ABSTRACT

The thesis is presented in 57 pages. It contains 1 appendix and bibliography of 16 references. 7 figures and 3 tables are given in the thesis.

The goal of the thesis is to develop mathematical and software tools for analyzing the uniqueness of text documents and digital images and autosummarization of texts.

In the thesis were analyzed existing solutions, of the above-mentioned tasks - Antiplagyat system, Advego Plagiatus program, Unplag service for detecting plagiarism in texts; for authoring texts, Extractive and Abstractive groups such as LexRank, LSA, Gensim's TextRank summarization module were considered and Gensim's TextRank was chosen. They are compared in terms of the accuracy of obtained results, algorithm efficiency and method adaptation to the data. To analyze the uniqueness of texts, the algorithm of shingle was chosen, and for images, the method of sectoral comparison of histograms.

An automated system that implements selected methods was developed. The testing of the developed system was carried out.

The work is implemented by order of LLC «DataRoot».

Keywords: analysis of uniqueness, plagiarism, autosummarisation, text processing, image processing.