

ABSTRACT

Relevance: to simplify the search for relevant information among scientific publications in Ukraine, a library classification is used. However, this system is not perfect at this time, because classification is erroneous, and in some cases it is executed for the journal as a whole, which results in partial discrepancies for some of its articles. Also, it takes a long time to perform the classification by a third party (such as a librarian or editor). The solution to this problem is to automate the classification process. By using machine learning, automatic classifier can be created, which will improve the accuracy of the classification compared to manual and accelerate the classification of new revenues.

Purpose: create a classifier of scientific articles by UDC categories based on machine learning.

To achieve this goal, the following tasks were formulated:

- systematization of existing text data classification algorithms;
- gathering sufficient training data, developing a classifier based on machine learning;
- testing and analysis of the efficiency of the obtained algorithm;
- determining the further direction of research.

Object of study: library classification of scientific articles.

Subject of study: algorithms for classification of text data.

Research methods: naive Bayes classifier, neural networks, backpropagation algorithm were used to solve this problem.

Scientific novelty: the most significant scientific results of a master's thesis are the study of the possibilities of automation of the classification of scientific texts; search for mistakes in already classified texts; creation of classification algorithms for distinguishing categories in texts of similar subjects.

The practical value of the obtained results is determined by the fact that the proposed algorithm allows to achieve the accuracy of library classification in 86%, which allows to use it for finding and correcting errors in the classification of texts, as well as an aid in the classification of new receipts.

Relationship with working with scientific programs, plans, topics: work was performed at the Department of Automated Information Processing and Management Systems of the Igor Sikorsky National Technical University of Ukraine «Kyiv Polytechnic Institute» within the topic «Mathematical Models and Technologies in DSS». State Registration Number 0117U000914

Approbation: the main provisions of the work were reported and discussed at the XIII Scientific and Practical Conference of undergraduate and graduate students «Applied Mathematics and Computing» (AMP-2019), as well as at the third all-Ukrainian scientific and practical conference of young scientists and students «Information Systems and Technologies of Management» (ISTM-2019).

Keywords: text classification, neural networks, UDC, machine learning.