M. Dli, *The Branch of National Research University MPEI in Smolensk, Russia, midli@mail.ru*
O. Bulygina, *The Branch of National Research University MPEI in Smolensk, Russia, baguzova_ov@mail.ru*
P. Kozlov, *The Branch of National Research University MPEI in Smolensk, Russia, originaldod@gmail.com*
G. Ross, *MIREA — Russian Technological University, Moscow, Russia, ross-49@mail.ru*

Developing the economic information system for automated analysis of unstructured text documents[1]

The study of tasks and methods of automated text rubrication was conducted and their prospects for the analysis of unstructured electronic text documents were evaluated taking into account the peculiarities of appeals received from citizens to the authorities. The architecture of the information system of automated analysis of such documents is developed. It implements the proposed multi-model approach to the rubrication based on the integrated use of intelligent and probabilistic-statistical methods. The procedure of processing citizens' appeals received by the authorities using the document management system and the developed information system is given.

Keywords: fuzzy logic, neuro-fuzzy classifier, fuzzy decision tree, unstructured electronic text document, rubrication, economic information system, automated analyzing text.

Intoduction

The analysis of electronic text documents written in natural language is one of the most important tasks implemented in systems of automated analysis of linguistic information.

Currently, a huge amount of information has been accumulated in numerous text databases stored in personal PCs, local and global networks. Every year the volume of such information is rapidly increasing due to the extensive informatization of various areas of human activity.

Appeals of citizens to various municipal, regional and federal government bodies are a type of such textual information.

The analysis of electronic text messages sent by citizens to the authorities (complaints, appeals, suggestions, etc.) has allowed formulating the following features:

• small size of the incoming document essentially complicates its statistical analysis;

• the lack of structuring (special markup for machine processing) complicates information retrieval procedures;

• the presence of grammatical and syntactic errors leads to the need for additional preprocessing procedures;

• thesaurus nonstationarity (variability of composition and importance of words), due to the constant change in the regulatory framework, leads to the need for dynamic rubric clustering.

It is obvious that these features of such unstructured electronic text documents (UETD) impose significant restrictions on the applied methods of automated analysis and processing of incoming appeals.

It should be noted that the known methods, models and algorithms of extracting knowledge

---