ABSTRACT

The thesis is presented in 68 pages. It contains two appendixes and bibliography of 42 references. Fourteen figures and two tables are given in the thesis.

Topic relevance. Nowadays, cyberattacks are used more and more together with common attacks. Therefore, the detection of information attacks and groups who commit such attacks is topical. Existing cyberattacks can be divided into two classes. Direct methods, such as DDOS attacks, using viruses, which intercept and modify the information flow or block computers. Indirect methods, such as the spread of fake information in news and social networks. Often, information attacks are carried out with the help of bots, but the biggest danger comes from users who first gain the trust and respect of the community, and then try to break it by spreading controversy and despair. As information spreads quickly in social networks, it is important to develop a system for identifying such accounts on Facebook and Twitter, as the most popular social networks.

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 system for collecting and analyzing the activity of user accounts in social networks.

To accomplish this goal, the following objectives were reached:

– systematize existing methods for collecting and analyzing data from social networks;

– develop an algorithm for collecting, filtering and storing data from various sources;

– develop methods for solving the problem of identifying phony accounts and implement them programmatically;

– conduct experimental studies using the collected data.

Object of research is methods of collecting and analyzing data in order to identify phony accounts.

Subject of research is applying data collection algorithms, statistical and natural language processing to identify phony accounts.

Methods of research. To solve the task, the following methods were used: data collection algorithms (for datasets creation); methods of probability theory and mathematical statistics (for automated data labeling); methods of natural language processing (for developing methods for solving the task of detecting phony accounts); methods of the theory of algorithms and programming (for software implementation of developed algorithms).

Scientific contribution consists of the following:

– for the first time, the task of detecting phony accounts is set, which, unlike the existing methods for searching fake accounts, involves taking into account the user behavior when they try to gain trust in the community;

– statistical and natural language processing methods for determining the openion of social network users, which, in contrast to the existing ones, take into account the reactions of all members of the community to the message, which makes it possible to track changes of the account behavior.

Practical value of obtained results. Methods are proposed that can be applied to collect and store large amoun of information from social networks, analyze data and detect phony accounts. The developed methods, mathematical approaches and software simplifies the search for information attacks and phony accounts, contributes to control and prompt response to propaganda and destabilization of the situation in the country and the world.

Approbation of the thesis results. Basic ideas and results of the research were presented at the International scientific and technical conference «International Conference on Military Communications and Information Systems» (2018) and X conference of young scientists «Прикладна математика та комп’ютинг» (2018).

Publications. Thesis results are published in two scientific papers:

– in one publication in the proceedings of the international scientific conference;

- in one publication in the proceedings of the Ukrainian scientific conference.

Keywords: phony account, social network, labeling, natural language processing, mathematical statistics.