

## ABSTRACT

Master dissertation: 99 pp., 19 fig., 22 tab., 1 app., 63 sources.

**Topicality.** The need for automatic processing of text documents is now extremely high and is steadily increasing. This is due to the daily increase in textual information on the expanses of the World Wide Web. According to March 2016, there are about 4.66 billion pages on the Internet, and this number only includes pages that are indexed in the most widely used search engines. Therefore, without computer processing, it is not possible to analyze such information volumes in a reasonable time.

One of the tasks of intellectual analysis of texts is their classification on the given categories, which needs to be solved in different spheres of human activity. So, for the purpose of securing information and public security, the analysis of SOCIAL NETWORK DATA, AND EVERYDAY BLOGS is important for the purpose of identifying data related to terrorism, drug trafficking, etc. Also, in commercial and public activities, there often is a need to process reviews and comments, in order to identify their emotional color (negative or positive), their distribution on further processing between different Subdivisions, etc. In the first and second examples, the task of classifying textual information among categories in terms of time constraints and resources of computing devices arises. Therefore, the task of automatic and qualitative classification of data at an acceptable time, without prior structuring of information, since structuring requires additional resources, time and can not always go through without loss of important information, is a task that is worth attention and research.

**Relationship of work with scientific programs, plans, themes.** The research was carried out at the Department of Computer-Aided Management And Data Processing Systems of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» within the theme «Methods and Technologies for High-Performance Computing and Processing of Large-Size Data Arrays. » State registration number 0117U000924.

**The aim of the research** is to improve the quality of monitoring of media activity by developing an algorithm for automatic analysis of text information, which increases the accuracy and completeness of the classification.

To achieve this goal, the following tasks must be performed:

- analyze algorithms and methods of machine learning for solving the problem of automatic classification of texts;
- choose a model for presenting textual information in the classifier;
- develop an algorithm for pre-processing texts in accordance with the selected model of presentation of textual information;
- develop a modified method for classifying textual information;
- implement the program realization of the developed algorithm of automatic classification of textual information;
- prepare data for the assessment of the quality of the classification;
- to conduct research on the effectiveness of the developed information technology.

**The object of research** is the process of classifying unstructured text arrays of information.

**The subject of research** there are technologies and methods of intellectual analysis of textual information.

**Research methods** are methods of machine learning, which based on text mining methods.

**Scientific novelty of the obtained results.** The modified indexing method on the basis of the statistical algorithm Viterbi with the connection of the grammatical rules of removal of morphological homonymy has been developed.

**Publications.** The results of the research are published in the journal "Scientific Review" [1] published in the theses of the scientific and practical conference mathematical and simulation modeling systems. MODIS "2017" [2], published in the theses of the scientific conference of students, graduate students and postgraduates "Informatics and Computing" - IOT-2018 [3].

MACHINE LEARNING, DATA MINING, TEXT MINING,  
CLASSIFICATION OF TEXTUAL INFORMATION, CONTENT ANALYSIS,  
CLASSIFICATION ALGORITHMS