ABSTRACT

Thesis: X p., X fig., X tabl., X append., X sources

The theme: "The system for analysis and categorization of textual medical data using SAS technologies".

INTELLECTUAL ANALYSIS OF DATA, INFORMATION SEARCH, TEXT ANALYTICS, MEDICINE, NATURAL LANGUAGE PROCESSING, CLUSTERIZATION, CLOUD COMPUTING.

Relevance of work: for today most of the information is unstructured, so using it with the usual analytical models is impossible. In the medical field, processing text data can greatly improve the quality of feedback from patients, and therefore the quality of drugs.

Object of research: medical text reports, namely patient reviews.

Subject of research: information search, intellectual analysis, method of maximum likelihood, latent semantic analysis, boolean rules.

Purpose: to study existing methods of processing unstructured text data and their implementation in the system of analysis and categorization of text medical reporting.

Research method: use of knowledge about natural language processing (NLP), mathematical methods and models for classification and clustering of text information.

The results of the work: the analysis of modern methods of information search, the possibilities of using existing tools of text analytics, their adjustment for a specific branch, a system developed that allows creating categories with a certain degree of probability to perform the classification task, receiving a large number of medical reports on the entrance is investigated.

Ways of further development of the subject of research - improvement of the chosen architecture of the model, expansion of the collection of texts, markup by experts.