# Rubrication of text documents based on fuzzy difference relations

*V. Borisov[1], O. Bulygina[1*], M. Dli[1], P. Kozlov[1]*
*[1]The Branch of National Research University MPEI in Smolensk, Russia*
*\* baguzova_ov@mail.ru*

**Resume.** One of the key areas of informatization of public authorities is to develop and implement the systems of automated processing the electronic appeals (applications, complaints, suggestions) of individuals and legal entities that arrive on official websites and portals of government. The rubrication plays an important role in solving this problem. It consists in the appeals' distribution according to thematic rubrics determining the directions of the activity of departments carrying out processing and preparation of the corresponding response. The results of the analysis of the specific features of such text messages (small size, markup lack, the errors' presence, thesaurus unsteadiness, etc.) confirmed the impossibility of using traditional approaches to rubrication and justified the feasibility of using data mining methods. The article proposes a new approach to the analysis and rubrication of electronic unstructured text documents arrived on official websites and portals of public authorities. It involves the formation of a tree-like structure of the rubric field, based on fuzzy relationships of differences between the syntactic characteristics of documents. The analysis is based on determining the fuzzy correspondence of these documents by their syntactic characteristics with the values of the clusters' centers. It is carried out sequentially from the root to the leaves of the constructed fuzzy decision tree. The proposed rubrication method is programmatically implemented and tested in the automated processing and analysis of appeals (applications, complaints and suggestions) of citizens entering the Administration of Smolensk Region. This made it possible to ensure prompt and high-quality updating of rubrics and document analysis under conditions of non-stationary composition of the thesaurus and the importance of rubric words.