

ABSTRACT

The thesis is presented in 73 sheets. It contains 3 appendixes and bibliography of 15 references. 22 figures and 7 tables are given in the thesis.

The main goal of the system is the automation of the user's personal knowledge management resources, ensuring its necessary mathematical and software for use specialized functions required to optimize knowledge management. The system is created for use in personal knowledge management with providing the basic functions of knowledge management and support functions optimizing the user experience.

During the project held significant comparative analysis of classification methods, including methods C4.5, method of support vectors and Bayes classification method based on Bayesian naive classification model, methods and algorithms of cluster analysis of data, including the EM-algorithm and k-means algorithm of clustering, algorithms of categorization and preliminary isolation of terms of the document. Were selected for use Bayesian classification method based on naive Bayesian classification model and algorithm k-means of cluster analysis.

As a result of diploma projects designed personal knowledge management system that represents the software on the Java programming language using database management systems Oracle 11gR2. The testing of the developed system was done, the testing of the algorithms of document classification and clustering data using documents specified topics, research results correspond to the proper working of algorithms.

The system is designed for personal use by users who own knowledge management system using additional specialized functionality to automate the work with their own data, perform quick document classification for certain topics, check compliance with the chosen theme.

Keywords: knowledge management system, document classification, clustering, document, theme, token, coincidence, probability.