

DOI: 10.24411/1993-8314-2019-10055

*O. Bulygina, the Branch of National Research University MPEI in Smolensk
Russia, baguzova_ov@mail.ru*

*T. Kakatunova, the Branch of National Research University MPEI in Smolensk
Russia, tatjank@yandex.ru*

*P. Kozlov, the Branch of National Research University MPEI in Smolensk
Russia, originaldod@gmail.com*

Analysis of short unstructured documents using fuzzy significance scales and special procedures for economic information integration¹

The article proposes a new approach to the automatic analysis of short messages arriving at Internet portals and e-mails of public authorities. The developed model allows to classify short unstructured text documents in a lack of statistical information and a low degree of thematic rubric intersection. The input data for the algorithm for constructing the model is the set of rubrics and the training sample. Its result is fuzzy scales of significant words in thesaurus of the rubrics, which ensures the correct presentation of the document characteristics and the operation of the classification (rubrication) algorithm.

Keywords: unstructured text document, text rubrication, fuzzy significance scales

Introduction

The formation of e-government is one of the key areas of the State program of the Russian Federation «Information society (2011-2020)». The active use of information and telecommunication technologies by public authorities should increase the effectiveness of state and municipal government and contribute to building partnerships with citizens and businesses.

According to the United Nations E-Government Survey 2018, Russia is one of the leading countries in e-government development and takes

35th place in the rating by EGDI (E-Government Development Index) [12].

One of the tasks of the e-government concept is the organization of effective interaction between public authorities and citizens, i.e. G2C model (government-to-citizen). In modern conditions, this model involves not only the provision of state and municipal services in electronic form but also the organization of feedback.

Every day, a large number of citizens' appeals in the form of an electronic text document are received on Internet portals and e-mails of public authorities and local self-government. The efficiency and completeness of the response to incoming messages directly depend on the correct choice of the addressee responsible for the issue under consideration. This raises the challenge of rubricating the electronic citizens'

¹ The reported study is funded by RFBR according to the research project 18-01-00558 A «Neuro-fuzzy methods of decision-making support for managing complex systems on the basis of dynamic information classification».