ABSTRACT

The thesis is presented in 63 pages. It contains 2 appendixes and bibliography of 43 references. 3 figures and 9 tables are given in the thesis.

Topic relevance. Nowadays, the media are an integral part of our lives, and representatives of the media market constantly compete with each other. It is very important for the user to submit the content that interests him.

The task of personalization is very relevant, it is important for user content to analyze their preferences and to offer content that potentially returns them to the media resource.

Methods of text mining and its classification provide a solution to a given problem.

Thesis connection to scientific programs, plans, and topics. The thesis was prepared according to the scientific research plan of the Applied Mathematics Department of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute."

Research goal and objectives. The goal of this thesis is to develop a decision support system for text content that is expedient or inappropriate to send to the user.

To accomplish this goal, the following objectives were reached:

- clustering content to identify clusters whose content is interested in a specific reader based on their past preferences;
- compare the new content with the obtained clusters;
- ranking of content for each reader according to his preferences and editorial policy.

Object of research is processes of clustering of text documents in order to provide the content that are interesting for users puts (in order to meet user requirements).

Subject of research is models and methods of text mining.

Methods of research. To solve the task, the following methods were used: methods of text mining; methods of text clustering; methods of the theory of algorithms and programming (for implementing the developed algorithms).

Scientific contribution consists of the following:

– improved methods of analysis of the text (due to the fact that the methods have been chosen for the solution of a particular task from the list of all considered, for which the advantages and disadvantages of the task are described in detail);

– It is proposed to cluster not only the articles for submission of the user, but also its revised articles, taking into account the CTR for each user.

Practical value of obtained results. Development of media resources in the direction of personalization. To present the user exactly the content that interests him is very important for this. This problem can be solved by classifying the preferences of users, the content of the resource.

Approbation of the thesis results. Basic ideas and results of the research were presented at PMK-2018.

Publications. Thesis results are published in 1 scientific works:

in 1 papers in proceedings of scientific conference.

Keywords: text clustering, text mining, *TF-IDF*.