

UDC 004.04

SYSTEM DEVELOPMENT FOR NETWORK TEXT DATA ANALYSIS AIMED TO IDENTIFY ADVERSE DRUG REACTIONS

Burlakova Y.A.

(ITMO University, Saint Petersburg)

Supervisor – senior lecturer Dobrenko N.V.

(ITMO University, Saint Petersburg)

Nowadays, detection of drug side effects is a thorny issue. In this work the development of a network data analysis system aimed to detect adverse drug reactions in text and analysis of the results received during system testing are proposed.

Introduction. Adverse drug reactions (ADR) cause big problems to public health, it does a great deal of damage to patient health and increases health care costs. In United States 2 million patients suffer from ADR annually. It is difficult to characterize all the ADRs during premarket trials, as many serious ADRs, particularly rare ones, are discovered years after a drug has been released to market. Clinical trials are limited in their abilities to detect ADRs because of the small test group sizes, short durations and lack of enrolled patient diversity. Meanwhile, a significant number of ADRs are recognized in post-market period, what results in patients' injuries and increases of health care costs. Drug side effects data collected at healthcare facilities are private, but there are many medical forums aimed to support victims of ADR and what is more important such forums are full of comprehensive information about drugs and their side effects in different groups of patients over a long period of time. Thus, the analysis of data collected from network sources is important for this research area.

Purpose of work. The aim of this work is to develop network data analysis system for identifying adverse drug reactions in text and obtained results review.

Main part. The first stage of work is system modeling, second stage is data preparation that includes compiling dictionaries of drug names and ADR. Next part of work includes writing script, testing it on prepared data and analyzing the achieved results. The main methods of detecting drug side effects were analyzed and data preparation methods were presented.

Conclusions. The developed system for ADR detecting will allow to analyze network text data, identify side effects of cancer medicines, detect the frequency of known ADR occurrence and possible causes of it in different groups of cancer patients. In addition, the results of this study can be used to improve the effectiveness of patient care.