ABSTRACT

Master dissertation: 94 pp., 22 fig., 19 tab., 1 app., 20 sources.

Topicality. Propaganda has existed in traditional media for many years, and with the development of the Internet, it is gradually penetrating social media. This is due to the daily increase in textual information on the Internet. Indeed, propaganda is so powerful that everyone is prone to it. The main channel through which the society accepts propaganda, are the means of mass communication. The statistics of the credibility of the media is impressive. Information dissemination is an industry with a turnover of more than four hundred billion dollars per year, half of them are spent on mass propaganda. Every person thinks verbally and is therefore more or less influenced by what sales professionals, politicians, journalists, fraudsters, sect organizers, special services and terrorists skillfully use.

Consequently, a necessary task is to create an informational and psychological security tool designed to check textual content for the presence of special linguistic structures and phrases that facilitate non-critical analysis of information.

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 is the development of an algorithm designed to check the text content for the presence of special linguistic structures and turns, which contribute to the non-critical analysis of information.

To achieve this goal it is necessary to perform the following tasks:

- analyze the algorithms and methods of machine learning to solve the problem of automatic text classification;

- select a model for the presentation of textual information in the classifier;

- develop an algorithm for preprocessing of texts in accordance with the selected model for the presentation of textual information;

- develop a modified method for classifying textual information;

- perform the software implementation of the developed algorithm for automatic classification of textual information;

- conduct research on the effectiveness of the developed information technology.

The object of research is the process of classifying text information.

The subject of research is the methods of classifying text data to identify propaganda.

Research methods are machine learning methods based on text mining techniques.

Scientific novelty of the obtained results. The modified text data classification method has been developed for identifying text data using the previous sentimental analysis, Dirichlet Latent Allocation and the TextRank algorithm. TextRank algorithm was improved by taking into account the position of the word in the text.

Publications. The research results were published in the journal "Modern directions of development of information and communication technologies and management tools" [1], published in the theses of the scientific conference of students, undergraduates and graduate students "Computer Science and Computer Engineering" - IOT-2018 [2], published in the journal "World Congress "Aviation in the 21st century" [3].

MACHINE LEARNING, TEXT MINING, DATA MINING, TEXT CLASSIFIER, CONTENT ANALYSIS, CLASSIFICATION ALGORITHMS