a keyword database describing the database was solved on the basis of specific methods and algorithms.

The study addresses processing applications of individuals and legal entities to government agencies through e-government portals [7]. At present, the process of processing appeals is carried out only by back-office staff. Currently, there is a need for a software module "automatic recognition of appeals" to reduce human labor in the analysis and referral of many appeals. This software mainly deals with the purpose of the application and the definition of the object (organization) where an application is referred.

The following text of the appeal can be cited as a model example for the process of analysis of appeals of individuals and legal entities:

I, Anvar Sodiqov, live in 1234 house, Uvaysi Street in Jizzakh. I would like you to help me get an internet connection to my apartment.

It is clear from the text that the issue in the appeal was “Internet connection”. So, how a decision made by human thinking is determined by software. The appeal in this sense can also be expressed in other words. For example:

I am Anvar Sodiqov, who lives in 1234 apartment, Uvaysi Street, I ask you to connect Uzonline for my house.

or

My house is not connected to the Internet via the landline. Help solve this problem.

It should also be borne in mind that the application may have several requirements. The main purpose of the field under consideration in this publication is to identify the things requested from the text of the appeal, i.e. the subject. Knowing that the request is a request or offer and clarifying the information about the applicant is a secondary goal.

Algorithm for solving the task of determining the purpose of natural Uzbek language appeals

In general, the algorithm for solving the task of determining the purpose of appeals of individuals and legal entities in the natural Uzbek language consists of the following steps:

- 1) creating a database of keywords, i.e.: a database of keywords and their synonyms; database of sentence member separators; word suffix database; database representing the interrogative part of the sentence; database representing the importance of words by fields, etc.;
- 2) creating a database of needs (classified template of appeals), which stores the formalized form of the text of the appeal;
- 3) the base of rules of interrogative sentence structure in the Uzbek language;
- 4) Methods of syntactic, morphological, semantic analysis of the text in the Uzbek language through the keywords and rules of sentence structure;
- 5) the method of indexing the parts that make up the text of the appeal;
- 6) index text formalization algorithm;
- 7) algorithm for searching the formalized text from the needs database;
- 8) algorithm for making decisions on the field the application belongs to on the basis of search results;

The tasks can be divided into two categories: the creation of specific databases and the development of word processing mechanisms (mathematical model, algorithmics, methods, etc.) on the basis of these databases.

We express it in a simple way so that the general process scheme of the matter is clear. Initially, the mathematical representation of these specific bases can be given as follows:

- The text of the application submitted – $X = \{x_i\}$
- Fields the applications refer to – $S = \{s_k\}$
- Keyword Database (KDB) – $A = \{a_j\}$
- Separation of words indicating the main subject from the text by KDB and indexing these words – $G = \{g_i^j\}$
- Needs Database (NDB) – $E = \{e_l = \langle \hat{g}_t, k, p \rangle, \bar{l}\}$. Here e_l – formalized needs are multi-parameter: $\hat{g}_t = \cup g_i^j$ – may be one or more objects, p is the degree of significance of the fact that the appeal belongs to the k field; $\bar{l} \Leftrightarrow l (l < \bar{l})$ – equivalence of needs

From these definitions it can be mathematically expressed that the conversion of the text to the need is carried out by the operator $R(A)$:

$$X \xrightarrow{R(A)} Y.$$

Then determine the result of the search for Y , which has reversed to a need, from E , which is needed. If the search does not yield results, it is registered in the database as a new need as follows:

$$r = \max_p NS(Y, E)$$

if $r = 0$ *then insert* (Y to E)
 $result(s_k)$

The above is an initial mathematical view of the process of determining which field the reference text belongs to through an automatic keyword database. In order to increase the real efficiency of work, it is necessary to develop special mathematical and algorithmic tools and process models to formalize the text in the form of need and search for it, covering all the tasks.

Conclusion

Specific required bases for the task of automatic identification of the organization serving text appeals of individuals and legal entities through the portals of the current e-government platform were mathematically expressed and the mathematical solution of the appeal and its search has been studied.

The issue of automatic processing of text appeals of individuals and legal entities in the e-government system was studied, and a logical mathematical and algorithmic support for the effective use of technical equipment and means of communication in natural language to provide information services within public administration reform was proposed.

A process model of processing text documents, which allows to classify the texts of written appeals of individuals and legal entities in automated systems of government agencies, was proposed.

An algorithm for solving the task of determining the purpose of appeals of individuals and legal entities in the natural Uzbek language was proposed.

REFERENCE

1. П.Ю.Козлов. Способы представления текстовой информации при автоматизированном рубрицировании коротких текстовых документов // Software & Systems. 4 (30) 2017. С.678-683. DOI: 10.15827/0236-235X.120.678-683.
2. Т.В.Батура Методы автоматической классификации текстов // Software & Systems. 2017. Т.30. №1. С.85–99. DOI: 10.15827/0236-235X.030.1.085-099.
3. Nishanov A.Kh., Akbaraliev B.B., Samandarov B.S., Akhmedov O.K., Tajibaev S.K. An algorithm for classification, localization and selection of informative features in the space of politypic data// Webology, Volume 17, No 1, 2020, Pages: 341-364.
4. A. Kh. Nishanov, O. B. Ruzibaev, J. C. Chedjou, K. Kyamakya, Kolli Abhiram, Perumadura De Silva, G.P. Djurayev and M.A. Khasanova Algorithm for the selection of informative symptoms in the classification of medical data// Proceedings of the 14th International FLINS Conference «Developments of Artificial Intelligence Technologies in Computation and Robotics». №12. - P.647-658 (2020). https://doi.org/10.1142/9789811223334_0078
5. Nishanov A.Kh., Saidrasulov Sh.N., Babadjanov E.S., Mamasaidov U.E., Toliev Kh.I. Mathematical Statement of Dynamic Factors Affecting the Development of Electron Government. International Journal of Engineering Research and Technology. 2020, 13(12), p. 5240-5246.
6. A.Kh.Nishanov, Akbaraliev B.B., Sh.Kh.Tajibaev. About one feature selection algorithm method in pattern recognition//11 th World Conference “Intelligent System for Industrial Autamation” (WCIS-2020). –P. 103–112
7. Nishanov A. Ruzibaev O. Tran N. Modification of decision rules 'ball Apolonia' the problem of classification// 2016 International Conference on Information Science and Communications Technologies, ICISCT 2016 (2016). DOI: 10.1109/ICISCT.2016.7777382.