

Theme of the Graduation Thesis: «Development of an intelligent system for clustering textual information (based on Russian-language texts) »

Author: Gatieva Anzhela Vladimirovna

Thesis tutor: Candidate of Technical Sciences, assistant professor of department of Information and Communication Technologies, Mathematics and Information Security Timchenko Olga Viktorovna.

Information about the contracting authority: Information about the contracting authority: Scientific and Educational Innovation Complex «Information and Communication and Mathematical Technologies» of Pyatigorsk State University.

Relevance of the research topic: New problems associated with the processing of such a large amount of data appear in conditions of rapid development of web-space and an increase in the amount of stored information.

Objective: to develop an intelligent system for clustering text documents. The system should automatically determine the tags of documents from the collection, which in the future allows you to distribute texts among clusters.

Tasks:

1. Review and analysis of modern methods of text clustering;
2. Statement of the problem of automatic tagging of text as a clustering task;
3. Development of the architecture of an intelligent system;
4. Development of a software product that implements the task of automatically tagging a text document.

Theoretical and practical significance of the research: consists in the development and testing of the proposed technology of clustering and tagging texts proposed as part of graduation qualification.

Results of the research: the result was an intelligent clustering and tagging system for text documents, a mandatory element of which is a user interaction system.

Recommendations: the algorithm for automatically extracting tags should contain a deeper preliminary processing of the text than that implemented in the framework of the graduation qualification work. This will probably improve tagging accuracy.